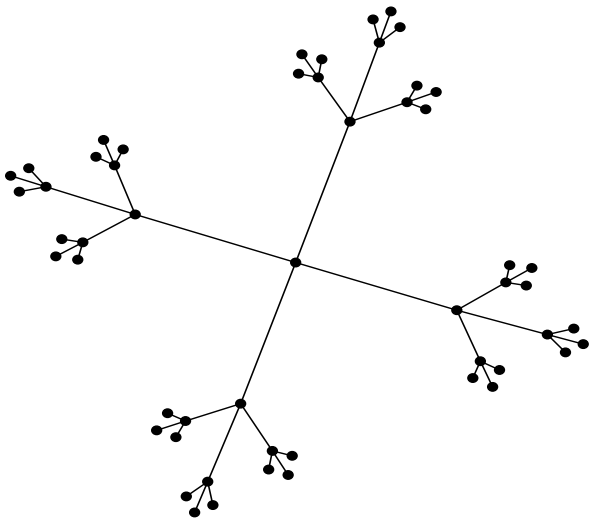
Université de Mons–Hainaut

Institut d'Informatique

On Automated and Computer Aided Conjectures in Graph Theory

Hadrien Mélot



Dissertation submitted in fulfilment of the requirements for the degree of Philosophiæ Doctor in Science (specialization Computer Science)

April 2006

Electronic Version

On Automated and Computer Aided Conjectures in Graph Theory

Hadrien MÉLOT
hadrien.melot@umh.ac.be

Institut d'Informatique
Faculté des Sciences
Université de Mons-Hainaut
Académie universitaire Wallonie-Bruxelles

Dissertation submitted in fulfilment of the requirements for the degree of Philosophiæ Doctor in Science (specialization Computer Science)

April 2006

Composition of the jury :

Professor J. WIJSEN, President, Université de Mons-Hainaut

Professor V. BRUYÈRE, Director, Université de Mons-Hainaut

Professor P. HANSEN, Co-Director, HEC Montréal, Canada

Professor A. HERTZ, Ecole Polytechnique, Montréal, Canada

Professor M. LABBÉ, Université Libre de Bruxelles

Professor T. MENS, Université de Mons-Hainaut

Note about this electronic version of the dissertation : The original printed edition of this dissertation is in a book format (B5) and does not fit good when printed on A4 paper or when viewed on a screen. It follows that the page numbering differ between this electronic edition and the printed edition. However, the contents is exactly the same.

The pictures of the cover and the part's covers are dendrimer representations, produced by the drawing routine of GraPHedron.

Pour Hélène, pour Sofian, pour Sarah

Acknowledgments

I would first like to thank my two co-directors, Professors V. Bruyère and P. Hansen, and my "non-official director", Professor M. Labbé. Chronologically, I would like to thank Professor M. Labbé who showed me confidence during my Master's degree and my DEA, even if I was not exactly the most studious person she met. This confidence became a collaboration which is still continued today. Professor M. Labbé is also the instigator of the project GraPHedron. In 2000, I went to Montréal, to work with Professor P. Hansen, during several weeks. This stay was really my initiation to research. Professor P. Hansen is continuously traveling all around the world. However, all our meetings – in Brussels, Montréal, Rutgers, Puerto de la Cruz or Ghent – were always intensive and fruitful. My sincere thanks go thus to Professor P. Hansen. Finally, I would like to thank Professor V. Bruyère. I started my PhD in Mons, working as her assistant, since 2001. I benefited greatly from her advices during these years. I would also like to thank her for her careful reading of the successive drafts of this dissertation.

I would like express my gratitude to Professors A. Hertz, J. Wijsen and T. Mens who accepted, showing interest, to be in my jury – even if it implies sometimes a quite long travel. I also thank Professor A. Hertz for corrections and discussion about the proofs of some theorems, and Professor T. Mens because he significantly improved the English of some important parts of this dissertation.

My sincere thanks go to my publication's co-authors. I also thank my colleagues and students who helped me in the development of GraPHedron. It was a pleasure to discuss, think, write or code with all of them. They contributed greatly towards the completion of this dissertation. In particular, I would like to thank Professors J. Cardinal, G. Caporossi, J.-P. Doignon, I. Gutman, S. Langerman and D. Stevanović and my colleagues S. Beauport, G. Fasbender, P. Grégoire and E. Levy.

I would like to thank the organizers and participants of the three DIMACS workshops *Computers in Scientific Discovery* (New Jersey 2001, Montréal 2004 and Ghent 2006). I have benefited from many discussions with leading researchers in the field of this dissertation.

I often noticed that research labs and academic departments are lost without the meticulous

and wonderful work of their secretaries. Special thanks are owed to Mrs. L. Bouchez (UMH), F. Van Brussel (ULB) and C. Dufour (GERAD).

I have always received encouragements from family, friends and colleagues from UMH and ULB. I gratefully thank them for it.

I dedicate this dissertation to Hélène, Sofian and Sarah, my tremendous and beloved little family. Hélène, I will show you that my hands are – actually – not sticked to a $\alpha!\otimes?\varphi$ computer. Thank you for your patience and understanding. Sofian and Sarah, you are born after the beginning of this work. You are so amazing. Everyday. You helped me to find motivation and to refresh my mind.

Contents

List of Figures

List of Tables

Preface

As already stressed by Archimedes [12], discovery and proof are different activities, which require different methods. One must first find what is to be proved, i.e., a *conjecture*, by any procedure, possibly aided by a physical model, then prove it or refute it by logical means. Both tasks can be assisted by computers in various fields of mathematics. In this dissertation, we focus on computerized support to graph theory research.

A *graph* G is an ordered pair (V,E), where

- V is a finite, non-empty set whose elements are called *nodes* and;
- E is a set of unordered pairs of distinct nodes of V, called *edges*.

Graphs considered in this dissertation are thus simple undirected and unlabeled graphs.

Objective

The main objective of this dissertation is to answer the following question :

How can computers help researchers to obtain new results in graph theory ?

More precisely we study how computers can help, in graph theory, to :

1. find new conjectures,
2. reinforce existing ones,
3. refute conjectures,
4. find proofs.

The vast majority of problems and results in graph theory are expressed as equalities or inequalities among graph *invariants*. A graph invariant is a numerical value which is preserved by isomorphism. The number of nodes n and the number of edges m of a graph G are examples of graph invariants.

Some computer systems can give conjectures in a fully automated way, i.e., without human intervention except for the problem statement. In general, the computer is used interactively to derive conjectures. For instance, one of the main principles is that the computer can compute very fast, and thus quickly give particular examples of graphs, useful for a given problem. These particular graphs are for instance extremal graphs for a specific graphical property, or simply a counter-example for a conjecture. Most of the computer systems used in graph theory apply this principle in the process of finding conjectures. Moreover, these particular graphs are generally important in the construction of a proof.

However, fully automated proofs in graph theory are still limited to simple properties [77, 78, 83, 84, 102, 103, 104]. In contrast, partly automated proofs, which use both human reasoning and specialized computer programs, have been used with much success since the proof of the 4-color theorem [8, 9, 10, 213] (and despite the controversy on the reliability of such proofs, see e.g., [23]). To illustrate, the fifth update of a "dynamic survey" on "Small Ramsey Numbers" [209] reviews results which were obtained with the aid of the computer in 71 papers among the 274 which are cited. Relying on the computer to complete a difficult proof is thus widespread.

With the intention of answering the above question, the dissertation is divided into three parts :

I. the study of existing *computer systems* and the classification of them in families;

II. the use of one of these systems, called AutoGraphiX, and the presentation of the theoretical results obtained with the help of this particular system;

III. the development of a new system, named GraPHedron, which is based on a polyhedral approach to find conjectures. Under some conditions, these conjectures can be considered as optimal. We also present the first theoretical results obtained with this system.

Part I. Computers and discovery in graph theory

In the first part, we survey the different computer systems which have been used to obtain new results in graph theory [162]. The term *computer system* is used here in a relatively large sense. It can be a software system, a programming language for symbolic computation, or a set of programs.

Basic definitions and notations of graph theory are recalled or introduced in the first chapter. Indeed, there is a wide divergence in graph theoretical terminology.

Chapter 2 is devoted to the existing computer systems which have been used to obtain new

results in graph theory. These systems can be grouped in five main families which are based on different approaches :

1. enumeration;
2. interactive computing;
3. invariant manipulation;
4. generation and selection;
5. heuristic optimization.

Each approach is illustrated and about thirty systems – the most prevalent ones in each family – are mentioned. Two of these approaches are applied in this dissertation.

i) The approach based on a heuristic optimization was introduced by Caporossi and Hansen. They have developed, since 1997, the system AutoGraphiX [55]. AutoGraphiX allows the study of graph invariants, by searching graphs for which these invariants have extremal values. Stating this problem as a combinatorial problem, Caporossi and Hansen use a metaheuristic to solve it. This metaheuristic is the *Variable Neighborhood Search* [198]. In the case of AutoGraphiX, the Variable Neighborhood Search allows, for a given problem, to obtain extremal graphs, or near extremal graphs since it is a heuristic. These extremal graphs can then be studied interactively, and conjectures about them derived, with the help of a graphical user interface. AutoGraphiX also uses several procedures [56], to give conjectures about extremal graphs, in a fully automated way. We present results obtained with AutoGraphiX in the second part of this dissertation.

ii) Using the approach based on enumeration, we wrote a program in *Matlab* allowing the generation of complete split graphs with up to 500 nodes, and the computation of the eigenvalues of their adjacency matrix. We can then deduce some infinite families of graphs for which all eigenvalues are integers [166]. These results are presented in Chapter 3.

Part II. Results obtained with AutoGraphiX

In the second part of the dissertation, we present theoretical results obtained with the help of the system AutoGraphiX. Most of these results have applications in chemistry [146, 163, 165] and constitute an important chapter of this dissertation (Chapter 4). Graphs can be used to model chemical structures in a straightforward way. Nodes represent atoms and edges represent links between them. The number of nodes adjacent to a given node v is termed the *degree* of v. In the case of graphs representing chemical structures, the node's degrees are generally bounded. The most frequent case is to have all degrees ≤ 4, this number being the valency of carbon. Invariants

are then topological indices frequently used in chemistry. We present in Section 4.1 several bounds on the Randić index, one of the most prevalent invariants in the literature on chemical graph theory. Other indices, expressing the level of irregularity (or branching) of a molecular structure, are also studied and compared in Section 4.2 : the index of Collatz-Sinogowitz, the variance of degrees and the irregularity. Finally, relations between the variance of degrees, the maximum degree, the order and the size of connected graphs, are studied in Section 4.3. The most important results, contained in this chapter, were obtained by the study of the extremal graphs generated by AutoGraphiX. Based on their observation, we have stated several conjectures, often quite easily. The vast majority of these conjectures have been proved. To achieve this, we follow the principle mentioned above : a family of particular graphs, significant for a given problem, are identified by the computer system. The characterization of them often led to new formulation of conjectures.

We go further in the study of the irregularity in Chapter 5. In a more general context, no more limited to chemical graph theory, we get, with AutoGraphiX, an upper bound on the irregularity of a graph [164]. Experiments with AutoGraphiX lead to automated conjectures which are presented in Section 5.1 and a conjectured upper bound on the irregularity is obtained in Section 5.2. This bound is tight for all possible values of n and m compatible with the existence of a graph G with n nodes and m edges. The proof of this result is particular (see Section 5.3). Indeed, it is based on the approach used by AutoGraphiX to find extremal graphs. The proof is written as a *simulated algorithm*. Each iteration of this algorithm strictly increases the value of the irregularity of a given graph, by constructing a new one. This algorithm stops when the upper bound is reached. The final resulting graph is one of the extremal graphs generated by AutoGraphiX.

Part III. The system GraPHedron : principles and results

The third and last part of this dissertation is devoted to the newly developed system GraPHedron, which is introduced in Chapter 6. In Section 6.1, we present the principles of GraPHedron and its outline in Section 6.2. This new system is particular as it uses an approach which tries to answer the following question :

What are all the best linear inequalities among a given set of invariants ?

The arguments used to answer this question are geometrical. Here, graphs are viewed as points in the euclidian space $\mathbb{S}^p$, whose coordinates are the values for a set of p invariants I. To the convex hull of these points corresponds a finite set of linear inequalities associated with its facets. This set of inequalities is a minimal description of a polytope (bounded polyhedron), and can be considered as optimal, among the set of invariants I. For graphs with a reasonable order, i.e., with a limited number of nodes n, GraPHedron computes these systems of inequalities (one for each chosen value of n) and presents this information in a report, generated using LaTeX. The convex hulls are also represented by drawing the polytopes (when the number of invariants in I is 2). A

generalization, in terms of n, of these inequalities is then possible in some cases, in two different ways :

a) interactively, by the study of the information contained in the report. Similarities between polytopes can be observed (number of facets, shape of the polytopes,...). Another important source of information about a specific problem comes from the observation and the characterization of the *vertex-graphs*[1]. A vertex-graph is a graph for which the associated point in the space $\mathbb{S}^p$ (vector of coordinates of invariants of I) is a vertex of the polytope. For instance, one can observe that all the vertex-graphs belong to a particular family of graphs. These graphs are then very useful in the formulation of a conjecture, but also in finding its proof.

b) by a fully automated procedure (see Section 6.5). The software uses a numerical procedure which tries to generalize, in term of any number of nodes n, the facets of the polytopes. The automated conjectures are given as inequalities expressed with the invariants of the set $I \cup n$, and representing the facets of the polytopes, for each fixed value of n.

Different applications of GraPHedron are presented in Section 6.3. The forms of conjectures derived with this system are particular and studied in Section 6.4.

Chapter 7 is devoted to the first results obtained with the help of GraPHedron. Indeed, this polyhedral approach has allowed, in particular, to solve a problem mentioned by Ore as open since 43 years [202] : "What is the minimum number of edges of a connected graph, with fixed number of nodes and stability number ?" (see Section 7.1 and [62]). The answer to this problem is a variant of a famous theorem due to Turán [228], which answers the same question in the case of general graphs, connected or not.

In Section 7.2, some relations between the irregularity, the diameter and the maximum degree of connected graphs are given.

The use of GraPHedron also allows to study approximation ratios of heuristics compared to exact algorithms. The heuristic known as the *maximal matching heuristic* gives a 2-approximation of two classical NP-hard problems : *minimum node cover* and *minimum maximal matching*. With the help of GraPHedron, we get more accurate approximation ratios [58], under some conditions on the density of the graphs (see Section 7.3).

Contributions

This dissertation is a collection of several publications [58, 146, 160, 162, 163, 164, 165, 166] and submitted papers [62, 196]. The contributions of this work are of two types :

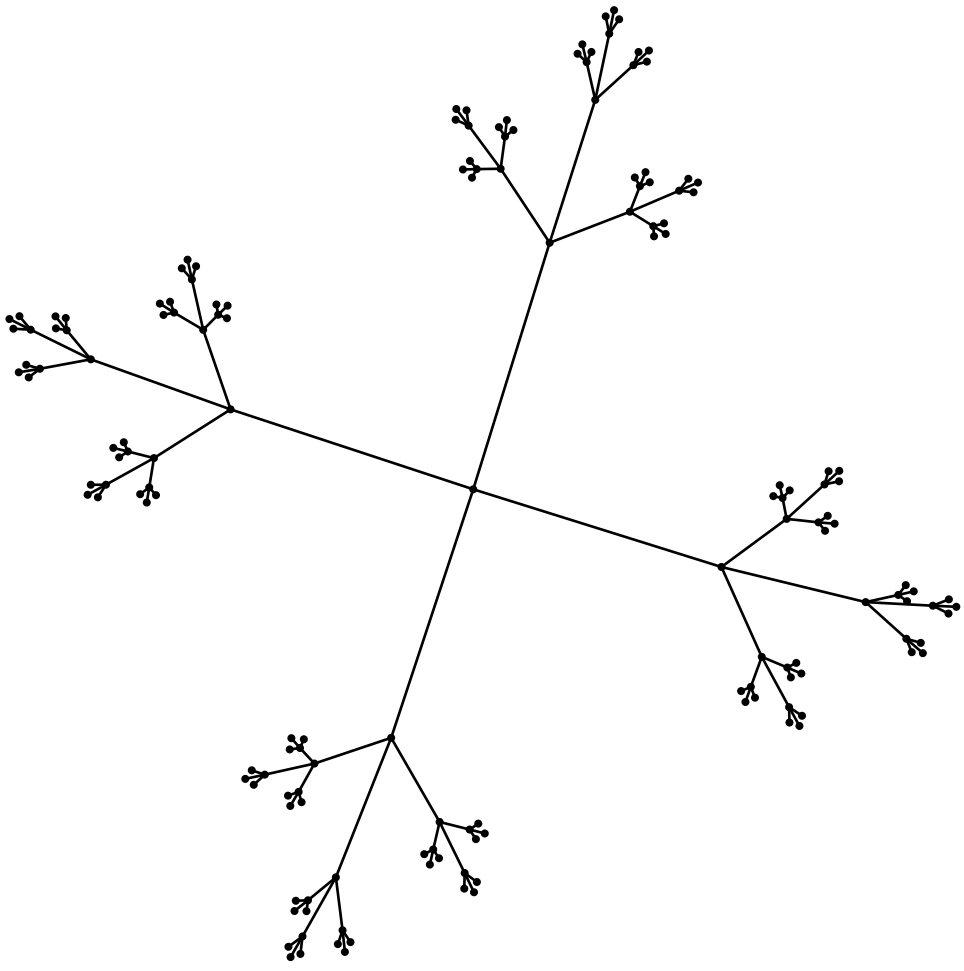
[1] To avoid confusion, the term *vertex* means a *polyhedral vertex* and we use always the term *node* for graphs.

a) *Study how computers can help in graph theory*. First, we surveyed what has been done since the pioneering system *Graph* [81, 82, 86] (see Chapter 2). Second, we used AutoGraphiX, and explained how it helped us in finding conjectures and ideas of proofs (see Chapters 4 and 5). Finally, we developed the system GraPHedron, which provides a formal framework to find optimal conjectures in a semi-automated way (see Chapter 6).

b) *Obtaining new graph theoretical results*. This dissertation contains many new theoretical results(see Chapters 3, 4, 5 and 7; Section 6.3.1 and Appendix A.1). The vast majority of these results were obtained with the help of AutoGraphiX and GraPHedron.

Part I

Computers and discovery in graph theory



Introduction

Bouvier and George [38] define a *conjecture* as "an a priori hypothesis on the exactness or falseness of a statement of which one doesn't know the proof". Actually, obtaining new results in science requires two steps : discovery and proof. One must first state what seems to be true, i.e., a conjecture. This can be done by observation, with the help of a model or computer software. Then, one must prove the conjecture by rigorous arguments.

This is also the case, of course, in graph theory. Since the early eighties, several computer systems have been used to makes conjectures in graph theory [159, 162].

It appears in [160] that each system explores specific types of conjectures and that some theorems, discovered without computer assistance, have forms that are not considered by these systems at this time. The forms of conjectures found with these systems and some unexplored forms of conjectures are studied in [160].

Chapter 1 recalls some basic notions about graph theory. Important definitions and notations are introduced in this chapter. New definitions, more specific, appear in subsequent chapters when needed.

Chapter 2 presents the main existing computers systems, used in graph theory, since the pioneering system *Graph* due to Cvetković et al. [76, 77, 78, 82, 83, 86]. For the main systems, examples of derived conjectures are given.

We regroup these systems in five families, based on the approaches they use. In Chapter 3, we present new theoretical results obtained by using one of these approaches : the enumeration. With an enumeration of particular graphs, we obtain some infinite families of integral graphs.

CHAPTER 1

Basics in graph theory

A model • Notations and definitions • Classes of graphs • Algebraic graph theory • Chemical graph theory • Extremal graph theory • Summary

This chapter recalls and defines important notions of graph theory, often together with simple examples. It is assumed in all the subsequent chapters that the reader is familiar with this material.

There is a wide divergence in graph theoretic terminology. We follow in general what is considered by many graph theorists to be of common use. The standard textbooks of Diestel [97] and Foulds [126] are the main references for such terminology.

1.1 A model

Königsberg Bridge Problem

Graph theory finds its roots in the famous and amusing Königsberg Bridge Problem. In the eighteenth century, two branches of the River Pregel meet in the Eastern Prussia city of Königsberg, to flow into the Baltic Sea. Seven bridges link the parts of the city which are separated by water (see Figure 1.1). Citizens ask themselves if it is possible to find a walk that crossed each bridge exactly once, coming back to the starting point.

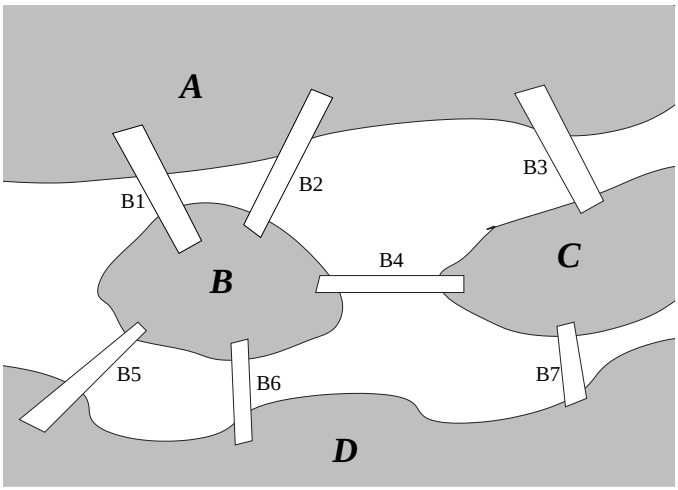


Figure 1.1: The seven bridges of Königsberg

This problem, apparently innocuous, raised the curiosity of the mathematician Léonard Euler. He proved in 1736 [106] that such a walk is unachievable. This problem can be represented with a picture composed of points and lines. Each point is a land area and a line is a bridge between two of them[1] (see Figure 1.2). The initial problem can be reformulated as follows. One has to find a sequence of points and lines which crosses each line exactly once, returning to its starting point. Euler [106] gave a condition for when it is possible on any such system of points and lines : each point has to be incident with an even number of lines. He proved that this condition is necessary and sufficient. It follows that the Königsberg Bridge Problem has no solution, which can be easily checked on Figure 1.2.

This mathematical model is today called a *graph*. The term *graph* should not be confused with the drawing of a mathematical function. Actually, this term, as the name of a system of

[1] Euler did not use such a system of points and lines in [106], but he used a letter to represent a land area and noted a walk from A to B by AB. However, Euler is commonly considered as the father of graph theory. For more information about history of graph theory, see [34], which include a translation of the publication [106] of Euler.

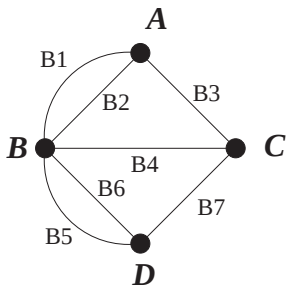


Figure 1.2: A graph representation of Königsberg

points and lines, comes from *graphic notation*, which was introduced in chemistry in 1878 by Sylvester [225]. Chemical structures are, of course, conveniently represented by a graph : atoms are points and atomic bonds are lines of the graph. Chemical applications in graph theory are numerous. The interested reader is referenced to the book [14]. We present in Chapter 4 several results obtained in chemical graph theory with the help of the computer system AutoGraphiX (see Section 2.5 and Part II).

Applications

The graph model has a lot of applications, not only in chemistry. See for instance the books [37, 126, 138]. We mention here some of them :

1. *Operational research.*

   Many problems of operational research use graphs in modeling and graph theoretic algorithms :

   - transportation networks problem;
   - assignment problem;
   - scheduling problem;
   - timetabling problem;
   - Traveling salesman problem;
   - shortest path problem;
   - location of centers;
   - communications networks;
   - ...

2. *Natural sciences* :

   - Physics (continuum statistical mechanics, percolation process, enumeration, ...);

- Biology (genetics, bioinformatics, ...);
- Chemistry (see above).

3. *Engineering* :
   - Electrical engineering (printed circuit design, electrical network analysis, ...);
   - Industrial engineering (production planning and control, facilities layout, ...);
   - Civil engineering (traffic networks design, pier construction planning, ...).
4. *Other applications* :
   - Computer science (network analysis, internet protocols, data structures, information retrieval, program analysis, ...);
   - Social sciences (voting, ranking, consistency of choices, urban studies, organizational structure, social hierarchy, ...);
   - Economics (input-output models, ...);
   - Geography (stream analysis, maps making, ...);
   - Mathematical puzzles and games;
   - ...

1.2 Notations and definitions

Definition 1.1. *A* graph *G, is an ordered pair (V,E), where V is a finite, nonempty set whose elements are called* nodes*, and where E is a set of unordered pairs of distinct nodes of V. Each element $\{v,w\} \in E$ is an* edge[2] *joining the nodes v and w.*

Remark 1.1. The term *vertex* is often used for a node. However, we avoid its use in this dissertation. Indeed, in Part III, a vertex means a *polyhedral vertex*.

Remark 1.2. As already said in the introduction of this chapter, there is variance in graph theoretic terminology, even for the term *graph* itself. A graph which is defined as in Definition 1.1 is often said to be a *simple, loop-free, undirected graph*.

If $e = \{v,w\} \in E$, then v and w are said to be *adjacent* to each other and *incident* with e. If two edges have a node in common, they are said to be *adjacent*.

An edge $\{v,w\}$ is often denoted simply by vw. The nodes v and w are called its *endpoints*.

A node (edge) of a graph is said to *cover* the edges (nodes) with which it is incident. The *neighborhood* $\mathcal{N}(v)$ of a node v is the set of nodes $\mathcal{N}(v) = \{w \in V : \{v,w\} \in E\} \cup \{v\}$.

[2] A synonym for edge is *link*.

Example 1.1. Let $G = (V, E)$ be defined as follows :

$$\begin{aligned} V &= \{v_1, v_2, v_3, v_4\} \\ E &= \{v_1v_2, v_1v_3, v_1v_4, v_2v_3, v_3v_4\} \end{aligned}$$

Then G can be represented as shown in Figure 1.3 and $\mathcal{N}(v_2) = \{v_1, v_2, v_3\}$.

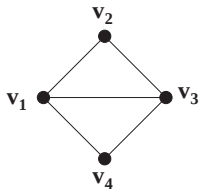


Figure 1.3: A graph with 4 nodes and 5 edges

The *number of nodes* of G, or its *order*, is denoted by $n(G) = |V|$. The *number of edges* of G, also called its *size*, is denoted by $m(G) = |E|$. The *degree* of a node v, denoted by $d_v(G)$, is the number of nodes w such that $\{v, w\} \in E$. We note $n_i(G)$ the *number of nodes of degree i* in G.

Example 1.2. Let G be the graph represented in Figure 1.3. Then its order is 4 and its size is 5. The degree of v_4 is equal to 2 and $n_3(G) = 2$.

Remark 1.3. When the context is clear, we often omit to mention G explicitly G in the notations. For instance, d_v is the degree of the node v, of the graph under consideration, and n_1 is the *number of pending nodes*, i.e., the number of nodes of degree 1.

Variants

There are several variants of related graph structures. For instance, some authors allow that V can be an empty set. In Definition 1.1, we ask that E contains only distinct edges. When this constraint is relaxed, the corresponding structure is called a *multigraph* (see for instance the multigraph depicted in Figure 1.2). In some problems, one wants to give an orientation to the edges, i.e., edges are *ordered* pairs of nodes. These types of graphs are called *directed graphs*. Another variant is the notion of *labeled graphs*. In this case a name is associated to each node and/or edge (see again Figure 1.2). One can also allow the presence of *loops* (edges vw such that $v = w$) in a graph, which is then called a *pseudograph*. There exist yet other types of graph-like structures (see [126]). Nevertheless, in this dissertation we consider only graphs which respect Definition 1.1.

Subgraphs

It is sometimes useful to consider only a part of a graph.

Definition 1.2. *Let $G = (V,E)$ be a graph.*

1. *A graph $H = (W,F)$ such that $W \subseteq V$ and $F \subseteq E$ is a* subgraph *of G. It is said a* proper subgraph *if $W \subset V$ or $F \subset E$.*
2. *Let W be a nonempty subset of V. Let $F = \{\{v,w\} \in E | v,w \in W\}$. The graph $H = (W,F)$ is an* induced subgraph *of G.*
3. *Let F be a subset of E. The graph $H = (V,F)$ is a* spanning subgraph *of G.*

Example 1.3. In Figure 1.4, the subgraphs H_1 and H_2 are both proper subgraphs of the graph G depicted in Figure 1.3. H_1 is an induced subgraph of G and H_2 is a spanning subgraph of G.

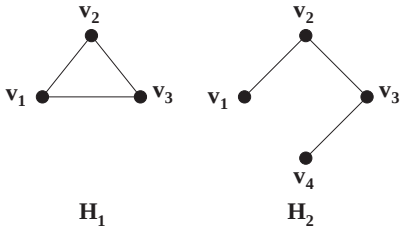


Figure 1.4: Two subgraphs of the graph in Figure 1.3

Path, distance and connectivity

Definition 1.3. *A* path *of length l in a graph G is a sequence of distinct nodes and edges of G of the form :*

$$(v_1, \{v_1,v_2\}, v_2, \{v_2,v_3\}, v_3, \ldots, v_l\{v_l,v_{l+1}\}, v_{l+1}).$$

The path is said to join *v_1 and v_{l+1}.*

A path is often noted

$$(v_1, v_2, v_3, \ldots, v_{l+1})$$

when there is no ambiguity. A graph which is solely a path with n nodes is denoted by P_n.

A closed path, i.e., a path $(v_1, v_2, v_3, \ldots, v_k)$ (where $k \geq 3$) with an additional edge $\{v_1, v_k\}$, is a *cycle*. A graph which is solely a cycle with n nodes is denoted by C_n.

The *distance* $d_{v,w}(G)$ between two nodes v and w is the length of a shortest path[3] between v and w.

One of the most important notions of graph theory is the *connectivity*.

Definition 1.4. *A graph G is* connected *if there exists a path between every pair of nodes in G.*

Example 1.4. The graphs H_1 and H_2 in Figure 1.4 are connected and $d_{v_1,v_3}(H_1) = 1$.

[3] In term of the number of edges.

Isomorphism

Another important notion is the *isomorphism* which expresses the fact that the labeling of the nodes does not make any difference between two graphs with the same node and edge structure.

Definition 1.5. *Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are said to be* isomorphic *(written $G_1 \cong G_2$) if there exist a one-to-one mapping*

$$f : V_1 \to V_2, \text{ such that } vw \in E_1 \iff f(v)f(w) \in E_2.$$

Example 1.5. The graph H_2 in Figure 1.4 is isomorphic to P_4.

Definition 1.6. *Two graphs are* homeomorphic *if they can be made isomorphic by inserting new nodes of degree 2 into edges.*

Example 1.6. The two graphs presented in Figure 1.5 are homemorphic.

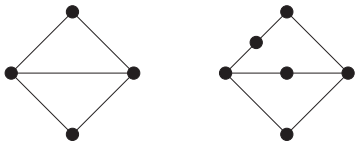


Figure 1.5: Two homeomorphic graphs

Invariants

Two graphs that are isomorphic share several properties which can be expressed in terms of *invariants*.

Definition 1.7. *A* graph invariant *$i(G)$ is a numerical value associated with a graph G and preserved by isomorphism.*

A graph invariant is often simply termed an *invariant*.

Example 1.7. The order n and the size m of a graph are invariants.

Remark 1.4. Invariants can have integer, boolean, rational or irrational values.

Graphs are used in many different applications (see Section 1.1). Therefore, there exist many invariants describing specific characteristics of graphs. Some of them are based on

- the notion of distance, as the *diameter* $D(G)$ which is the maximum distance between any pair of nodes of G, i.e., $D(G) = \max_{v,w \in V} d_{v,w}(G)$; the *average distance* $\bar{l}(G)$ which is the average of the distances between distinct nodes of G;

- the degrees, e.g., the *maximum degree* $\Delta(G) = \max_{v \in V} d_v$; the *minimum degree* $\delta(G) = \min_{v \in V} d_v$;
- the size of a particular set as the *stability number*[4] $\alpha(G)$ which is the maximum cardinality of an *independent set* of G, i.e. a set of non-adjacent nodes; the *chromatic number* $\chi(G)$, which is the minimum number of colors needed to assign a color to each node of G such that no pair of adjacent nodes have the same color;
- the membership in a given class of graphs (see Section 1.3). These invariants have thus often boolean values;
- algebraic properties (see Section 1.4);
- chemical properties (see Section 1.5);
- ...

Another example of invariant based on the degrees is the *irregularity* introduced by Albertson [1]. This invariant is intensively studied in the next chapters (see Section 4.2, Chapter 5 and Section 7.2).

Definition 1.8. *The* irregularity $A(G)$ *of a graph G is defined by :*

$$A(G) = \sum_{\{v,w\} \in E} |d_v - d_w|.$$

Example 1.8. Let G be the graph described in Figure 1.3, then

$$n(G) = 4,\ m(G) = 5,\ D(G) = 2,\ \bar{l}(G) = \frac{7}{6},\ \Delta(G) = 3,$$

$$\delta(G) = 2,\ \alpha(G) = 2,\ \chi(G) = 3 \text{ and } A(G) = 4.$$

1.3 Classes of graphs

In the previous section we recall specific classes of graphs, like P_n and C_n. Other classical classes of graphs are used in this dissertation. A graph in which every pair of its n nodes are connected by an edge is a *complete graph* and is denoted by K_n. A *clique* of a graph is its maximal complete subgraph and the *clique number* $\omega(G)$ of a graph G is the maximum number of pairwise adjacent nodes.

A *bipartite graph* $B_{p,q}$ is composed of two independent sets of nodes (partitions), with p and q nodes respectively, and some edges joining pairs of nodes $\{v, w\}$ such that v and w are not in the

[4] A synonym for the stability number is the *independence number*.

same partition. $K_{p,q}$ is a *complete bipartite graph* if it contains such edges for all pairs. The *star* of order n, denoted by S_n, is the graph $K_{1,n-1}$.

The *complement* $\overline{G}$ of a graph $G = (V,E)$, is the graph with the same node set V such that two nodes are adjacent in $\overline{G}$ if and only if these nodes are not adjacent in G. The *empty graph* of order n is the graph $\overline{K}_n$.

A graph G is *cyclic* if it contains at least one subgraph which is a cycle. It is *acyclic* otherwise. If G contains exactly one cycle (resp. two cycles), it is termed *unicyclic* (resp. *bicyclic*).

Example 1.9. Let H_1 and H_2 be the graphs shown in Figure 1.4, then,

$$H_1 \cong K_3,\ \omega(H_1) = 3,\ H_2 \cong \overline{H}_2 \text{ and } \omega(H_2) = 2.$$

Moreover H_1 is unicyclic and H_2 is an acyclic bipartite graph.

A *tree* is an acyclic connected graph. The following theorem is proven in [28].

Theorem 1.1. [28] *Let $G = (V,E)$ be a graph of order $n \geq 2$. The following properties are equivalent to characterize a tree ;*

1. *G is connected and acyclic;*
2. *G is acyclic and has $n-1$ edges;*
3. *G is connected and has $n-1$ edges;*
4. *G is acyclic and adding an edge creates exactly one cycle;*
5. *G is connected and removing an edge breaks connectivity;*
6. *each pair of nodes is joined by one and only one path.*

Example 1.10. The graph represented on the cover of this dissertation is a tree.

A graph is *k-regular* if all its nodes have a degree equal to k. A 3-regular graph is a *cubic graph*.

Some classes of graphs are defined by means of forbidden subgraphs. For instance, a *triangle-free* graph is a graph which does not contain any *triangle*, i.e., the graph K_3.

Example 1.11. In Figure 1.4, H_1 is 2-regular and H_2 is triangle-free.

If the class of a graph is unspecified, it is said to be a *general* graph.

Constructions

One can also define a graph or a class of graphs using some constructions.

Definition 1.9. *The* union $G_1 \cup G_2$ *of graphs* $G_1 = (V_1, E_1)$ *and* $G_2 = (V_2, E_2)$ *with* $V_1 \cap V_2 = \emptyset$ *is the graph* $G = (V, E)$ *for which* $V = V_1 \cup V_2$ *and* $E = E_1 \cup E_2$.

The notation kG is short for

$$kG = G_1 \cup G_2 \cup \cdots \cup G_k,$$

where $G_i \cong G$, for all $i = 1, 2, \ldots, k$.

Example 1.12. A union of (at least two) trees is a (non-connected) acyclic graph and is also called a *forest*.

Definition 1.10. *The* complete product $G_1 \nabla G_2$ *of graphs* $G_1 = (V_1, E_1)$ *and* $G_2 = (V_2, E_2)$ *with* $V_1 \cap V_2 = \emptyset$ *is the graph obtained from* $G_1 \cup G_2$ *by joining every node of* G_1 *with every node of* G_2.

Example 1.13. Let G be the graph described in Figure 1.3, then

$$G \cong K_2 \nabla \overline{K}_2$$

Definition 1.11. *The* sum $G_1 + G_2$ *of graphs* $G_1 = (V_1, E_1)$ *and* $G_2 = (V_2, E_2)$ *is the graph with the node set* $V_1 \times V_2$ *in which two nodes* (u_1, u_2) *and* (v_1, v_2) *are adjacent if and only if* ($u_1 = v_1$ *and* $(u_2, v_2) \in E_2$) *or* ($u_2 = v_2$ *and* $(u_1, v_1) \in E_1$).

Example 1.14. The graph $K_2 + K_3$ is represented in Figure 1.6.

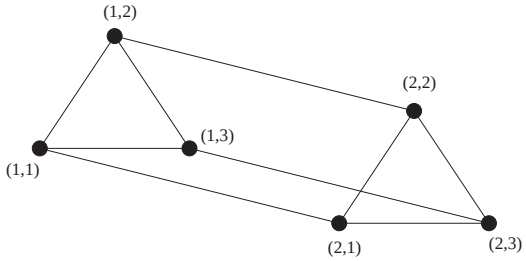


Figure 1.6: The graph $K_2 + K_3$

1.4 Algebraic graph theory

Definition 1.12. *The* adjacency matrix $\mathcal{A}(G) = (a_{ij})$ *of a graph* G *of order* n *is a* $n \times n$ *symmetric matrix with coefficients* $a_{ij} \in \mathbb{N}$ *defined by*

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

Several invariants are based on the algebraic properties of the adjacency matrix.

The polynomial $P_G(\lambda) = det(\lambda I - \mathcal{A})$ is called the *characteristic polynomial* of G. The *spectrum* of G is the set of solutions to $P(\lambda) = 0$, called *eigenvalues*, and denoted $S_p(G) = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The first eigenvalue λ_1 is called the *index* or *spectral radius*.

Remark 1.5. It is known that if G is an undirected graph, then eigenvalues of G are *real* and the maximum eigenvalue is non-negative [80].

Definition 1.13. *A graph is called* integral *if all its eigenvalues are integer values.*

Another important $n \times n$ symmetric matrix is the *Laplacian matrix* $\mathcal{L}(G) = (l_{ij})$ with coefficients $l_{ij} \in \mathbb{N}$ defined by

$$l_{ij} = \begin{cases} d_i & \text{if } i = j, \\ -1 & \text{if } i \neq j \text{ and } \{i, j\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Remark 1.6. The eigenvalues of $\mathcal{L}(G)$ are always real [80].

The *Laplacian eigenvalues* $S_p^{\mathcal{L}}(G)$ of G are the eigenvalues of $\mathcal{L}(G)$ and are denoted by $(\lambda_1^{\mathcal{L}}, \lambda_2^{\mathcal{L}}, \ldots, \lambda_n^{\mathcal{L}})$ with $\lambda_1^{\mathcal{L}} \geq \lambda_2^{\mathcal{L}} \geq \cdots \geq \lambda_n^{\mathcal{L}}$. The smallest Laplacian eigenvalue $\lambda_n^{\mathcal{L}}$ is always equal to 0. The second smallest eigenvalue $\lambda_{n-1}^{\mathcal{L}}$ is called the *algebraic connectivity* and $\lambda_1^{\mathcal{L}}$ is the *largest Laplacian eigenvalue*.

Example 1.15. Let G be the graph in Figure 1.3. Then,

$$\mathcal{A}(G) = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}, \quad \mathcal{L}(G) = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix},$$

$$S_p(G) = (\frac{1+\sqrt{17}}{2}, 0, -1, \frac{1-\sqrt{17}}{2}) \text{ and } S_p^{\mathcal{L}}(G) = (4, 4, 2, 0).$$

So the index of G is $\frac{1+\sqrt{17}}{2} \simeq 2.5615528$ and its algebraic connectivity is equal to 2.

The interested reader is referenced to [80], a standard book of algebraic graph theory. In Chapter 2, most examples of conjectures come from algebraic graph theory.

1.5 Chemical graph theory

Another important part of graph theory is the so-called chemical graph theory. The main goal of chemical graph theory is to evaluate chemical invariants as well as correlations between chemical

activities or properties and those invariants or functions of them in *Quantitative Structure Activity Relationship* (*QSAR*) and *Quantitative Structure Property Relationship* (*QSPR*) studies [66, 92, 94, 99, 125, 156, 179, 181, 182, 183, 190, 227, 229, 230].

Graphs representing molecular structures often have a maximum degree Δ which is bounded. In particular, we recall

Definition 1.14. *A* chemical graph *is a graph with a maximum degree* $\Delta \leq 4$.

It comes from the fact that the carbon atom has a valency of 4. For instance, the *fullerenes* are molecules composed entirely of carbon, which take the form of a hollow sphere, ellipsoid, or tube.

The following invariants are very common in chemical graph theory.

Definition 1.15. *The* Randić index *[210] or* connectivity index *of a graph* $G = (V,E)$ *is defined as*

$$Ra(G) = \sum_{\{v,w\} \in E} \frac{1}{\sqrt{d_v d_w}}. \tag{1.1}$$

Definition 1.16. *The* energy *of a graph [141, 144] is defined as*

$$E = \sum_{i=1}^{n} |\lambda_i|.$$

Definition 1.17. *The* HOMO-LUMO gap *is defined as the difference between the smallest positive and the largest negative eigenvalue of the adjacency matrix*[5]

Example 1.16. Let G be the graph in Figure 1.3. From its spectrum, computed in the previous example, we get that its HOMO-LUMO gap is $\frac{3+\sqrt{17}}{2}$ and its energy is $\sqrt{17}+1$. Moreover, $Ra(G) = \frac{1}{3} + \frac{4}{\sqrt{6}}$.

Chapter 4 presents several results obtained in chemical graph theory, with the help of AutoGraphiX.

1.6 Extremal graph theory

Extremal graph theory studies the graphs which are extremal among graphs with a certain property. There are various meanings for the word extremal: with the largest stability number, the largest minimum degree, the smallest irregularity, etc. The term *extremal graph theory* can thus

[5] One should take care, in this definition, that 0 is neither positive nor negative. *HOMO* means Highest Occupied Molecular Orbital and *LUMO* means Lowest Unoccupied Molecular Orbital.

encompass a large part of graph theory. Actually, the vast majority of the results presented in this dissertation can be considered as belonging to extremal graph theory. Indeed, most of the conjectures or theorems presented in the next chapters are bounds on invariants, i.e., inequalities among invariants.

Example 1.17. Connected graphs with n nodes and with *extremal* diameter are easily characterized :

$$D(K_n) \leq D(G) \leq D(P_n).$$

These two inequalities mean that the complete graph K_n has the smallest diameter ($= 1$), and the path P_n the largest one ($= n-1$), among connected graphs with n nodes.

1.7 Summary

In this chapter, we presented the *graph* model and mentioned some of its numerous applications. Basic notations and definitions are recalled or introduced about graph theory, classes of graphs, algebraic graph theory, chemical graph theory and extremal graph theory.

We presented several invariants but other ones are introduced when needed in the subsequent chapters.

CHAPTER 2

Existing computer systems

Enumeration • Interactive computing • Invariant manipulation • Generation and selection • Heuristic optimization • Summary

Quite a few systems for computer-aided graph theory have been developed over the last 25 years. The term *computer system* is used here in a relatively large sense. It can be a software system, a programming language, or a set of programs. Some systems presented here were not initially written for graph theoretical conjecture-making. However, we mention their use in this field.

Hansen [159] divides such systems into two classes: automated systems which provide conjectures in a fully automated way (i.e., without human intervention except for the problem statement), and computer-assisted systems. Derived conjectures are then said to be obtained *by* an automated system or *with* a computer-assisted one. Notice that in practice an automated system can often also be used interactively and leads to both automated and/or computer-assisted conjectures.

The existing computer systems are based on different principles, which can be grouped as follows [162] :

(i) enumeration;

(ii) interactive computing;

(iii) invariant manipulation;

(iv) generation and selection;

(v) heuristic optimization.

Representative systems of each family are discussed in the next five sections.

2.1 Enumeration

Enumeration refers to two distinct operations performed in mathematics, when examining a set of objects or structures :

- one can find how many there are, i.e., *count* them. Methods to do so are well developed in combinatorial mathematics and graph theory (see, e.g., the books [171, 204]).
- one may consider each of them in turn, i.e., *list* these objects or structures (this is sometimes referred to as *constructive enumeration*).

Constructive enumeration

The computer systems which use enumeration mostly address constructive enumeration. It is the case of the systems mentioned below.

Basic problems of enumeration are to avoid duplication and to efficiently exploit properties of the class of graphs under study to curtail the search. Often graphs are generated by adding one node at a time, and some adjacent edges. To avoid duplication, graphs are encoded and a unique *father* is assigned to any *son*. This principle of *orderly generation* is proposed in [113, 114, 211, 212]. Variants and extensions, including the *canonical path method* are discussed in [192, 193]. Moreover, symmetry can be exploited both to avoid duplication and to accelerate the search. A recent survey of isomorphism rejection methods is [45]. Several systems do this, e.g., *geng* [192, 193]. This system and others such as *CoCo* [115, 116] are used in the package *GAP* (Groups, Algorithms and Programming) [131]. When applied to problems on groups and graphs, the program *GRAPE*, which is a part of *GAP*, led to several results [68, 69, 70, 71, 72, 205, 221, 222].

Some systems for enumerating graphs are specialized, e.g., *MOLGEN* [139] which is designed for molecular graphs, *Fullgen* for fullerenes [46] and *minibaum* for cubic graphs [44]. *CaGe* [47] generates graphs of different types often related to interesting chemical molecules.

Examples

The enumeration of families of graphs defined by given properties often leads to conjectures about them, or to refutations of such conjectures.

Example 2.1. In [79] spectra of all graphs with up to 7 nodes are given. This list refutes 5 conjectures of *Graffiti* ([107, 109], see Section 2.4) as noticed in [119].

Example 2.2. In Part III, we present a new system, called GraPHedron, which uses a polyhedral approach to suggest conjectures. The first step of its outline is a generation of graphs using *geng* (see Section 6.2).

Example 2.3. In Chapter 3, we present results obtained by enumeration of particular graphs (written with the programming language *Matlab*). With this approach, we obtained some infinite families of graphs, for which all eigenvalues are integers.

2.2 Interactive computing

Numerous conjectures of graph theory are obtained by drawing small graphs on paper or blackboard, making hand or pocket calculator computations of invariants under study, reasoning about their values, then modifying these graphs and computing the consequences. Such a process can be aided by the computer, exploiting its abilities to make very quick computations and represent graphs in a clear way.

The system *Graph*

A pioneering system in graph theoretical interactive computing is *Graph* [81, 82, 86, 87] developed during the period 1980 – 1984 in Belgrade. This system comprises three main components : *Algor* which implements graph algorithms for computing a series of invariants, as well as *Bibli* and *Theor*, a bibliographic and a theorem-proving component respectively. These last two are not discussed here.

Graph also displays the graph currently under study and allows interactive modifications on screen by the addition or deletion of edges and nodes (this, of course allows any transformation). While *Graph* is not an automated system (it does not provide conjectures or proofs in an entirely automated way), it has proved to be very successful in suggesting conjectures through analysis of examples, and also in helping to complete proofs by checking particular cases. The survey papers [87] mention 92 papers with results obtained up to 2001 with the help of *Graph*.

Examples

We next give a few examples of results obtained with the aid of the system *Graph*. Further theorems of algebraic graph theory obtained in this way are listed in Table 2.1. There, an *outerplanar graph* is a graph that can be embedded in the plane such that all nodes lie on the unbounded outside

Table 2.1: Some theorems obtained with the help of the *Graph* system

Formula	Ref.
If G is a tree (with $n \geq 3$), then $$\lambda_1(P_n) \leq \lambda_1(G) \leq \lambda_1(S_n)$$	[189]
If G is a maximal outerplanar graph (with $n \geq 4$), then $$\lambda_1(P_n^2) \leq \lambda_1(G) \leq \lambda_1(K_1 \nabla P_{n-1})$$	[214]
If G is a connected graph and if G' is obtained from G by splitting a node, then $$\lambda_1(G') \leq \lambda_1(G)$$	[219]

face. Moreover, P_n denotes the path on n nodes, P_n^2 its square, i.e., the graph obtained by joining by an edge pairs of nodes of P_n at distance 2 :

$$V(P_n^2) = V(P),$$

$$\{v,w\} \in E(P_n^2) \text{ iff } d_{v,w}(P_n) = 2, \forall v,w \in V(P_n).$$

Example 2.4. Unicyclic graphs with extremal index are characterized by Simić [218] in the following theorem.

Theorem 2.1. [218] *Let G denote a unicyclic graph; then*

$$\lambda_1(C_n) \leq \lambda_1(G) \leq \lambda_1(S_n + e)$$

with equality if and only if G is the n-cycle C_n (lower bound) or the star S_n with an additional edge e between two pending nodes (upper bound).

Example 2.5. Combining graph theoretical results and a computer search with *Graph*, all connected, non-regular, non-bipartite integral graphs with maximum degree four were determined in [207, 208] (see Theorem 2.2 below). This method was used also to find the connected non-regular bipartite integral graphs with $\Delta \leq 4$ [20, 21] as well as one class of connected 4-regular integral graphs [224]. Note that the search was not carried out by brute force but as a man-machine interaction, many parts of the search space being discarded for graph theoretical reasons or by computational results.

Theorem 2.2. [207, 208] *There are exactly 13 integral graphs which are connected non-bipartite and non-regular with maximum degree 4.*

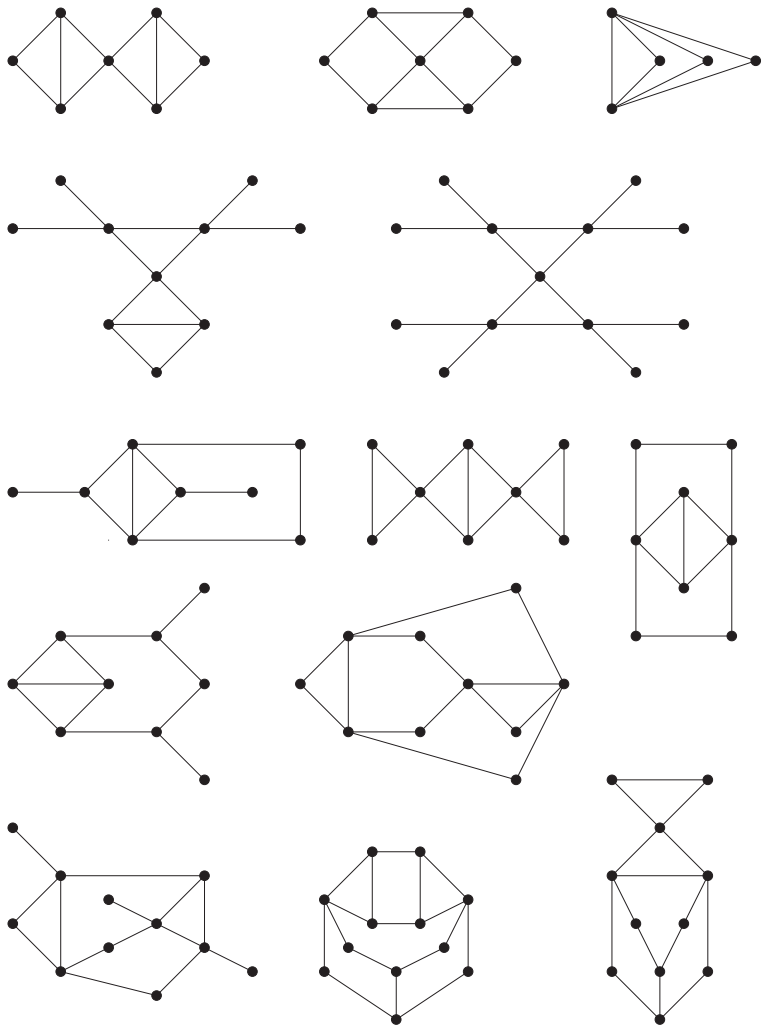


Figure 2.1: The 13 integral graphs of Theorem 2.2 [207, 208]

These graphs are shown in Figure 2.1.

Other interactive computing systems

In the last decade libraries of graph algorithms as well as systems for graph visualization and/or editing have proliferated. We mention a few. *GraphBase* [184], *Leda* [194] and *Vega* [231] comprise efficient implementations of many graph algorithms. *GraphEd* [174] and its successor *Graphlet* [175], *VCG* [215], *CABRI-Graph* [57], *Link* [30, 31, 32, 33], *GGCL* (Generic Graph Component Library) [186] and other systems possess in addition an editor. *EDGE* [200], *Da Vinci* [129], *Grappa* [24] and other systems focus on graph visualization and editing. Note that graph drawing is a well developed research area, see, e.g., the surveys [95, 100] and the book [96].

Recent systems are often the outcome of a merging of several previous ones. This is for instance the case with *Link* which built upon the experience of the authors of *Combinatorica*

[220], *NETPAD* [197], *SetPlayer* [29] and *GraphLab* [216]. The *Link* system led to a conjecture which helped to find several new infinite families of *antichains of tournaments*[1] [137, 185].

2.3 Invariant manipulation

Graph theory contains a large number of *relations* between graph invariants. A few of them are equalities, the others are inequalities, often nonlinear in one or more parameters. They may involve *conditions* which are themselves relations or properties defining classes of graphs, e.g., bipartite, tree, *planar*[2], etc. Logical variables associated with these classes may also be considered as invariants. Generalized relations are then obtained and are of the forms:

(i) "If relation R_1 holds then relation R_2 holds"
e.g., "If $\alpha = 2$ and $\chi \geq 4$ then $n \geq 11$" [64], where α denotes the stability number and χ the chromatic number of G.

(ii) "If condition c holds then relation R holds"
e.g., "If G is planar then $a \leq 3$" [59] where a denotes the *point arboricity* of G, i.e., the minimum number of forests needed to cover G.

(iii) "If condition c_1 holds then condition c_2 holds"
e.g., "If G is a tree then it is bipartite" (obvious).

(iv) "If relation R holds then condition c holds"
e.g., "If $n \geq 6\delta$ and $m > \frac{1}{2}(n-\delta)(n-\delta-1)+\delta^2$ then G is hamiltonian" [48] where δ denotes the minimum degree of G and a *hamiltonian graph* contains a path going once and only once through every node.

The system *Ingrid*

In order to build the graph invariant manipulator system *Ingrid*, 458 relations between graph invariants were gathered in [41, 42]. They involve 37 graph invariants, 27 of which are integer-valued, 1 real-valued and 9 boolean. A representative subset of these relations, involving λ_1, is given in Table 2.2. Here χ_1 denotes the *edge chromatic number* or smallest number of colors needed to color edges so that no two incident edges have the same color; g denotes the *girth* or length of the smallest cycle of the graph considered.

Ingrid [40, 43] was designed to assist researchers in obtaining precise information, in the form of intervals, on invariant values for incompletely specified graphs or classes of graphs. To this

[1] A *tournament* is a complete directed graph. Tournament *embedding* is an order relation on the class of tournaments. An *antichain* is a set of tournaments that are pairwise incomparable in this ordering.

[2] Recall that a graph is planar if it can be drawn in a plane without graph edges crossing.

Table 2.2: Some relations involving λ_1 in the *Ingrid* system

Formula	Ref.	Formula	Ref.
$\lambda_1 \geq \frac{2m}{n}$	[25]	$\lambda_1 \geq \sqrt{\Delta}$	[188]
$\lambda_1 \leq \Delta$	[25]	$\chi \geq \frac{n}{n-\lambda_1}$	[75]
$a \leq 1 + \lfloor \frac{\lambda_1}{2} \rfloor$	[187]	$\alpha \geq \frac{n}{n-\lambda_1} - \frac{1}{3}$	[101]
$\chi \geq \frac{2m}{(2m-\lambda_1^2)}$	[101]	if $\lambda_1 \leq \frac{\Delta}{2}$ then $\chi_1 = \Delta$	[121]
if $\lambda_1 \geq \sqrt{m}$ then $g = 3$	[80]	if G is connected, then	
$\chi \leq \lambda_1 + 1$	[25]	$\lambda_1 \geq 2\cos[\pi/(n+1)]$	[25]
$\lambda_1 \geq \delta$	[25]		

end, some parameters are given specified values or intervals containing their values. Then rules deduced from the relations are applied to tighten intervals for all invariants. This is done in a systematic way, until stability is attained. A tracing function allows the listing of those relations which have led to the lower or upper bound of the final interval for an invariant.

Moreover, conjectures may be considered temporarily as theorems (proved relations), added to the system and the consequences tested. If the interval of feasible values for some invariant becomes empty, a contradiction has been found and the conjecture refuted.

Examples

Ingrid can contribute to graph theory in several ways :

(i) *Detecting existence of relations between invariants and sets of relations leading to them.*

Consider a pair of invariants for which a relation is sought; vary one of them and check if the feasible interval of the other given by *Ingrid* varies also. If it is the case, a relation exists. To find it, consider which relations have been used, with the tracing function. Then derive the relation by algebraic manipulation. This last step is done by hand, but could be automated, for instance with *Mathematica*^{TM} [232].

Example 2.6. In [43] a relation is sought between the index λ_1 and the *node clique cover*

number θ_0, i.e., the smallest number of cliques which cover all nodes. While both parameters had been much studied it did not appear that any relation between them was yet published. Keeping the order n fixed and varying λ_1, *Ingrid* detected a change in the upper bound of the interval for θ_0. This was due to the use of the four relations

$$\Delta \leq \lambda_1^2,$$

$$\varphi \geq \frac{n}{(\Delta+1)},$$

$$\theta_0 \leq \alpha_1,$$

and

$$\alpha_1 \leq n - \varphi$$

where Δ, α_1 and φ denote maximum degree, *edge covering number* (minimum number of edges needed to cover the nodes) and *matching number* (maximum number of independent edges) respectively. It is then easy to derive the theorem

$$\theta_0 \leq n\left\lfloor \frac{\lambda_1^2}{(1+\lambda_1^2)} \right\rfloor$$

which is useful for small λ_1.

(ii) *Refuting conjectures*

Example 2.7. It was asked in [73] whether there exist planar triangle-free graphs with exactly 3α nodes. Conjecturing that this was the case, using the temporary theorem feature of *Ingrid* as explained above, led to a negative answer.

(iii) *Exploring dominance between relations*

An inequality between graph invariants may be implied by one or several other inequalities. When that is the case, there is no need to add it to the system. To check this, the bound it gives can be compared with that given by *Ingrid* for various feasible values of the invariants involved.

Example 2.8. The bound

$$\lambda_1 \leq -1 + \sqrt{1+8m}$$

was proposed in [223]. Varying m in *Ingrid* and observing the upper bound on λ_1 together with the relations used, it was found that the pair of relations

$$\chi \leq \left\lfloor 1 + \frac{\sqrt{1+8m}}{2} \right\rfloor$$

and

$$\lambda_1 \leq \sqrt{2m(\chi-1)/\chi}$$

provided bounds which were usually better and never worse. This can then be proven analytically.

The question of which relations are undominated is considered in [169]. An inequality $i_1 \geq (\leq) f(i_2, i_3, \ldots, i_n)$ between an invariant i_1 and one or several others $i_2, \ldots, i_n$ is sharp if for all values of the independent invariants compatible with the existence of a graph, there is a graph for which equality holds. A *complete set of sharp inequalities* between invariants $i_1, i_2, \ldots, i_n$ is composed of $2n$ sharp lower and upper bounds for each invariant i_j as a function of the others. Such a complete set for the order, size and stability number of (non-necessarily connected) graphs has been obtained in [169].

Ingrid can also be used to help to solve practical problems in network design and for pedagogical purposes [43]. Discovery-based pedagogy in graph theory is also discussed in [57], [60] and [89].

2.4 Generation and selection

The system *Graffiti*

The *Graffiti* system [107, 108, 110, 111, 112, 91] is designed to automate generation of conjectures in graph theory (as well as in geometry, number theory and mathematical chemistry). It contains a database of relations and a database of examples, which are graphs that refute some conjecture. *Grafitti* proceeds in two steps :

(i) Graph invariants $i_1, i_2, \ldots i_p$ are selected and a large number of *a priori* relations (or conjectures) between them are generated. They have simple forms, e.g. :

$$i_k \leq i_l \text{ or } i_k \leq i_l + i_m \text{ or } i_k + i_l \leq i_m + i_n;$$

one invariant may also be replaced by a constant, usually 1; sometimes ratios or products of invariants are also considered. In fact, as an algebraic expression involving one or several graph invariants is itself a graph invariant, any such relation can be used.

Classes of graphs to be considered, e.g., general, triangle-free, bipartite, tree and so on are also specified.

(ii) Selection is performed among the relations (or conjectures) obtained in (i). They may be discarded or provisionally set aside.

The former happens

(a) when a new relation does not appear to be informative, i.e., does not provide a sharper value for some invariant than all other relations in the database for at least one of the stored examples, or

(b) when a new relation is shown to be false for at least one of these graphs.

The latter happens

(c) when a new relation is implied by an existing conjecture, or

(d) when a new relation for a given class of graphs (e.g., trees) is not refuted by any example of a larger class (e.g., bipartite graphs), or

(e) when invariants in a relation are too close to one another (i.e., $i_k \leq i_l = i_k + 1$).

Note that the test for informativeness in the selection procedure removes most but not all such relations.

To speed up the procedure, both databases are kept to a moderate size. When a counter-example is found it is added to the database, the refuted relations removed and possibly others, which become informative, added. When a new relation is added, those which are no longer informative are set aside.

False conjectures play an important role as the systematic addition to the database of graphs refuting them leads to increasingly strong conjectures. The aim is to find this strongest conjecture for which no counter-example is known. Selected conjectures are proposed to the mathematical community in the large *Written on the Wall* [107] file. Their status (proved, refuted or open) is specified and regularly updated. Indications of partial proofs and generalizations of the conjectures are also given, with references.

Graffiti has attracted the attention of more than 80 graph theorists and has led to the publication of several dozen papers [90], some well-known ones being [63, 117, 118, 119]. Initially, refutation was easy; in [39] 200 conjectures of *Graffiti* were tested on all graphs with up to 10 nodes and over 40 of them were refuted. The remaining early open conjectures seem more likely to hold, and some of them appear to be hard to prove. Relations of *Graffiti* within algebraic graph theory were studied in depth in [119].

The paper [91] gives an insight in the history of the development of *Graffiti*.

Examples

We next give a few examples. Further relations and their status are presented in Table 2.3. There E^* denotes the vector whose i^{th} component is the number of nodes at even distance from the i^{th} node and *Ra* is the Randic index [210].

Example 2.9. [Conjecture WOW 747, open] If b is the order of the largest bipartite subgraph of a connected graph G, then the average distance between distinct nodes of G is not more than $\frac{b}{2}$.

Fajtlowicz observes that if true this conjecture would generalize the previous **Conjecture WOW 2** that the average distance is not more than the stability number proven in [63].

Table 2.3: Some conjectures of *Graffiti*

Num.	Formula	Status
WOW 19	$-\lambda_n \leq Ra$	Proved in [119]
WOW 43	If G is regular, $-\lambda_n \leq \varphi$	Proved in [119]
WOW 44	If G is regular, $\lambda_2 \leq \alpha$	Refuted by N. Alon
WOW 45	If G is regular, $\lambda_2 \leq \varphi$	Proved by N. Alon
WOW 116	If G is triangle-free, $\lambda_1 \leq Ra$	Proved in [119]
WOW 195	$\lambda_n \leq \max(E^*)$	Open

Example 2.10. [Conjecture WOW 776, refuted] Let p be the sum of positive eigenvalues of G. If G is cubic then the stability number of G is greater than or equal to $-1+\frac{p}{2}$.

An 18-node counter-example was found in [206] using the AutoGraphiX system (see Section 2.5 and Part II). For that graph $\alpha = 6$ and

$$-1+\frac{p}{2} > 6.04.$$

Example 2.11. [Conjecture WOW 256, proved] Let the *dual degree* of a node be the average of the degrees of its adjacent nodes. Then the maximum eigenvalue of the adjacency matrix of a graph G is not more than its maximum dual degree.

The same short and elegant proof for this result was found independently by a researcher in the U.S. and a group of three researchers in France, see [107], page 78. The result generalizes the well known property that the largest eigenvalue of G is not more than its maximum degree. The French group noticed that equality holds if and only if every node has the same dual degree.

2.5 Heuristic optimization

Conjectures in graph theory can be viewed as combinatorial optimization problems on an infinite family of graphs (of which only those of moderate order are explored). Indeed, given a relation $i_k \leq i_l$, one can minimize $i_l - i_k$ over all graphs, parameterizing for instance by graph order. As soon

as a graph such that $i_l - i_k < 0$ is found the conjecture is refuted. Conversely if extremal or near-extremal values of an invariant (which may be an expression involving several other invariants) are found for all small values of parameters such as order and size, this may lead, automatically or with the aid of the computer, to the discovery of new conjectures.

The system AutoGraphiX

This is the approach on which the AutoGraphiX system (AGX) [2, 3, 4, 5, 6, 7, 26, 50, 51, 52, 54, 55, 56, 88, 127, 146, 163, 164, 165, 168] is based. AutoGraphiX addresses the following problems :

(a) find a graph satisfying given constraints;

(b) find optimal or near-optimal values for an invariant subject to constraints;

(c) refute a conjecture;

(d) find (or suggest) a new conjecture (or sharpen an existing one);

(e) suggest a proof strategy.

AutoGraphiX uses extensively the *Variable Neighborhood Search* (VNS) metaheuristic (or framework-for-building heuristics) [167]. This metaheuristic exploits the relatively unexplored idea of systematic change of neighborhood within a local search. VNS starts with a given randomly generated initial solution (or graph) x_0, then applies a descent routine (when minimizing) until a local optimum x is reached. Then a set of nested neighborhoods centered around x are considered and a point x' is randomly generated from the first neighborhood. A descent is performed from x', leading to a local optimum x''. If $x'' = x$ or if the value of x'' is not better than that of x, the solution x'' is ignored and another solution x' is generated from the next neighborhood. Otherwise, as a better local optimum x'' than x has been found, the search is recentered there. When the last neighborhood has been considered one begins again with the first one until a stopping condition is met.

The descent routine may itself use several neighborhoods (of types of moves). AutoGraphiX uses ten simple graph transformations for that purpose : *addition* of an edge, *removal* of an edge, *rotation* of an edge (i.e., change of one of its endpoints), *move* of an edge (i.e., deletion followed by addition, but not in the same position), and similar more complex changes.

Nested neighborhoods of a graph are defined by the *Hamming distance* between edge-sets : the first one consists of all graphs obtained by deletion or addition of a single edge, then two, and so on.

The first version of AutoGraphiX has led, partly in conjunction with a program for enumerating cubic graphs [45], to the refutation of 9 conjectures of *Graffiti* [55] [206] and to the discovery of

Table 2.4: Some conjectures obtained with AutoGraphiX

Num.	Conjecture	Status	Ref.
Co. AGX 13	If G is a graph with n nodes, $m \leq \lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil$ edges and minimum energy, then (i) if there are positive integers a and b such that $ab = m$ and $a+b \leq n$, then G is a complete bipartite graph $K_{a,b}$ possibly with additional isolated nodes; (ii) otherwise G is a complete bipartite graph $K_{a',b'}$ with $a'b' \leq m$ and $a'+b' \leq n$, modified by the addition of $m-a'b'$ edges joining a node on the smallest side of $K_{a',b'}$ to other nodes on that side, possibly with additional isolated nodes.	Open	[50]
Th. AGX 12	Let $\mathcal{T}_{a,b}$ denotes the family of trees with a black and b white nodes, such that $a \geq b$ and no edge has two endnodes with the same color. Then, for a fixed number of nodes n and $T \in \mathcal{T}_{a,b}$, the minimal value of $\lambda_1(T)$ increases monotonically with $a-b$.	Proved	[88]
Th. AGX 13	For $a = b+2$ and $n \geq 6$, the trees $T^* \in \mathcal{T}_{a,b}$ with minimum index λ_1 are comets $Co(n,3)$. Moreover $\lim_{n\to\infty} \lambda_1(T^*) = 2.$ For $a = b+3$ and $n \geq 7$, the trees $T^* \in \mathcal{T}_{a,b}$ with minimum index λ_1 are double comets $D(3,n-6,3)$ and $\lambda_1(T^*) = 2$.	Proved	[88]
Co. AGX 4	Let G be a graph with $n \geq 3$ nodes, then $r + Ra - mode(d) \geq \sqrt{n-1} - 1.$ (Reinforcement of **Conjecture WOW 7**)	Open	[56]

over 50 new conjectures, 15 of which have been proved.

AutoGraphiX2 and automated comparison of graph invariants

Recently, Aouchiche, Caporossi and Hansen proposed new types of fully automated conjectures [157], using the new version of the system, called AutoGraphiX 2 . They made a systematic comparison of 20 graph invariants : for each pair of them, AutoGraphiX 2 considers the four operations $+$, $-$, $\times$, $/$ and derives best possible lower and upper bounding functions of the number of vertices, as well as extremal graphs. For instance, for two invariants i_1 and i_2, AutoGraphiX 2 tries to find functions f and g such that:

$$f^+(n) \leq i_1 + i_2 \leq g^+(n),$$

$$f^-(n) \leq i_1 - i_2 \leq g^-(n),$$

$$f^\times(n) \leq i_1\ i_2 \leq g^\times(n),$$

and

$$f^/(n) \leq \frac{i_1}{i_2} \leq g^/(n).$$

As 20 invariants are considered, it makes 190 different pairs of invariants and a total of 1520 problems. AutoGraphiX 2 recognizes 35 known results and derives automatically conjectures in 1341 cases. Among the other cases, 84 conjectures have been found with assistance of AutoGraphiX 2 but not automatically (*assisted conjectures*) and 60 cases lead to no exploitable results. The 1341 automated conjectures can be classified as follows :

i) *Results proved automatically* by AutoGraphiX 2 (often easy results).

ii) *Automated conjectures*. AutoGraphiX 2 derives algebraic conjectures and corresponding extremal graphs (e.g., a class of graph as paths or stars).

iii) *Structural conjectures*. AutoGraphiX 2 derives only conjectures about the extremal graphs.

These results are summarized here :

Known results	35 (2.30 %)
Results proved automatically	839 (55.20 %)
Automated conjectures	422 (27.76 %)
Structural conjectures	80 (5.26 %)
Assisted conjectures	84 (5.53 %)
No results	60 (3.95 %)

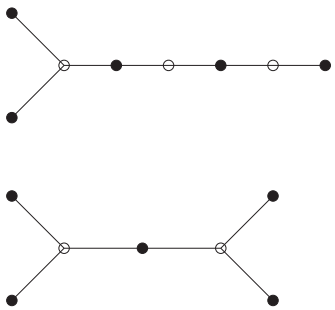


Figure 2.2: Color-constrained trees with minimum index : comets and double comets

Among the 586 conjectures[3] derived by AutoGraphiX 2 or with its help, 352 were proved by hand, 13 refuted and 221 conjectures remain open.

Examples

We next give a few examples of the results obtained with AutoGraphiX. Others are listed in Table 2.4 and Part II of this dissertation is devoted to the personal results, also obtained with AutoGraphiX. All these results were obtained with the initial version of AutoGraphiX.

We recall a few definitions used there. A *comet* $Co(n,n_1)$, of order n and with n_1 pending nodes, is obtained from a star S_{n_1+1} by appending a path with $n-n_1-1$ edges to a pending node. A *double comet* $D_{p,t,q}$ is obtained from two stars on p and q nodes by joining a pending node of each of them with a path of $t+1$ edges (see Figure 2.2). The *radius* r of graph G is the minimum over all nodes of the maximum distance from that node to any other.

Observe that the conjectures in Table 2.4 are of two types, concerning

(a) relations between graph invariants and

(b) structure of extremal graphs.

Example 2.12. A study of graphs with extremal energy [50], parameterizing by n and m led, among others, to the conjectures

$$E \geq 2\sqrt{m} \quad \text{and} \quad E \geq \frac{4m}{n}.$$

Both of them have been easily proved.

Example 2.13. In the same study, unicyclic graphs with maximum energy were investigated. This led to the following structural result.

[3]The number 586 is the sum of the numbers of automated conjectures, structural conjectures and assisted conjectures.

Conjecture AGX 16 *Among unicyclic graphs with n nodes the cycle C_n has maximum energy if $n \leq 7$ or $n = 9$, 10, 11, 13 and 15; otherwise the unicyclic graph with maximum energy is $C_6 + P_{n-6}$ i.e., C_6 with an appended path with $n-6$ edges.*

Partial results towards the proof of this conjecture were recently obtained : it is shown in [147] that among bipartite unicyclic graphs those with maximum energy are either C_n or $C_6 + P_{n-6}$.

Automated conjectures

The results of these two examples were obtained interactively. However, there are several ways to use AutoGraphiX in an entirely automated way [56]. Indeed, conjectures can be found by

(i) a numerical procedure which exploits the mathematics of principal components analysis in order to find resemblances instead of differences between extremal graphs. This leads, in polynomial time, to a basis of affine relations between graph invariants;

(ii) a geometric procedure, i.e., finding the convex hull of the set of extremal graphs viewed as points in invariant space: facets of this convex hull give linear relations, i.e., lower and upper bounds on the invariants associated with each of the axes;

(iii) an algebraic procedure, i.e., recognizing the class of extremal graphs found, and if it is a well-defined one for which formulae relating graph invariants are known, eliminating variables to obtain simple relations between the invariants under study.

Example 2.14. In [88] *color-constrained trees* (i.e., trees with given numbers of black or white nodes) with minimum index are investigated. A further study of the extremal graphs found was performed in [54, 56]. To this effect 15 graph invariants were computed and the numerical method applied. In addition to some trivial relations it led to the following outcome.

Conjecture AGX 9 *In all color-constrained trees with minimum index,*

$$\alpha = \frac{1}{2}(m + n_1 + D - 2r)$$

where α denotes the stability number, m the size, n_1 the number of pending nodes, D the diameter and r the radius.

It is unlikely that a relation with as many invariants could have been obtained without a computer.

This conjecture is open; it does not hold for all trees. However, it can be shown [56] that the right-hand side is an upper bound on the stability number of trees. So **Conjecture AGX 9** implies that color-constrained trees with maximum index have maximum stability number.

AutoGraphiX can also obtain conjectures with real numbers (approximated to a reasonable extent, as computations are made by machine). The following conjecture was also found by the numerical procedure.

Conjecture 2.1. *For any tree T of size m with maximum degree $\Delta \leq 3$ and maximum irregularity $A(T)$, Randic index $Ra(T)$, and n_1 pending nodes,*

$$Ra(T) = -0.027421\,A(T) + 0.538005\,m - 0.1104848\,n_1 + 0.614014.$$

This conjecture is proved in Appendix A.1. Extremal trees have nodes of degree 3 and 1 alternatingly, as far as possible.

Example 2.15. Use of the geometric approach [56] to study chemical graphs (i.e., graphs with a maximum degree of 4) led to the result :

Theorem AGX 5 *In all chemical graphs,*

$$Ra \geq \frac{1}{4}(n_1 + m).$$

Example 2.16. Conjecture 8 of *Graffiti* is that in a graph G,

$$\bar{l} + Ra - \text{ mode}(d) \geq 0$$

where $\bar{l}$ denotes the average distance between two distinct nodes and mode(d) is the statistical mode of the vector of degrees of G. Use of the algebraic approach led to the following strengthening of that result :

Conjecture AGX 5 *In a graph G,*

$$\bar{l} + Ra - \text{ mode}(d) \geq \frac{2(n-1)}{n} + \sqrt{n-1} - 2 \text{ if } n \geq 3.$$

2.6 Summary

Several discovery systems in graph theory have been very successful in helping mathematicians to formulate and explore conjectures, or to suggest interesting conjectures in an entirely automated way. Moreover, new systems sometimes based on new principles are being developed. The underlying paradigms, i.e., enumeration, interactive computing, formula manipulation, generation and

selection, heuristic optimization, are varied. They appear to be largely complementary. So one may expect much activity and the advent of more comprehensive systems in the near future.

The forms of conjectures found with these systems and some unexplored form of conjectures are studied in [160].

CHAPTER 3

Enumeration and integral graphs

Conjectured families of integral graphs • Integrality condition for the complete product of graphs • Integral complete split graphs • Integral extended complete split-like graphs • Summary

In this chapter, several results, obtained partially using enumeration, are given. We used enumeration to characterize some infinite families of integral graphs. Recall that a graph is called integral if all its eigenvalues are integer. The search for integral graphs, initiated by F. Harary and A. Schwenk in [172] and continued in many papers thereafter, revealed that not only the number of integral graphs is infinite, but that one can find them in almost all classes of graphs. For a recent survey on integral graphs, see [17]. In [16, 18] it is shown that there are only 263 non-isomorphic connected integral graphs with up to 11 nodes. These graphs can be determined by an enumeration of connected graphs and computation of their spectra. Larger integral graphs, but possibly not all of them for given n, can be obtained with an evolutionary algorithm, using as fitness function the sum of distances from eigenvalues to their closest integer (and variants thereof) [19].

Among other results, examination of the 263 integral graphs suggested a new infinite family of integral graphs [18]. These graphs are constructed from the complete bipartite graphs $K_{a,a+2}$ for $a = 1, 2, \ldots$ by appending an edge to each node of the smallest independent set (see Figure 3.1).

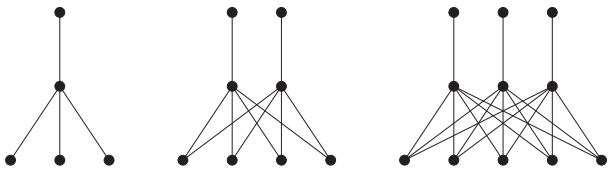


Figure 3.1: A new family of bipartite integral graphs [18]

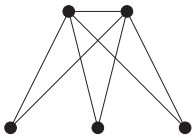


Figure 3.2: Small integral complete split graph CS_2^3

A *split graph* SP_b^a is composed of an independent set on a nodes, a clique on b nodes and some edges joining pairs of nodes one in each of those sets. It is *complete* if it contains edges for all such pairs (see Figure 3.2). Complete split graphs are noted CS_b^a. Notice that

$$CS_b^a \cong \overline{K}_a \nabla K_b.$$

It was observed in [166, 206] that complete split graphs are sometimes but not always integral (see again Figure 3.2). This particular class of graphs appears often in this dissertation. Indeed, it is again used in Sections 5.2, 7.1.2 and 7.3.

For $a, b, n \in \mathbb{N}$ we define the following related classes of graphs:

- the *multiple complete split-like graph* $MCS_{b,n}^a \cong \overline{K}_a \nabla (nK_b)$;
- the *extended complete split-like graph* $ECS_b^a \cong \overline{K}_a \nabla (K_b + K_2)$;
- the *multiple extended complete split-like graph*

$$MECS_{b,n}^a \cong \overline{K}_a \nabla \Big(n(K_b + K_2) \Big).$$

In Fig. 3.3 we have shown examples of graphs from all of these classes.

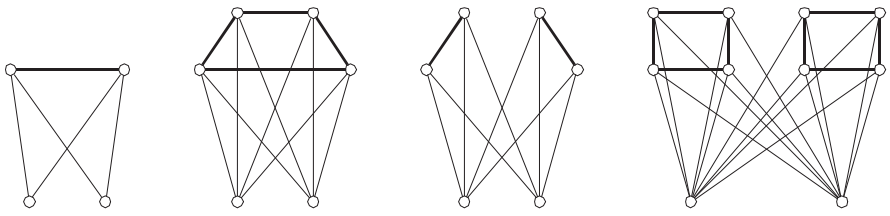


Figure 3.3: Examples of split-like graphs.

In the next section, we present the conjectures derived by a generation of complete split graphs, and a computation of their spectrum. In Section 3.2, we give an integrality condition for the complete product of graphs. Then, in Section 3.3, we characterize integral graphs in the families

of complete split graphs and multiple complete split-like graphs. Finally, in Section 3.4, we also characterize integral graphs in the families of extended complete split-like graphs and multiple extended complete split-like graphs.

3.1 Conjectured families of integral graphs

In [166], we proposed a generation of all complete split graphs with $n = a + b \leq 500$ and $b \leq 50$, and a computation of their spectrum. This program was written with the *Matlab*^{TM} [191] programming language . These computations lead to conjecture two types of integral graphs (in terms of a and b), in the family of complete split graphs.

Theorem 3.1. *All complete split graphs CS_b^a with*

$$a = \left\lceil \frac{i}{2} \right\rceil^2 + (b-1) \left\lfloor \frac{i}{2} \right\rfloor \left\lfloor \frac{i+2}{2} \right\rfloor \tag{3.1}$$

where i is a positive integer, are integral. Moreover, if b is a power of a prime, then there are no other integral complete split graph CSa_b with this value of b.

Theorem 3.2. *All complete split graphs CS_b^a satisfying $b = 4k+2$ for $k \in \mathbb{N}$ and*

$$a = -\left\lfloor \frac{b}{4} \right\rfloor + (b-1) \left\lceil \frac{i}{2} \right\rceil^2 + \left\lfloor \frac{i}{2} \right\rfloor \left\lfloor \frac{i+2}{2} \right\rfloor \tag{3.2}$$

where i is a positive integer, are integral.

These theorems are proven in Section 3.3.

3.2 Integrality condition for the complete product of graphs

The following theorem is proven in [120].

Theorem 3.3. [120] *For $i = 1,2$ let G_i be r_i-regular graphs with n_i nodes. The characteristic polynomial of the complete product of graphs G_1 and G_2 is given by the relation*

$$P_{G_1 \nabla G_2}(\lambda) = \frac{P_{G_1}(\lambda) P_{G_2}(\lambda)}{(\lambda - r_1)(\lambda - r_2)} [(\lambda - r_1)(\lambda - r_2) - n_1 n_2].$$

In the following we give a necessary and sufficient condition for the complete product of two regular graphs to be integral.

Corollary 3.1. *For $i = 1,2$ let G_i be r_i-regular graphs with n_i nodes. The complete product $G_1 \nabla G_2$ is an integral graph if and only if both G_1 and G_2 are integral graphs and there exist $k \in \mathbb{N}$ such that the* integrality condition

$$n_1 n_2 = k(k + |r_1 - r_2|). \tag{3.3}$$

holds.

Proof. Since G_i is r_i-regular graph, its largest eigenvalue is equal to r_i [80] (with the eigenvector equal to all-1 vector) and the fraction $\frac{P_{G_i}(\lambda)}{\lambda - r_i}$ is the polynomial with roots equal to the remaining eigenvalues of G_i with the same multiplicities. Thus in order that the complete product $G_1 \nabla G_2$ of two regular graphs is integral both G_1 and G_2 must be integral and the expression

$$Q(\lambda) = (\lambda - r_1)(\lambda - r_2) - n_1 n_2$$

must have integer roots. The roots of $Q(\lambda)$ are equal to

$$\lambda_{1,2} = \frac{r_1 + r_2 \pm \sqrt{(r_1 - r_2)^2 + 4n_1 n_2}}{2},$$

and they are integers if and only if $r_1 + r_2$ and $\sqrt{(r_1 - r_2)^2 + 4n_1 n_2}$ are integers of the same parity. The last fact means that there exists $k \in \mathbb{N}$ such that $(r_1 - r_2)^2 + 4n_1 n_2 = (|r_1 - r_2| + 2k)^2$, wherefrom we get the integrality condition (3.3). □

3.3 Integral complete split graphs

Using Corollary 3.1, we now characterize integral graphs in the families of complete split graphs and multiple complete split-like graphs, by giving the explicit formulae for their parameters a, b and n.

The greatest common divisor of two integers p and q is denoted (p,q).

Theorem 3.4. *The complete split graph CS_b^a is integral if and only if there exist $p,q \in \mathbb{N}$ with $(p,q) = 1$ and $c \in \mathbb{Z}$ such that*

$$x + cq > 0, \quad a = (x + cq)(y + p + cp), \quad \textit{and} \quad b = pq,$$

where $x,y \in \mathbb{Z}$ are determined by the Euclidean algorithm such that $px - qy = 1$.

Proof. We have that $CS_b^a \cong \overline{K}_a \nabla K_b$. Graph $\overline{K}_a$ is 0-regular ($n_1 = a, r_1 = 0$), while K_b is $b-1$-regular ($n_2 = b, r_2 = b-1$). By Corollary 3.1, graph CS_b^a is integral if and only if there exists $k \in \mathbb{N}$ such that $ab = k(k + b - 1)$, hence

$$a = k + \frac{k(k-1)}{b} \quad \text{and} \quad b \text{ divides } k(k-1). \tag{3.4}$$

Since the greatest common divisor $(k, k-1)$ of k and $k-1$ is equal to 1, we have that

$$b = (b, k(k-1)) = (b,k) \cdot (b, k-1).$$

Let $p = (b,k)$ and $q = (b,k-1)$. Thus $b = pq$, $(p,q) = 1$, $p|k$ and $q|k-1$; where $|$ means *divides*. Let $x,y \in \mathbb{Z}$ be determined by the Euclidean algorithm such that $px - qy = 1$, and let $k' = k - px =$

$k-1-qy$. From $p|k$ it follows that $p|k'$, while from $q|k-1$ it follows that $q|k'$. Since $(p,q)=1$ we have that $pq|k'$ and

$$k = px + cpq \quad \text{for some } c \in \mathbb{Z},$$

while $k-1 = qy+cpq$. Now from (3.4) it follows that

$$a = px+cpq+(x+cq)(y+cp) = (x+cq)(y+p+cp). \qquad \square$$

Let us see how to apply Theorem 3.4 to the conjectured families (3.1) and (3.2) of integral complete split graphs.

Proof of Theorems 3.1 and 3.2. The family (3.1) for odd i corresponds to the case $p=1$, $q=b$, $x=1$, $y=0$ and $c=\left\lfloor \frac{i}{2} \right\rfloor$, while for even i it corresponds to the case $p=b$, $q=1$, $x=0$, $y=-1$ and $c=\left\lfloor \frac{i}{2} \right\rfloor$.

Moreover, if b is a power of prime, we have only two possibilities. The first possibility is that $p=b$ and $q=1$, and thus $x=1$ and $y=(b-1)$. In this case we have that

$$a = (1+c)(b-1+b+cb) = (c+1)^2+(c+1)(c+2)(b-1),$$

which corresponds to $i=2c+2$. The second possibility is that $p=1$ and $q=b$, wherefrom $x=b+1$ and $y=1$. In this case we have that

$$a = (b+1+cb)(1+1+c) = (c+2)^2+(c+1)(c+2)(b-1),$$

which corresponds to $i=2c+3$.

The family (3.2) for odd i corresponds to the case $p=2k+1$, $q=2$, $x=1$, $y=k$ and $c=\left\lfloor \frac{i}{2} \right\rfloor$, while for even i it corresponds to the case $p=2$, $q=2k+1$, $x=k+1$, $y=1$ and $c=\left\lfloor \frac{i}{2} \right\rfloor - 1$.

Theorems 3.1 and 3.2 are thus proved. $\square$

We characterize now integral multiple complete split-like graphs.

Theorem 3.5. *The multiple complete split graph $MCS^a_{b,n}$ is integral if and only if there exist $p,q \in \mathbb{N}$ with $(p,q)=1$ and $c \in \mathbb{Z}$ such that*

$$x+cq>0, \quad an = (x+cq)(y+p+cp), \quad \textit{and} \quad b = pq,$$

where $x,y \in \mathbb{Z}$ are determined by the Euclidean algorithm such that $px-qy=1$.

Proof. Multiple complete split-like graphs provide a simple generalization of complete split graphs. We have that $MCS^a_{b,n} \cong \overline{K}_a \nabla (nK_b)$. Graph $\overline{K}_a$ is 0-regular ($n_1=a, r_1=0$), while nK_b is $b-1$-regular ($n_2=nb, r_2=b-1$). By Corollary 3.1, graph $MCS^a_{b,n}$ is integral if and only if there exists $k \in \mathbb{N}$ such that $anb = k(k+b-1)$, hence

$$an = k + \frac{k(k-1)}{b} \quad \text{and} \quad b \text{ divides } k(k-1).$$

Repeating the consideration from above we get the result. □

3.4 Integral extended complete split-like graphs

One can also characterize integral graphs in the families of extended complete split-like graphs and multiple extended complete split-like graphs.

Theorem 3.6. *Let $b = p_1^{x_1} p_2^{x_2} \dots p_r^{x_r}$ be a prime factorization of b. The extended complete split graph ECS_b^a is integral if and only if there exists $c \in \mathbb{N}$ such that*

$$a = \frac{1}{2} \cdot c \cdot \left(p_1^{\lceil \frac{x_1}{2} \rceil} p_2^{\lceil \frac{x_2}{2} \rceil} \cdots p_r^{\lceil \frac{x_r}{2} \rceil} + p_1^{\mathrm{odd}(x_1)} p_2^{\mathrm{odd}(x_2)} \cdots p_r^{\mathrm{odd}(x_r)} \cdot c \right).$$

and either c is even or b is odd or the highest power of 2 which divides b is odd.

Proof. We have that $ECS_b^a \cong \overline{K}_a \nabla (K_b + K_2)$. Graph $\overline{K}_a$ is 0-regular ($n_1 = a, r_1 = 0$), while $K_b + K_2$ is b-regular ($n_2 = 2b, r_2 = b$). By Corollary 3.1, graph ECS_b^a is integral if and only if there exists $k \in \mathbb{N}$ such that $2ab = k(k+b)$. From this it follows that

$$a = \frac{1}{2} \left(k + \frac{k^2}{b} \right) \quad \text{and} \quad 2b \text{ divides } k^2. \tag{3.5}$$

Let

$$b = p_1^{x_1} p_2^{x_2} \dots p_r^{x_r}$$

be a prime factorization of b, such that $p_1 \le p_2 \le \cdots \le p_r$. From $b | k^2$ it follows that $p_i^{\lceil \frac{x_i}{2} \rceil} | k$ for each $i = 1, 2, \dots, r$, and we can write

$$k = p_1^{\lceil \frac{x_1}{2} \rceil} p_2^{\lceil \frac{x_2}{2} \rceil} \cdots p_r^{\lceil \frac{x_r}{2} \rceil} \cdot c, \quad c \in \mathbb{N}.$$

Let $\mathrm{odd}(n)$, $n \in \mathbb{N}$, be the characteristic function of the set of odd numbers, i.e.

$$\mathrm{odd}(n) = \begin{cases} 1, & \text{if } n \text{ is odd} \\ 0, & \text{if } n \text{ is even} \end{cases} = 2 \left\lceil \frac{n}{2} \right\rceil - n.$$

From (3.5) it follows that

$$a = \frac{1}{2} \cdot c \cdot \left(p_1^{\lceil \frac{x_1}{2} \rceil} p_2^{\lceil \frac{x_2}{2} \rceil} \cdots p_r^{\lceil \frac{x_r}{2} \rceil} + p_1^{\mathrm{odd}(x_1)} p_2^{\mathrm{odd}(x_2)} \cdots p_r^{\mathrm{odd}(x_r)} \cdot c \right). \tag{3.6}$$

We see that $a \in \mathbb{N}$ if either c is even or b is odd. In this second case, all primes p_1, p_2, …, p_r are odd and the right hand side of (3.6) contains the product of two numbers of distinct parities, which is always even.

In the remaining case when both c is odd and b is even, it must hold that

$$2 \mid p_1^{\lceil \frac{x_1}{2} \rceil} p_2^{\lceil \frac{x_2}{2} \rceil} \cdots p_r^{\lceil \frac{x_r}{2} \rceil} + p_1^{\text{odd}(x_1)} p_2^{\text{odd}(x_2)} \cdots p_r^{\text{odd}(x_r)} \cdot c.$$

Since b is even, we have that $p_1 = 2$ and $\lceil \frac{x_1}{2} \rceil \geq 1$, so that the above condition, taking into account that $p_2, \ldots, p_r$ are all odd, becomes $2 \mid p_1^{\text{odd}(x_1)}$, which is equivalent to the condition that x_1 is odd. □

Theorem 3.7. *Let $b = p_1^{x_1} p_2^{x_2} \ldots p_r^{x_r}$ be a prime factorization of b. The multiple extended complete split graph $MECS_{b,n}^a$ is integral if and only if there exists $c \in \mathbb{N}$ such that*

$$an = \frac{1}{2} \cdot c \cdot \left(p_1^{\lceil \frac{x_1}{2} \rceil} p_2^{\lceil \frac{x_2}{2} \rceil} \cdots p_r^{\lceil \frac{x_r}{2} \rceil} + p_1^{\text{odd}(x_1)} p_2^{\text{odd}(x_2)} \cdots p_r^{\text{odd}(x_r)} \cdot c \right).$$

and either c is even or b is odd or the highest power of 2 which divides b is odd.

Proof. Similar as with complete split graphs, multiple extended complete split-like graphs provide a simple generalization of extended complete split-like graphs. We have that $MECS_{b,n}^a \cong \overline{K}_a \nabla \Big(n(K_b + K_2) \Big)$. Graph $\overline{K}_a$ is 0-regular ($n_1 = a, r_1 = 0$), while $n(K_b + K_2)$ is b-regular ($n_2 = 2nb, r_2 = b$). By Corollary 3.1, graph $MECS_{b,n}^a$ is integral if and only if there exists $k \in \mathbb{N}$ such that $2anb = k(k+b)$. From this it follows that

$$an = \frac{1}{2} \left(k + \frac{k^2}{b} \right) \quad \text{and} \quad 2b \text{ divides } k^2.$$

Repeating the consideration from above we get the result. □

3.5 Summary

In this chapter, we presented a characterization of integral graphs in several families of split-like graphs. This work was started by an enumeration of complete split graphs and a computation of their spectra.

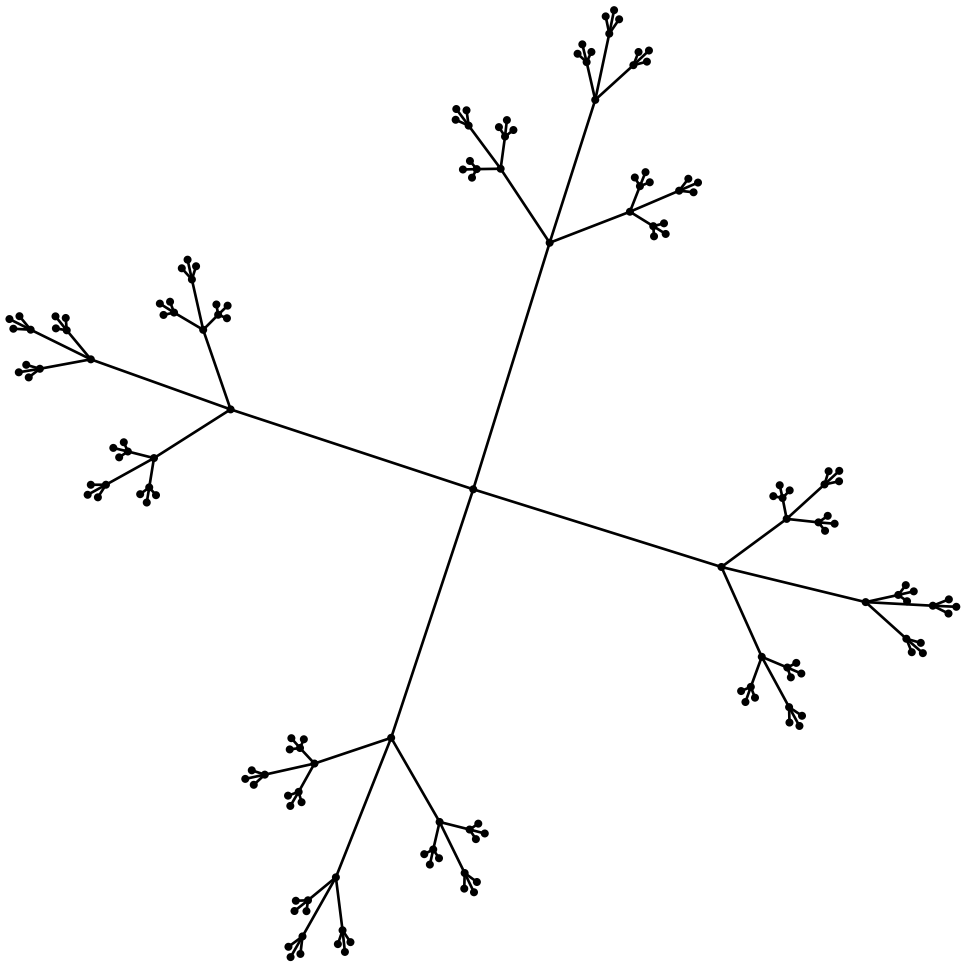
We introduced an integrality condition for the complete product of graphs and used it intensively in the proofs, since each studied family can be expressed as a complete product of two graphs. One can also consider a more general case of graph $\overline{K}_a \nabla (G + K_n)$, where G is an integral, r-regular graph with m nodes. Here we have that $n_1 = a, r_1 = 0$ and $n_2 = mn, r_2 = r + n - 1$, and the integrality condition (3.3) says that $\overline{K}_a \nabla (G + K_n)$ is an integral graph if and only if there exists $k \in \mathbb{N}$ such that

$$amn = k(k + r + n - 1).$$

If $a = mn + r + n - 1$ then $k = mn$ satisfies the above condition, but the task of characterization of set of parameters a, m, r, n for which the graph $\overline{K}_a \nabla (G + K_n)$ is integral appears to be difficult.

Part II

Results obtained with AutoGraphiX



Introduction

In Section 2.5 we presented AutoGraphiX, a computer system developed by Caporossi and Hansen. In [55], its main features are illustrated; it can

(i) find graphs satisfying various constraints;

(ii) find graphs with extremal values for some invariant, possibly subject to constraints;

(iii) refute conjectures;

(iv) suggest conjectures;

(v) give ideas of proofs, or show that some ideas are not likely to succeed.

Three ways to automate the generation of conjectures obtained by AutoGraphiX are outlined in [56] :

(a) a numerical one, based on the mathematics of principal component analysis (data mining), where instead of explaining differences between extremal graphs (observations), resemblances between them are captured in the form of a basis of affine relations between invariants;

(b) a geometric one, in which the convex hull of points representing extremal graphs in invariants space is computed and provides linear inequalities corresponding to facets;

(c) an algebraic one, which recognizes families of extremal graphs and manipulates known formulae for invariants on these families to obtain new relations.

Further papers describing mathematical applications of AutoGraphiX are [4, 51, 88]. We also refer to the PhD dissertation of Caporossi [49] and to a recent survey on AutoGraphiX is [5]. Several papers on the use of AutoGraphiX in chemical graph theory have also appeared. In [50] a study is made of the energy of a graph. New relations between energy, size and order of graphs were obtained and easily proven (see Example 2.12, p. 39). They appear to be simpler in most cases than those found in literature and yet previously unnoticed. Conjectures on the structure of

extremal graphs have also be found even though they appear to be harder to prove. For instance, we recall Conjecture AGX 16 presented in Section 2.5 : "Among unicyclic graphs on n nodes, the cycle C_n has maximum energy if $n \leq 7$ and $n = 9, 11, 13, 15$. For all other values of n the unicyclic graph with maximal energy is composed of a 6-cycle and an appended path with $n-6$ edges". Partial results towards a proof of this conjecture were recently obtained by Gutman and Hou [147].

In [52] chemical trees with minimum Randić index are characterized; they belong (or are equivalent through a transformation) to three families, two of which are caterpillars[1]. These families were easily obtained with AutoGraphiX. These results led to several further papers in which cyclic graphs are considered [53, 142, 143, 148, 149, 150]. The proof of Theorem 2 of [52] relies on the use of linear and mixed-integer programming, where variables are associated with numbers of edges with given endpoint's degrees instead of with individual edges, as in most graph-theoretic models. We use the same technique in some of the proofs of the next sections (see the proofs of Theorems 4.5, 4.7 and 4.12).

In [127] chemical trees with maximum HOMO-LUMO gap are studied (see Definition 1.17, p. 22). For even order they are easily characterized, which is a result already obtained by Shao and Hong [217]. For odd order their form is more complicated and does not appear to be unique.

In Chapter 4, we present several results in chemical graph theory, obtained using AutoGraphiX, in collaboration with Hansen and partially with Gutman [146, 165, 163].

Chapter 5 presents results on the irregularity also obtained with the help of AutoGraphiX , in collaboration with Hansen, but no more restricted to chemical graph theory [164]. In this chapter, we introduce a *simulated algorithm* type of proof. This proof was derived by the use of the *VNS* heuristic implemented in AutoGraphiX.

[1] A caterpillar is a tree composed of a path and pending edges appended to some or all of its nodes.

CHAPTER 4

Results in chemical graph theory

Bounds on the Randić index of chemical graphs • Comparison of irregularity indices for chemical trees • Variance of bounded degrees in graphs • Summary

In Section 1.5, we introduced chemical graph theory. Computers are much used, in this field of graph theory, to evaluate invariants as well as correlations between chemical activities or properties and those invariants or functions of them in *QSAR* and *QSPR* studies [66, 92, 94, 99, 125, 156, 179, 181, 182, 183, 190, 227, 229, 230]. They can be also used to enhance chemical graph theory *per se*, e.g., in finding extremal graphs for invariants or new relations between invariants.

We present in this chapter results in chemical graph theory obtained with the help of AutoGraphiX. In the next section, bounds on the Randić index of chemical graphs are given [163]. We compare three irregularity indices for chemical trees in Section 4.2 [146]. Finally, relations between the variance of degrees VAR, the maximum degree Δ, the order n and the size m of connected graphs, are studied in Section 4.3. From a general result on these invariants (see Theorem 4.13, p. 105), special cases of chemical interest are considered [165].

4.1 Bounds on the Randić index of chemical graphs

The Randić index *Ra* (see Definition 1.15, p. 22) was conceived in 1975 by Randić [210]. It is also called the *connectivity index*. Together with its generalizations it is certainly the molecular-graph-based structure-descriptor, that found the most numerous applications in organic chemistry, medicinal chemistry and pharmacology.

Recall that for given adjacent nodes v_1 and v_2 of a graph, the *weight* of the edge $\{v_1, v_2\}$ is defined by

$$\frac{1}{\sqrt{d_{v_1} d_{v_2}}}. \tag{4.1}$$

The Randić index of a graph is thus the sum of the weights of all edges of the graph.

In this section we study the Randić index of trees, chemical trees and chemical graphs as functions of the order and number of pending nodes or cyclomatic number (defined below, p. 57) of the tree or graph. Our point of departure is a recent paper of Araujo and De la Peña [11] on these problems, which presents several lower and upper bounds on the Randić index . With the help of AutoGraphiX all these results are improved (and in one case corrected). Let us now review the results of [11].

4.1.1 Bounds of Araujo and De la Peña

In order to present the bounds of Araujo and De la Peña about the Randić index, we recall the notion of the ramification index.

Ramification index

In [11] the *ramification index* of a tree T is defined as

$$ram(T) = \sum_{v \in V,\ d_v \geq 3} (d_v - 2).$$

We prove below that this index is equivalent, up to a constant, to the number of pending nodes n_1 of T. To this effect, we establish a more general result.

Theorem 4.1. *For all graph G, with n nodes, n_1 pending nodes, m edges and ramification index $ram(G)$,*

$$ram(G) = 2(m-n) + n_1.$$

Proof. By definition of the ramification index,

$$ram(G) = \sum_{v \in V} (d_v - 2) - \sum_{v \in V,\ d_v < 3} (d_v - 2). \tag{4.2}$$

By easy observations when $d_v = 2$ or 1,

$$\sum_{v \in V,\ d_v < 3} (d_v - 2) = -n_1, \tag{4.3}$$

and (4.2) becomes

$$ram(G) = \sum_{v \in V} d_v - 2n + n_1. \tag{4.4}$$

As for every graph G,

$$\sum_{v \in V} d_v = 2m, \tag{4.5}$$

we get the result

$$ram(G) = 2(m-n)+n_1. \tag{4.6}$$

□

Corollary 4.1. *If T is a tree,*

$$ram(T) = n_1 - 2.$$

Proof. Replacing m by $n-1$ in Theorem 4.1 gives the result. □

Recall that the *cyclomatic number* $\mu = m-n+1$ of a connected graph G represents the maximum number of fundamental cycles of G.

Corollary 4.2. *If G is a connected graph,*

$$ram(G) = n_1 + 2\mu - 2.$$

Proof. Replacing $m-n$ by $\mu-1$ in Theorem 4.1 gives the result. □

Randić index of P_n and $Co(n,n_1)$

Recall that a *comet* is a tree composed of a star and an appended path. For any number n and $2 \leq n_1 \leq n-1$, we denote by $Co(n,n_1)$ the comet of order n with n_1 pending nodes, i.e., a tree formed by a path P_{n-n_1-1} of which one end node coincides with a pending node of a star S_{n_1+1} of order n_1+1 (see Figure 4.1).

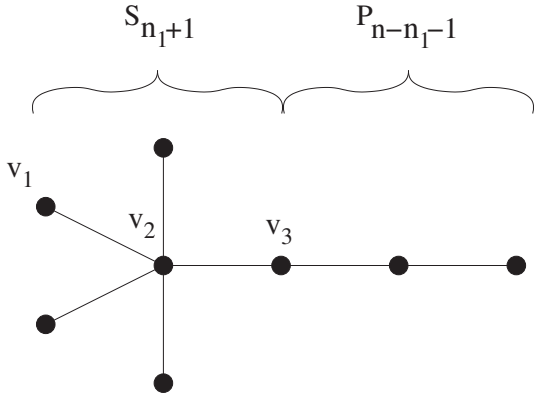


Figure 4.1: A comet of order 8 with 5 pending nodes

By definition of the Randić index, we have if $n > 2$,

$$Ra(P_n) = \frac{n}{2} - \left(\frac{3}{2} - \sqrt{2}\right), \tag{4.7}$$

and if $2 \leq n_1 < n-1$,

$$Ra(Co(n,n_1)) = \sqrt{n_1} + \left(\frac{1}{\sqrt{2}} - 1\right)\frac{1}{\sqrt{n_1}} + \frac{n-n_1-2}{2} + \frac{1}{\sqrt{2}}. \tag{4.8}$$

If $n_1 = n-1$, T is a star S_n and $Ra(T) = \sqrt{n-1}$.

Randić index of chemical trees

A first theorem of [11] bounds the Randić index of a chemical tree T of order n in terms of its ramification index and of the Randić indices of a comet and a path of the same order.

Theorem 4.2. ***[11]*** *Let T be a chemical tree with n nodes and $ram(T) = t-2$. Then*

$$Ra(Co(n,t)) - c_0(ram(T)-1) \leq Ra(T) \leq Ra(P_n) - a_0\, ram(T)$$

where $a_0 = 1 - \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{6}}$ ($\simeq 0.0144$) and $c_0 = 0$ if $ram(T) = 0$ and otherwise $c_0 = \frac{\sqrt{3}}{2} - \frac{3}{4}$ ($\simeq 0.1160$).

With the help of AutoGraphiX, we improve these two bounds, and characterize the corresponding extremal chemical trees. These results are presented in Section 4.1.3.

It is known that $Ra(P_n)$ is maximum for all trees with n nodes (see [52, 53, 234]) and that $Ra(S_n)$ is minimal for all graphs of order n [36]. One may wonder if a similar property holds for $Ra(n,n_1)$. AutoGraphiX suggests that the trees with minimal Randić index and a fixed number of pending nodes are the comets. This is proven in Section 4.1.2 (as this result is needed in Section 4.1.3).

Randić index of chemical graphs

A second theorem of [11] gives bounds for the Randić index of chemical graphs.

Theorem 4.3. ***[11]*** *Let G be a chemical graph and T be a maximal subgraph of G which is a tree. Then*

$$Ra(T) - d_0\mu(G) \leq Ra(G) \leq Ra(T) + b_0\mu(G)$$

where $b_0 = -\frac{1}{2\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{6}} - \frac{1}{2}$ ($\simeq 0.1320$) and $d_0 = \sqrt{2} + \frac{1}{2\sqrt{3}} - \frac{3}{2}$ ($\simeq 0.2029$). In particular if G has n nodes and $ram(T) = t-2$, then

$$Ra(Co(n,t)) - c_0(ram(T)-1) - d_0\,\mu(G) \leq Ra(G),$$

and

$$Ra(G) \leq Ra(P_n) - a_0\, ram(T) + b_0\mu(G).$$

Note that, in view of Corollary 4.1, G has t pending nodes. The coefficients a_0 and c_0 are the same as in Theorem 4.2. We give counter-examples, correct and slightly strengthen Theorem 4.3 in Section 4.1.4.

4.1.2 A new bound on the Randić index of trees

As explained above, we used AutoGraphiX to find chemical trees with minimal Randić index. The trees obtained are comets for $n_1 \leq 4$. This fact led us to run AutoGraphiX on trees that are not chemical, i.e., we minimized $Ra(T)$ where T is a tree with n and n_1 fixed. The system gave systematically extremal graphs which are comets, for all values of n and n_1 considered. We now prove that comets indeed minimize the Randić index.

Theorem 3 of [36] says that every graph G of order n, containing no isolated node has a Randić index greater than or equal to $\sqrt{n-1}$ with equality if and only if G is a star.

Let $\kappa = \{v_1, v_2\}$ be a pending edge of a graph G; we note $G - v_1v_2$ the graph obtained by removing the node of degree 1 of κ and the edge κ.

The following lemma is a variant of the Lemma 1 of [36], which applies when G is a tree, and which uses the number of pending nodes.

Lemma 4.1. *Let $\{v_1, v_2\}$ be an edge of a tree T of order n with $n_1 \leq n-2$ pending nodes. If $d_{v_1} = 1$ and $d_{v_2} \geq 2$ then*

$$Ra(T) - Ra(T - v_1v_2) \geq \sqrt{n_1} - \sqrt{n_1 - 1} + \left(\frac{1}{\sqrt{2}} - 1\right)\left(\frac{1}{\sqrt{n_1}} - \frac{1}{\sqrt{n_1 - 1}}\right),$$

with equality if and only if T is a comet and v_2 has maximum degree.

Proof. Denote by W_2 the sum of the weights of the edges, other than $\{v_1, v_2\}$, incident with node v_2. Note that, by definition of the Randić index ,

$$Ra(T) - Ra(T - v_1v_2) = \frac{1}{\sqrt{d_{v_2}}} + W_2 - W_2\sqrt{\frac{d_{v_2}}{d_{v_2} - 1}}. \tag{4.9}$$

If all nodes adjacent to v_2, other than v_1, have a degree of 1, then W_2 is maximal. But in this case, we have a star and $n_1 = n - 1$, which is excluded by assumption. So, at least one node adjacent to v_2 has a degree ≥ 2. Thus

$$W_2 \leq \frac{d_{v_2} - 2}{\sqrt{d_{v_2}}} + \frac{1}{\sqrt{2d_{v_2}}}, \tag{4.10}$$

with equality if and only if all the nodes adjacent to v_2 are pending nodes, except one, noted v_3, with $d_{v_3} = 2$ (see again Figure 4.1).

As

$$1-\sqrt{\frac{d_{v_2}}{d_{v_2}-1}}<0,$$

majorizing W_2 with (4.10) in (4.9) yields

$$Ra(T)-Ra(T-v_1v_2)\geq \frac{1}{\sqrt{d_{v_2}}}+$$

$$\left(\frac{d_{v_2}-2}{\sqrt{d_{v_2}}}+\frac{1}{\sqrt{2d_{v_2}}}\right)\left(1-\frac{\sqrt{d_{v_2}}}{\sqrt{d_{v_2}-1}}\right), \tag{4.11}$$

or, after easy manipulations,

$$Ra(T)-Ra(T-v_1v_2)\geq \sqrt{d_{v_2}}-\sqrt{d_{v_2}-1}+$$

$$\left(\frac{1}{\sqrt{2}}-1\right)\left(\frac{1}{\sqrt{d_{v_2}}}-\frac{1}{\sqrt{d_{v_2}-1}}\right). \tag{4.12}$$

Let $f(d_{v_2})$ denote the right hand side of (4.12). Then

$$f'(d_{v_2})=\frac{1}{2}\left[\frac{1}{\sqrt{d_{v_2}}}\left(1+\frac{\sqrt{2}-1}{\sqrt{2}}\frac{1}{d_{v_2}}\right)-\frac{1}{\sqrt{d_{v_2}-1}}\left(1+\frac{\sqrt{2}-1}{\sqrt{2}}\frac{1}{d_{v_2}-1}\right)\right],$$

and as

$$\frac{1}{\sqrt{d_{v_2}}}\left(1+\frac{\sqrt{2}-1}{\sqrt{2}}\frac{1}{d_{v_2}}\right)$$

decreases when d_{v_2} increases, $f'(d_{v_2})<0$. Hence, as $d_{v_2}\leq\Delta\leq n_1$, where Δ denotes the maximum degree of T,

$$Ra(T)-Ra(T-v_1v_2)\geq \sqrt{n_1}-\sqrt{n_1-1}+$$

$$\left(\frac{1}{\sqrt{2}}-1\right)\left(\frac{1}{\sqrt{n_1}}-\frac{1}{\sqrt{n_1-1}}\right), \tag{4.13}$$

with equality if and only if T is a comet, by construction of the graph obtained by taking equality in (4.10) and (4.11), where v_2 is the center of the star, and v_3 the end node of the path which belongs to the star (see once more Figure 4.1). □

We now give our result for trees with minimal Randić index and fixed number of pending nodes.

Theorem 4.4. *Let T be a tree of order $n\geq 3$, with n_1 pending nodes. Then if $n_1<n-1$,*

$$Ra(T)\geq \sqrt{n_1}+\left(\frac{1}{\sqrt{2}}-1\right)\frac{1}{\sqrt{n_1}}+\frac{n-n_1-2}{2}+\frac{1}{\sqrt{2}}, \tag{4.14}$$

with equality if and only if T is the comet $Co(n,n_1)$.

Proof. If T is a comet, then we have equality, since we have seen in expression (4.8) that the Randić index of a comet is equal to the right-hand side of (4.14).

We apply induction on n. It is easy to check that assertion (4.14) holds for the 4 trees of order 3, 4 and 5 which are not stars ($n_1 < n-1$). These trees are P_3, P_4, P_5 and $Co(5,3)$, which are all comets. So let us assume that $n \geq 6$ and that the result holds for all smaller values of n.

It is well known that every tree has at least two pending nodes and that if we remove a pending edge of a tree, the subgraph so obtained remains a tree. Let $\{v_1, v_2\}$ be a pending edge of T (with $d_{v_1} = 1$). We next prove the induction, discussing the degree of v_2.

i) If $d_{v_2} = 1$, then T is the only tree with two nodes connected by one edge, which contradicts the assumption $n \geq 6$.

ii) If $d_{v_2} = 2$, we note v_3 ($\neq v_1$) the node adjacent to v_2. We have,

$$Ra(T) - Ra(T - v_1v_2) = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{d_{v_3}}}\left(\frac{1}{\sqrt{2}} - 1\right).$$

As $\frac{1}{\sqrt{2}} - 1$ is negative, this expression is minimal when $\frac{1}{\sqrt{d_{v_3}}}$ is maximal, i.e., when d_{v_3} is minimal.

If $d_{v_3} = 1$, T is the path P_3, which is impossible by the assumption $n \geq 6$. So, $d_{v_3} = 2$ is minimal and

$$Ra(T) - Ra(T - v_1v_2) \geq \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}} - 1\right) = \frac{1}{2}, \tag{4.15}$$

with equality if and only if $d_{v_3} = 2$.

As $d_{v_2} = 2$, the tree $T - v_1v_2$ has n_1 pending nodes and $n-1$ nodes. In the case of $n_1 = n-2$, the tree $T - v_1v_2$ is a star, so T is a comet and we can stop the induction here. If $n_1 < n-2$, by induction the theorem holds and

$$Ra(T - v_1v_2) \geq \sqrt{n_1} + \left(\frac{1}{\sqrt{2}} - 1\right)\frac{1}{\sqrt{n_1}} + \frac{n - n_1 - 3}{2} + \frac{1}{\sqrt{2}}, \tag{4.16}$$

with equality if and only if $T - v_1v_2$ is a comet. So, by expressions (4.15) and (4.16)

$$Ra(T) \geq \sqrt{n_1} + \left(\frac{1}{\sqrt{2}} - 1\right)\frac{1}{\sqrt{n_1}} + \frac{n - n_1 - 2}{2} + \frac{1}{\sqrt{2}},$$

with equality if and only if $T - v_1v_2$ is a comet and $d_{v_3} = 2$, and that implies that we have equality if and only if T is also a comet.

iii) If $d_{v_2} \geq 3$, the tree $T - v_1v_2$ has $n_1 - 1$ pending nodes and $n-1$ nodes. Then, by the induction hypothesis,

$$Ra(T - v_1v_2) \geq \sqrt{n_1 - 1} + \left(\frac{1}{\sqrt{2}} - 1\right)\frac{1}{\sqrt{n_1 - 1}} + \frac{n - n_1 - 2}{2} + \frac{1}{\sqrt{2}}, \tag{4.17}$$

with equality if and only if $T - v_1v_2$ is a comet. By Lemma 4.1, we have

$$Ra(T) \geq Ra(T - v_1v_2) + \sqrt{n_1} - \sqrt{n_1 - 1} +$$

$$\left(\frac{1}{\sqrt{2}}-1\right)\left(\frac{1}{\sqrt{n_1}}-\frac{1}{\sqrt{n_1-1}}\right), \tag{4.18}$$

with equality if and only if T is a comet. Thus, by (4.17) and (4.18)

$$Ra(T) \geq \sqrt{n_1}+\left(\frac{1}{\sqrt{2}}-1\right)\frac{1}{\sqrt{n_1}}+\frac{n-n_1-2}{2}+\frac{1}{\sqrt{2}},$$

with equality if and only if T is a comet. □

An immediate consequence of Theorem 4.4 is that the term

$$-c_0(ram(T)-1)$$

can be disposed of in Theorem 4.2. However, we can do better, as shown in the next section.

4.1.3 Improvements of Theorem 4.2

We are now looking for the extremal chemical trees for the Randić index .

Lower bound

We have used the system AutoGraphiX on chemical trees, iterating on the number n of nodes and the number n_1 of pending nodes, minimizing the function

$$f_1(T) = Ra(T) - Ra(Co(n,n_1)) \tag{4.19}$$

where T is a chemical tree (if T is not chemical, we know by Theorem 4.4 that the minimal trees are the comets).

Presumably minimal values for $f_1(T)$, depending of n and n_1 and found by AutoGraphiX, are represented in Figure 4.2, for $5 \leq n \leq 21$ and $2 \leq n_1 \leq \min(n-2,12)$. Examples of the corresponding extremal trees are shown in Figure 4.3, for $21 \leq n \leq 22$ and $3 \leq n_1 \leq 12$.

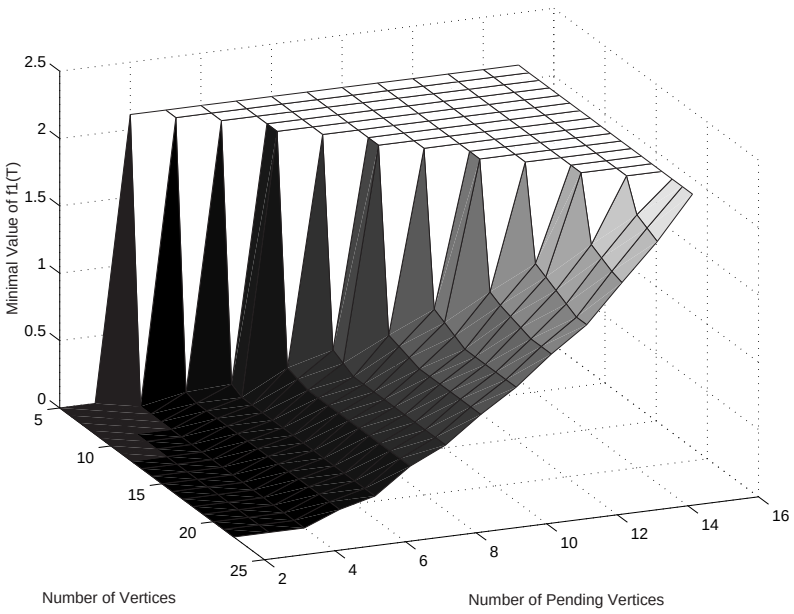


Figure 4.2: Minimal values for $f_1(T)$ found by AutoGraphiX

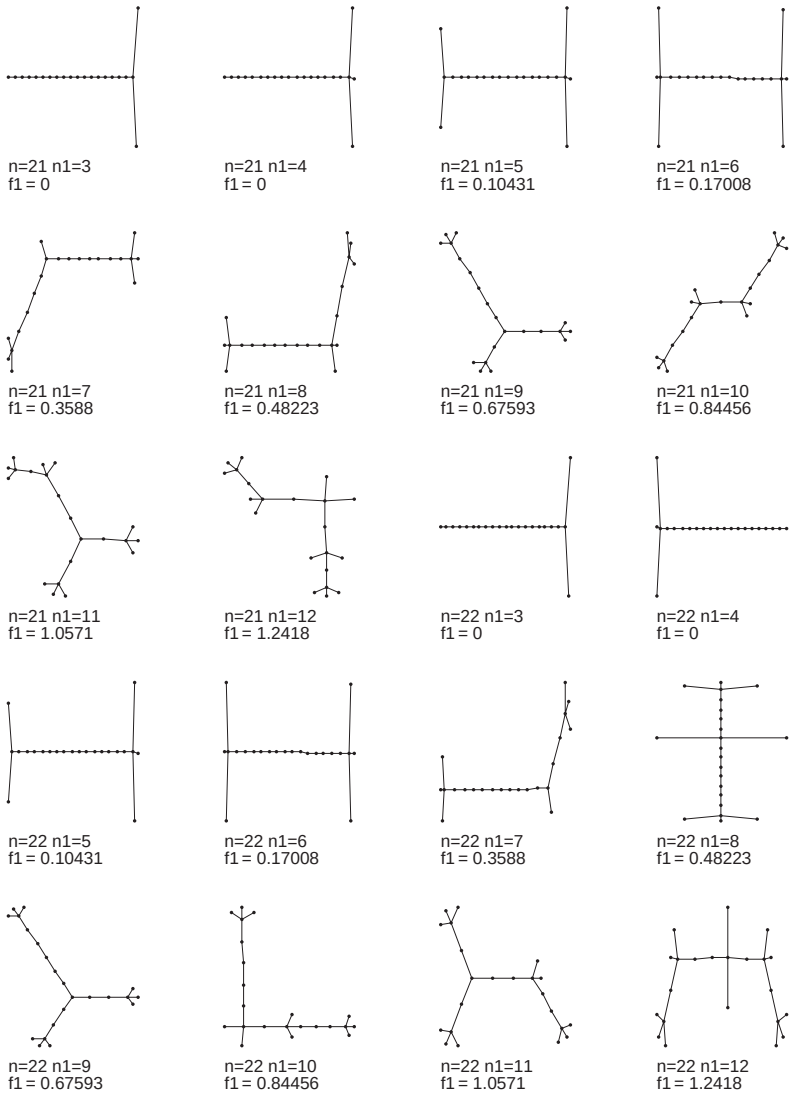


Figure 4.3: Chemical trees with presumably minimal values for $f_1(T)$ found by AutoGraphiX

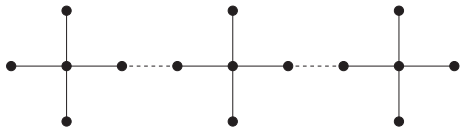


Figure 4.4: Structure of $L_e(n,n_1)$

An examination of these trees reveals some special structures. When $n_1 \leq 4$, these graphs are comets, which are minimal by Theorem 4.4. When $n_1 \geq 5$ the comets are no more chemical trees, and AutoGraphiX gives trees with specific structures, depending of the parity of n_1. When n_1 is even, we denote them $L_e(n,n_1)$ and when n_1 is odd $L_o(n,n_1)$.

The structure of $L_e(n,n_1)$ is depicted in Figure 4.4. They are composed by subgraphs which are the stars S_5 and these stars are connected by paths (the dotted lines in the figure), for which the lengths can be zero. The configuration is complete if $n \geq 9$ and $6 \leq n_1 \leq \lfloor \frac{n+3}{2} \rfloor$ (and even).

We can compute $Ra(L_e(n,n_1))$. We see in Figure 4.4 that $L_e(n,n_1)$ is formed by $\frac{n_1-2}{2}$ stars S_5. This chemical tree has n_1 pending edges of weight $\frac{1}{2}$, n_1-4 edges between the centers of the stars and the paths joining these stars of weight $\frac{1}{\sqrt{8}}$. The other edges are on the paths between the stars and have a weight of $\frac{1}{2}$ also. As any tree has $n-1$ edges, there are $n-2n_1+3$ edges of this type (and so n_1 has to be less than or equal to $\lfloor \frac{n+3}{2} \rfloor$).

So,

$$\begin{aligned} Ra(L_e(n,n_1)) &= \tfrac{n_1}{2} + \tfrac{n_1-4}{2\sqrt{2}} + \tfrac{n-2n_1+3}{2} \\ &= \tfrac{n}{2} + \tfrac{n_1}{2}\left(\tfrac{1}{\sqrt{2}} - 1\right) + \tfrac{3}{2} - \sqrt{2}. \end{aligned}$$

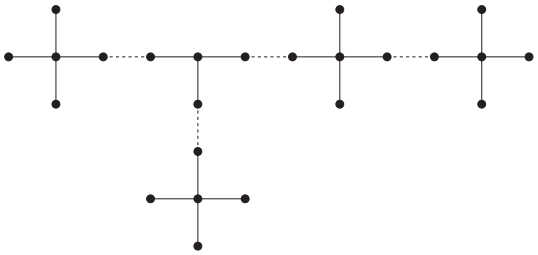


Figure 4.5: Structure of $L_o(n,n_1)$

The structure of $L_o(n,n_1)$ is depicted in Figure 4.5. This structure is similar to $L_e(n,n_1)$ but in this case, we have not only 2, but 3 *pending stars* S_5, i.e., stars connected to the remainder of the tree through one node only. That implies that one node has degree 3. The dotted lines in Figure 4.5 are paths, for which the length can be zero. This configuration is not always complete : we should have $n \geq 16$ and $9 \leq n_1 \leq \lfloor \frac{n+2}{2} \rfloor$ (and odd).

We can compute $Ra(L_o(n,n_1))$. We see in Figure 4.5 that this chemical tree has n_1 pending edges of weight $\frac{1}{2}$, n_1-6 edges between the centers of the stars S_5 and the paths joining these stars

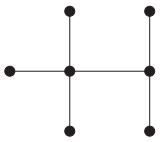


Figure 4.6: The only chemical tree with 7 nodes and 5 pending ones

of weight $\frac{1}{\sqrt{8}}$, and 3 edges adjacent to the node of degree 3 of weight $\frac{1}{\sqrt{6}}$. The other edges are on the paths represented by the dotted lines, and have a weight of $\frac{1}{2}$ also. The tree containing $n-1$ edges, there are $n-2n_1+2$ edges of this type (and so n_1 has to be less than or equal to $\lfloor \frac{n+2}{2} \rfloor$).

So,

$$\begin{aligned} Ra(L_o(n,n_1)) &= \tfrac{n_1}{2} + \tfrac{n_1-6}{2\sqrt{2}} + \tfrac{3}{\sqrt{6}} + \tfrac{n-2n_1+2}{2} \\ &= \tfrac{n}{2} + \tfrac{n_1}{2}\left(\tfrac{1}{\sqrt{2}} - 1\right) + 1 + \tfrac{3}{\sqrt{6}} - \tfrac{3}{\sqrt{2}}. \end{aligned}$$

Dropping the parity condition on n_1, we see that

$$Ra(L_e(n,n_1)) < Ra(L_o(n,n_1)),$$

because

$$\frac{3}{2} - \sqrt{2} \simeq 0.0858 < 1 + \frac{3}{\sqrt{6}} - \frac{3}{\sqrt{2}} \simeq 0.1034.$$

All these observations on the results of AutoGraphiX suggest the following theorem : $L_e(n,n_1)$ are the minimal chemical trees, when $n_1 \geq 5$.

Theorem 4.5. *Let T be a chemical tree of order n with $n_1 \geq 5$ pending nodes. Then*

$$Ra(T) \geq \frac{n}{2} + \frac{n_1}{2}\left(\frac{1}{\sqrt{2}} - 1\right) + \frac{3}{2} - \sqrt{2}, \tag{4.20}$$

with equality if and only if n_1 is even and T is isomorphic to $L_e(n,n_1)$.

Proof. First we apply induction, as in the proof of Theorem 4.4, to prove inequality (4.20).

By assumption $n_1 \geq 5$, and as $n_1 \leq n-1$, we have $n \geq 6$. If $n = 6$, we have the star S_6 which is not chemical. If $n = 7$, n_1 is equal to 5 or 6. The latter case is again a star (S_7) which is not chemical. It is easy to check that the only chemical tree T^* with 7 nodes and 5 pending nodes is that one of Figure 4.6. In this case, condition (4.20) is satisfied as

$$Ra(T^*) = \frac{3}{2} + \frac{2}{\sqrt{3}} + \frac{1}{\sqrt{12}} \simeq 2.9434 \geq \frac{7}{2} + \frac{5}{2}\left(\frac{1}{\sqrt{2}} - 1\right) + \frac{3}{2} - \sqrt{2} \simeq 2.8536.$$

So let us assume that $n \geq 8$ and (4.20) holds for all smaller values of n.

Let $\{v_1, v_2\}$ be a pending edge of a tree T with n nodes (with $d_{v_1} = 1$). If $d_{v_2} = 1$, T has 2 nodes which contradicts $n \geq 8$.

If $d_{v_2} = 2$,

$$Ra(T) - Ra(T - v_1v_2) = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{d_{v_3}}}\left(\frac{1}{\sqrt{2}} - 1\right),$$

where d_{v_3} is the degree of v_3, the node adjacent to v_2 other than v_1. Thus,

$$Ra(T) - Ra(T - v_1v_2) \geq \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}} - 1\right) = \frac{1}{2},$$

as $d_{v_3} \geq 2$.

As $d_{v_2} = 2$, the tree $T - v_1v_2$ has n_1 pending nodes and $n-1$ nodes.

By induction,

$$Ra(T) \geq \frac{n-1}{2} + \frac{n_1}{2}\left(\frac{1}{\sqrt{2}} - 1\right) + \frac{3}{2} - \sqrt{2} + \frac{1}{2},$$

and (4.20) holds.

If $d_{v_2} \geq 3$, the tree $T - v_1v_2$ has $n_1 - 1$ pending nodes and $n-1$ nodes. Then, by induction,

$$Ra(T - v_1v_2) \geq \frac{n}{2} + \frac{n_1}{2}\left(\frac{1}{\sqrt{2}} - 1\right) + \frac{3}{2} - \sqrt{2} + \frac{1}{2} - \frac{1}{\sqrt{2}}. \tag{4.21}$$

We are in the same conditions as for Lemma 4.1, so we can use Equation (4.13), replacing n_1 by 4 hence $\Delta \leq 4$ in chemical trees. We obtain

$$Ra(T) - Ra(T - v_1v_2) \geq 2 - \sqrt{3} + \left(\frac{1}{\sqrt{2}} - 1\right)\left(\frac{1}{2} - \frac{1}{\sqrt{3}}\right) =$$

$$\frac{3}{2} - \frac{2}{\sqrt{3}} - \frac{1}{\sqrt{6}} + \frac{1}{\sqrt{8}}. \tag{4.22}$$

Combining expressions (4.21) and (4.22), we have

$$Ra(T) \geq \frac{n}{2} + \frac{n_1}{2}\left(\frac{1}{\sqrt{2}} - 1\right) + \frac{3}{2} - \sqrt{2} + \frac{1}{2} - \frac{1}{\sqrt{2}} + \frac{3}{2} - \frac{2}{\sqrt{3}} - \frac{1}{\sqrt{6}} + \frac{1}{\sqrt{8}},$$

and as $2 - \frac{1}{\sqrt{2}} - \frac{2}{\sqrt{3}} - \frac{1}{\sqrt{6}} + \frac{1}{\sqrt{8}}$ is positive, expression (4.20) holds again.

We have shown above that $Ra(L_e(n,n_1))$ is equal to the lower bound of (4.20). We now prove that trees $L_e(n,n_1)$ are the only ones to reach that bound.

For all chemical tree we have by definition of the Randić index that

$$Ra(T) = \frac{x_{12}}{\sqrt{2}} + \frac{x_{13}}{\sqrt{3}} + \frac{x_{14}}{2} + \frac{x_{22}}{2} + \frac{x_{23}}{\sqrt{6}} + \frac{x_{24}}{2\sqrt{2}} + \frac{x_{33}}{3} + \frac{x_{34}}{2\sqrt{3}} + \frac{x_{44}}{4}, \tag{4.23}$$

where x_{ij} is the number of edges connecting a node of degree i with a node of degree j.

Denoting by n_i the number of nodes of degree i, the following six (linearly independent) relations are obeyed (see [52]):

$$\begin{aligned} x_{12}+x_{13}+x_{14} &= n_1 && (4.24)\\ x_{12}+2x_{22}+x_{23}+x_{24} &= 2n_2 && (4.25)\\ x_{13}+x_{23}+2x_{33}+x_{34} &= 3n_3 && (4.26)\\ x_{14}+x_{24}+x_{34}+2x_{44} &= 4n_4 && (4.27)\\ n_1+2n_2+3n_3+4n_4 &= 2(n-1) && (4.28)\\ n_1+n_2+n_3+n_4 &= n. && (4.29) \end{aligned}$$

If $Ra(T)$ is equal to the lower bound of (4.20), there are no terms of the form $\frac{1}{\sqrt{3}}$, $\frac{1}{\sqrt{6}}$, $\frac{1}{2\sqrt{3}}$, and so $x_{13}=x_{23}=x_{34}=0$ for T. This implies that $x_{33}=0$ because otherwise all nodes of degree 3 are connected together, and only to nodes of degree 3, which is impossible in a tree, so $n_3=0$.

We get the relations

$$\begin{aligned} x_{12}+x_{14} &= n_1 && (4.30)\\ x_{12}+2x_{22}+x_{24} &= 2n_2 && (4.31)\\ x_{14}+x_{24}+2x_{44} &= 4n_4 && (4.32)\\ n_1+2n_2+4n_4 &= 2(n-1) && (4.33)\\ n_2 &= n-n_1-n_4. && (4.34) \end{aligned}$$

Replacing n_2 in (4.33) by (4.34) gives

$$n_4=\frac{n_1-2}{2}; \tag{4.35}$$

we can replace n_4 by this expression in (4.34) to obtain :

$$n_2=n-\frac{3n_1}{2}+1. \tag{4.36}$$

Eliminating x_{14} by (4.30) and n_4 by (4.35) in (4.32) yields :

$$x_{24}=n_1-4-2x_{44}+x_{12}. \tag{4.37}$$

Then replacing x_{24} by (4.37) and n_2 by (4.36) in (4.31) gives :

$$x_{22}=n-2n_1+3+x_{44}-x_{12}. \tag{4.38}$$

We rewrite the Randić index (4.23), replacing x_{13}, x_{23}, x_{33} and x_{34} by 0 :

$$Ra(T)=\frac{x_{12}}{\sqrt{2}}+\frac{x_{14}}{2}+\frac{x_{22}}{2}+\frac{x_{24}}{2\sqrt{2}}+\frac{x_{44}}{4}.$$

Using Equations (4.31), (4.37) and (4.38) to replace x_{14}, x_{24} and x_{22} respectively, we can write this index with the unknowns n, n_1, x_{12} and x_{44}. We get

$$Ra(T) = \frac{x_{12}}{\sqrt{2}} + \frac{n_1}{2} - \frac{x_{12}}{2} + \frac{x_{44}}{2} - n_1 + \frac{n}{2} + \frac{3}{2} - \frac{x_{12}}{2} + \frac{n_1}{2\sqrt{2}} - \frac{2}{\sqrt{2}} - \frac{x_{44}}{\sqrt{2}} + \frac{x_{12}}{2\sqrt{2}} + \frac{x_{44}}{4}.$$

So,

$$Ra(T) = \frac{n}{2} + \frac{n_1}{2}\left(\frac{1}{\sqrt{2}} - 1\right) + \frac{3}{2} - \sqrt{2} + x_{12}\left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{8}} - 1\right) + x_{44}\left(\frac{3}{4} - \frac{1}{\sqrt{2}}\right).$$

As $\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{8}} - 1 \simeq 0.0607$ and $\frac{3}{4} - \frac{1}{\sqrt{2}} \simeq 0.0429$ are strictly positive, we have to put $x_{12} = x_{44} = 0$, otherwise the lower bound will not be obtained.

In short, with all the information above we have minimal graphs with $x_{12} = x_{13} = x_{23} = x_{33} = x_{34} = x_{44} = 0$, $x_{14} = n_1$, $x_{22} = n - 2n_1 + 3$ and $x_{24} = n_1 - 4$. By Equation (4.35) we know that the number of nodes of degree 4 is $\frac{n_1-2}{2}$, forming the stars S_5. Moreover, there is no node of degree 3, all pending nodes are connected to a node of degree 4, and the other nodes (of degree 2) form paths connecting the stars S_5. Because $n_4 = \frac{n_1-2}{2}$ has to be integer, n_1 has to be even. This is the configuration of $L_e(n, n_1)$. □

This theorem is illustrated in Figure 4.7 (when $n = 22$), where the circles represent the value of the minimal chemical trees obtained by AutoGraphiX, the dotted line is the lower bound of Theorem 4.2 and the black line is the new lower bound proposed in Theorem 4.5. Observe that the bound is attained for four values of n_1, all even.

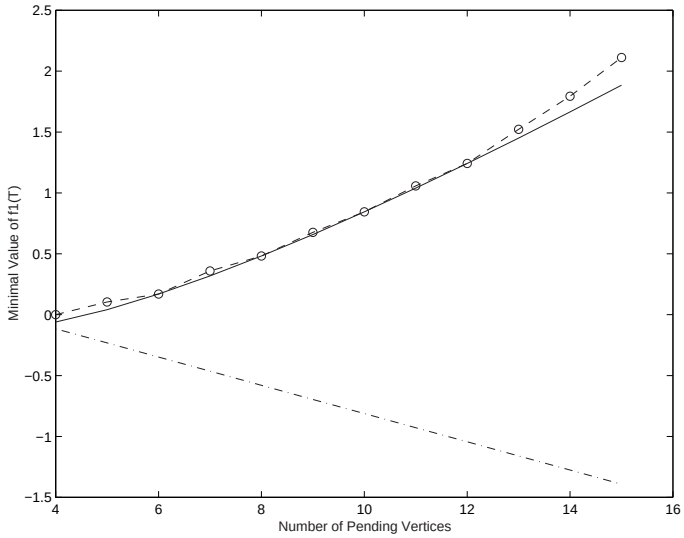


Figure 4.7: Minimal values for $f_1(T)$, $n = 22$

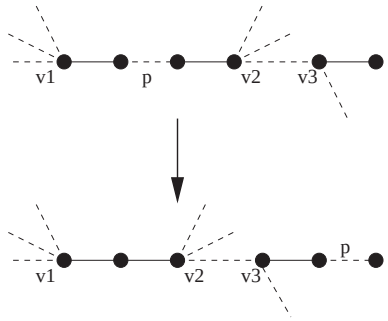


Figure 4.8: Moving the p inner edges of the path

Upper bound

We first introduce a new concept. The *ramification subgraph* of a tree T is the subgraph induced by the nodes of degree greater than or equal to 3.

Theorem 4.6. *The ramification subgraph of a chemical tree T is a tree if $Ra(T)$ is maximal.*

Proof. If $ram(T) = 1$, by definition of the ramification index, T has only one node of degree 3 forming the ramification subgraph, which is a tree.

If $ram(T) \geq 1$ and the ramification subgraph of T is not a tree, we have at least two nodes of T of degree 3 or 4, say v_1 and v_2 that are connected by a path of length at least 2. Suppose that the number p of inner edges on this path is positive, we note T' the chemical tree obtained by moving these p edges to a pending node (adjacent to a node v_3) as shown in Figure 4.8. It is easy to see that

$$Ra(T') - Ra(T) = \frac{1}{\sqrt{2}} - \frac{1}{2} + \frac{1}{\sqrt{d_{v_3}}}\left(\frac{1}{\sqrt{2}} - 1\right),$$

as $\frac{1}{\sqrt{2}} - 1$ is negative, the previous expression is minimal when d_{v_3} is minimal, i.e., $d_{v_3} = 2$. Thus

$$Ra(T') - Ra(T) \geq \frac{1}{\sqrt{2}} - \frac{1}{2} + \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}} - 1\right) = 0.$$

If $p \geq 1$ and $Ra(T)$ is maximal, then we can apply this transformation on T until a chemical tree T^* is obtained such that $ram(T) = ram(T^*)$, $Ra(T) = Ra(T^*)$ and $p = 0$. So we can suppose that $p = 0$. We can move one of the two edge between v_1 and v_2, say e, to a pending node, as shown in Figure 4.9. We note T'' the chemical tree obtained by this operation. By definition of the Randić index ,

$$Ra(T'') - Ra(T) = \frac{1}{\sqrt{d_{v_3}}}\left(\frac{1}{\sqrt{2}} - 1\right) + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{d_{v_1} d_{v_2}}} - \frac{1}{\sqrt{2d_{v_1}}} - \frac{1}{\sqrt{2d_{v_2}}},$$

as $\frac{1}{\sqrt{2}} - 1$ is negative, this expression is minimal when d_{v_3} is minimal, i.e., $d_{v_3} = 2$. Moreover as d_{v_1} and d_{v_2} are equal to 3 or 4, there are three different cases : the two nodes have a degree 3; the

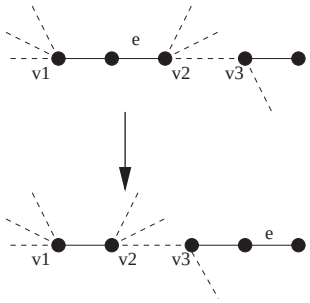


Figure 4.9: Moving edge e

two nodes have a degree 4; one node has a degree 3 and the other a degree 4. It is easy to check that the difference between $Ra(T'')$ and $Ra(T)$ is minimal when $d_{v_1} = d_{v_2} = 3$. Thus,

$$Ra(T'') - Ra(T) \geq \frac{1}{2} + \frac{1}{3} - \frac{2}{\sqrt{6}} \simeq 0.0168,$$

and T is not maximal for the Randić index . The result follows. □

Again, we have applied the system AutoGraphiX on chemical trees, iterating on the number n of nodes and the number n_1 of pending nodes. We maximize the following objective function. Let T be a chemical tree,

$$f_2(T) = Ra(T) - Ra(P_n). \tag{4.39}$$

The maximal values found by AutoGraphiX for $f_2(T)$ are represented in Figure 4.10, for $5 \leq n \leq 24$ and $2 \leq n_1 \leq \min(n-2, 14)$. We see that a plane can be taken for an upper bound, except if $n_1 = 2$, which is an easy case, i.e., the paths.

Examples of extremal graphs are shown in Figure 4.11, for $21 \leq n \leq 22$ and $3 \leq n_1 \leq 12$. An examination of the corresponding extremal trees reveals again a special structure. We denote $U(n,n_1)$ these specific chemicals trees. The trees $U(n,n_1)$ have a subgraph of $n_1 - 2$ nodes of degree 3 which is a tree; we note this node set V_3. In Figure 4.12, the nodes of V_3 are on a path, but it can be different. All these nodes are connected to another node of V_3, or to a path of length at least 2. The number of paths adjacent to the nodes of V_3 is $|V_3| + 2$, and the number of nodes of degree 2 is $n - 2n_1 + 2$.

This configuration is complete if $n \geq 7$ and $3 \leq n_1 \leq \lfloor \frac{n+2}{3} \rfloor$, and that explains that we do not have an upper plane, going through many points in Figure 4.10 if the case $n_1 = 2$ is included.

We can compute $Ra(U(n,n_1))$. We see in Figure 4.12 (p. 74) that $U(n,n_1)$ has n_1 pending edges of weight $\frac{1}{\sqrt{2}}$, n_1 edges connecting the paths and the nodes of V_3 of weight $\frac{1}{\sqrt{6}}$, $n_1 - 3$ edges joining the nodes of V_3 of weight $\frac{1}{3}$. The $n - 3n_1 + 2$ other edges are the inner edges of the paths (the dotted lines) of weight $\frac{1}{2}$ (and thus $n_1 \leq \lfloor \frac{n+2}{3} \rfloor$). So

$$Ra(U(n,n_1)) = \frac{n}{2} + n_1 \left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{6}} - \frac{7}{6} \right). \tag{4.40}$$

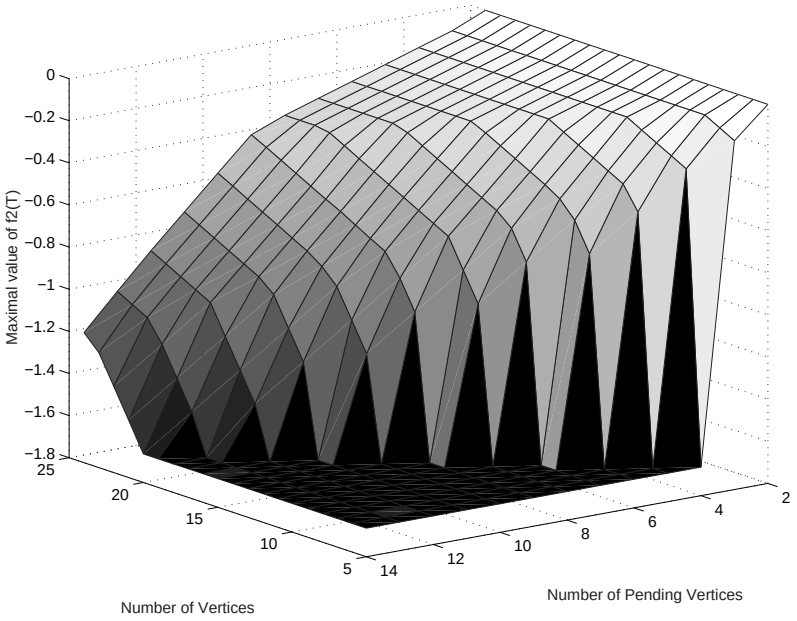


Figure 4.10: Maximal values for $f_2(T)$ found by AutoGraphiX

In order to prove that the right-hand side of Equation (4.40) is an upper bound on the Randić index of chemical trees, with fixed numbers of nodes n and pending nodes n_1, we first introduce four lemmata.

Lemma 4.2. *Let T be a chemical tree with n nodes and n_1 pending nodes. If $Ra(T)$ is maximal, then either T has no node of degree 4, or T has no edge joining two nodes of degree 2.*

Proof. By contradiction. Suppose that $n_4 > 0$, $x_{22} > 0$ and $Ra(T)$ is maximal. Let v be a node of T of degree 4 and $e = \{x, y\}$ be an edge of T such that $d_x = d_y = 2$. We note the 4 nodes adjacent to v by a, b, c and d. One can always choose v such that

$$d_a \leq 3,\ d_b \leq 3,\ d_c \leq 3 \text{ and } d_d \leq 4. \tag{4.41}$$

Indeed, the subgraph H, induced by the nodes of T of degree 4, is a forest. Thus, there always exists a node v in H with a degree ≤ 1. That means that v respect the conditions (4.41) in G.

We note T' the chemical tree obtained by applying the tranformation on T depicted in Figure 4.13. Remark that $n(T) = n(T')$ and $n_1(T) = n_1(T')$. We can then compute

$$Ra(T') - Ra(T) = \left(\frac{1}{\sqrt{d_a}} + \frac{1}{\sqrt{d_b}} + \frac{1}{\sqrt{d_c}} + \frac{1}{\sqrt{d_d}}\right)\left(\frac{1}{\sqrt{3}} - \frac{1}{2}\right) + \frac{1}{3} - \frac{1}{2}.$$

As $\left(\frac{1}{\sqrt{3}} - \frac{1}{2}\right)$ is positive, this expression is minimal when d_a, d_b, d_c and d_d are maximal, i.e., when $d_a = d_b = d_c = 3$ and $d_d = 4$ by the conditions (4.41). It follows that $Ra(T)$ is not maximal as

$$Ra(T') - Ra(T) \geq \left(\frac{3}{\sqrt{3}} + \frac{1}{2}\right)\left(\frac{1}{\sqrt{3}} - \frac{1}{2}\right) + \frac{1}{3} - \frac{1}{2} \simeq 0.0059831 > 0.$$

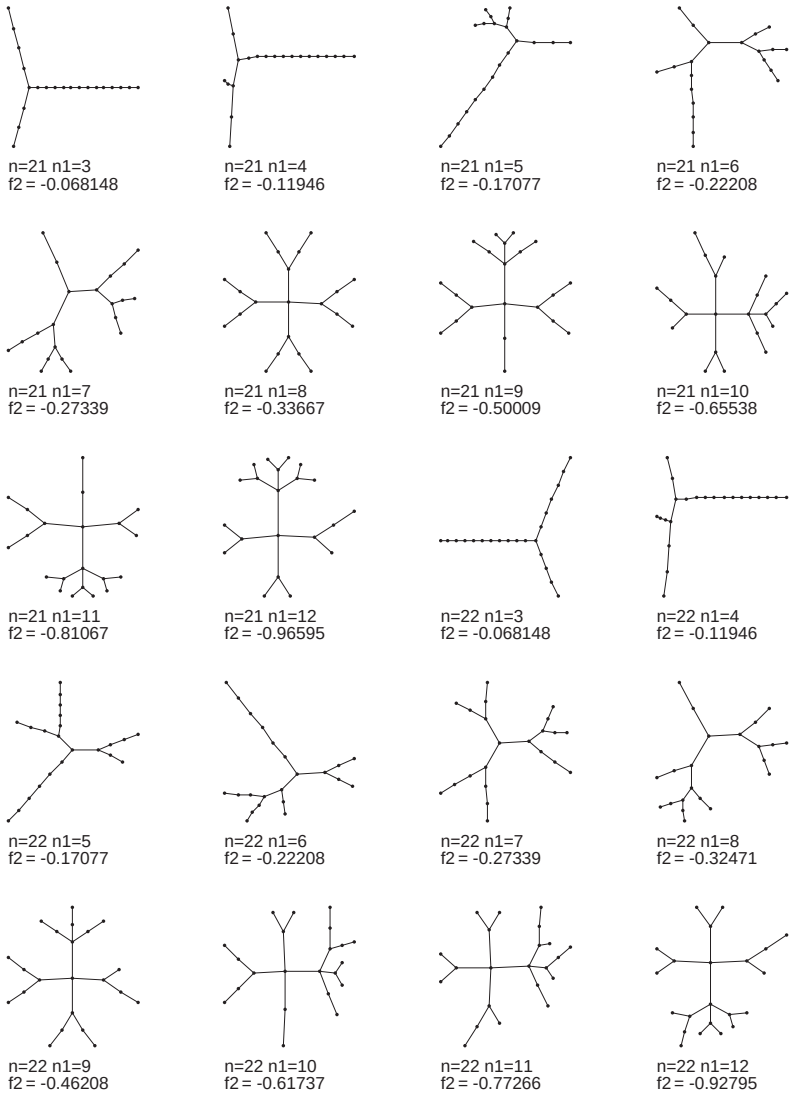


Figure 4.11: Graphs with presumably maximal values for $f_2(T)$ found by AutoGraphiX

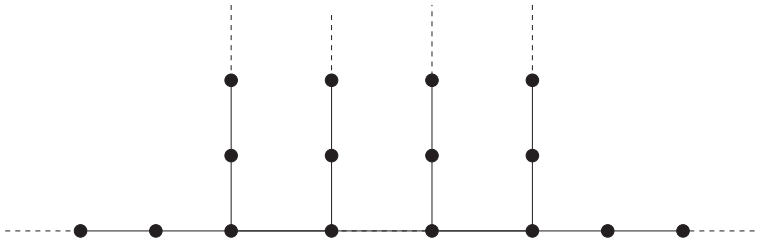


Figure 4.12: Structure of $U(n,n_1)$

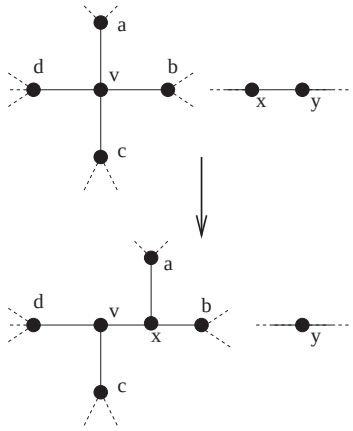


Figure 4.13: Transformation used in the proof of Lemma 4.2

It is easy to check that $Ra(T')-Ra(T)>0$ even if x is one of the 4 nodes adjacent to v. □

Lemma 4.3. *Let T be a chemical tree with n nodes, n_1 pending nodes and with maximal Randić index. Then,*

$$x_{22} = n-3n_1+n_4+2+x_{13}+x_{14}.$$

Proof. As $Ra(T)$ is maximal and by Theorem 4.6, we know that the nodes of T of degree 3 or 4 form a induced subtree H of order n_3+n_4. There are exactly n_1 paths appended to nodes of H. Each path is composed

- either by an edge counted in x_{12}, inner edges counted in x_{22}, and one edge counted in x_{23} or x_{24},
- or by a pending edge counted in x_{13} or x_{14}.

Other edges, forming tree H are counted in x_{33}, x_{34} and x_{44}. It follows that if $Ra(T)$ is maximal, then

$$x_{23}+x_{24} = x_{12} = n_1-x_{13}-x_{14}. \tag{4.42}$$

From the observation above, we count the number of edges joining two nodes of degree 2 :

$$x_{22} = m - (n_3 + n_4 - 1) - n_1 - x_{23} - x_{24}, \tag{4.43}$$

and as T is a tree,

$$x_{22} = n - n_3 - n_4 - n_1 - x_{23} - x_{24}. \tag{4.44}$$

By Equation (4.42), we can replace $-x_{23} - x_{24}$ by $x_{13} + x_{14} - n_1$:

$$x_{22} = n - n_3 - n_4 - 2n_1 + x_{13} + x_{14}. \tag{4.45}$$

From Equations (4.28) and (4.29) we get

$$n_3 = n_1 - 2 - 2n_4. \tag{4.46}$$

Substituting n_3 in (4.45) by (4.46) gives

$$x_{22} = n - 3n_1 + n_4 + 2 + x_{13} + x_{14}, \tag{4.47}$$

and the thesis is proved. □

Lemma 4.4. *Let T be a chemical tree with n nodes and n_1 pending nodes. If $Ra(T)$ is maximal, then either T has no edge joining one node of degree 1 and one node of degree 2, or T has no edge joining one node of degree 1 and one node of degree 4.*

Proof. By contradiction. Suppose that $x_{14} > 0$, $x_{12} > 0$ and $Ra(T)$ is maximal. Let $e = \{a,b\}$ be an edge of T such that $d_a = 1$, $d_b = 4$ and $f = \{c,d\}$ be an edge of T such that $d_c = 1$, $d_d = 2$. We note the 3 nodes $\neq a$ adjacent to b by w, x, y and the node $\neq c$ adjacent to d by z (see the first situtation in Figure 4.14).

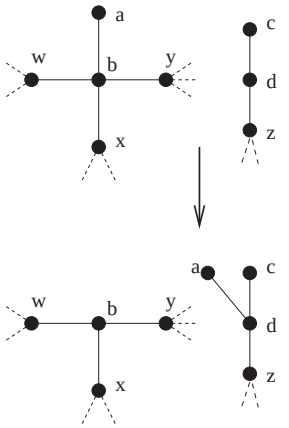


Figure 4.14: Transformation used in the proof of Lemma 4.4

By the same argument than the one used in the proof of Lemma 4.2, there exists a node v of degree 4 in T, such that three of its adjacent nodes have a degree ≤ 3 and one has a degree ≤ 4. It

is easy to see that the transformation which consists to exchange edge e and one branch adjacent to v does not change the value of $Ra(T)$. So, applying such transformation if needed, we can suppose that

$$d_w \leq 3,\ d_x \leq 3,\ \text{and } d_y \leq 4. \tag{4.48}$$

Since $Ra(T)$ is maximal, we can apply Lemma 4.2. As $n_4 \geq x_{14} > 0$, it follows that there are no edge in T joining two nodes of degree 2 and thus

$$d_z \geq 3. \tag{4.49}$$

We note T' the chemical tree obtained by applying the tranformation on T depicted in Figure 4.14. Remark that $n(T) = n(T')$ and $n_1(T) = n_1(T')$. We can then compute

$$Ra(T') - Ra(T) = \left(\frac{1}{\sqrt{d_w}} + \frac{1}{\sqrt{d_x}} + \frac{1}{\sqrt{d_y}}\right)\left(\frac{1}{\sqrt{3}} - \frac{1}{2}\right)$$

$$+\frac{1}{\sqrt{d_z}}\left(\frac{1}{\sqrt{3}} - \frac{1}{\sqrt{2}}\right) + \frac{2}{\sqrt{3}} - \frac{1}{\sqrt{2}} - \frac{1}{2}.$$

As $\left(\frac{1}{\sqrt{3}} - \frac{1}{2}\right)$ is positive, this expression is minimal when d_w, d_x and d_y are maximal, i.e., when $d_w = d_x = 3$ and $d_y = 4$ by the conditions (4.48). As $\left(\frac{1}{\sqrt{3}} - \frac{1}{\sqrt{2}}\right)$ is negative, the same expression is minimal when d_z is minimal, i.e., when $d_z = 3$ by condition (4.49). It follows that

$$Ra(T') - Ra(T) \geq \left(\frac{2}{\sqrt{3}} + \frac{1}{2}\right)\left(\frac{1}{\sqrt{3}} - \frac{1}{2}\right) + \frac{1}{\sqrt{3}}\left(\frac{1}{\sqrt{3}} - \frac{1}{\sqrt{2}}\right) + \frac{2}{\sqrt{3}} - \frac{1}{\sqrt{2}} - \frac{1}{2}$$

$$\simeq 0.00067 > 0.$$

This observation contradicts the hyptotheses as $Ra(T)$ is not maximal. □

Lemma 4.5. *Let T be a chemical tree with n nodes and n_1 pending nodes. If there are no node of degree 2 in T, then,*

$$Ra(T) \leq \frac{n}{2} + n_1\left(\frac{1}{\sqrt{3}} + \frac{1}{3}\right).$$

Proof. As $n_2 = 0$, we get from Equations (4.24) – (4.29) the following system of five linear equations :

$$\begin{aligned}
x_{13} + x_{14} &= n_1 && (4.50)\\
x_{13} + 2x_{33} + x_{34} &= 3n_3 && (4.51)\\
x_{14} + x_{34} + 2x_{44} &= 4n_4 && (4.52)\\
n_1 + 3n_3 + 4n_4 &= 2(n-1) && (4.53)\\
n_1 + n_3 + n_4 &= n. && (4.54)
\end{aligned}$$

Solving this system in the unknowns x_{13}, x_{33}, x_{44}, n_3 and n_4 leads to :

$$\begin{aligned} x_{13} &= n_1 - x_{14}, && (4.55)\\ x_{33} &= 3n - 5n_1 + \frac{1}{2}x_{14} - \frac{1}{2}x_{34} + 3, && (4.56)\\ x_{44} &= 4n_1 - 2n - \frac{1}{2}x_{14} - \frac{1}{2}x_{34} - 4, && (4.57)\\ n_3 &= 2n - 3n_1 + 2, && (4.58)\\ n_4 &= 2n_1 - n - 2. && (4.59) \end{aligned}$$

Substitution of the relations (4.55) – (4.57) into Equation (4.23) where $x_{12} = x_{22} = x_{23} = x_{24} = n_2 = 0$ yields

$$Ra(T) = \frac{n}{2} + n_1\left(\frac{1}{\sqrt{3}} + \frac{1}{3}\right) + x_{14}\left(\frac{13}{24} - \frac{1}{\sqrt{3}}\right) + x_{34}\left(\frac{1}{2\sqrt{3}} - \frac{7}{24}\right). \quad (4.60)$$

As $\left(\frac{13}{24} - \frac{1}{\sqrt{3}}\right) \simeq -0.03568$ and $\left(\frac{1}{2\sqrt{3}} - \frac{7}{24}\right) \simeq -0.00299$ are negative, the maximal value for $Ra(T)$ is obtained when $x_{14} = x_{34} = 0$ and

$$Ra(T) \leq \frac{n}{2} + n_1\left(\frac{1}{\sqrt{3}} + \frac{1}{3}\right).$$

□

We can now establish the upper bound on $Ra(T)$.

Theorem 4.7. *Let T be a chemical tree of order n and with $n_1 \geq 3$ pending nodes. Then,*

$$Ra(T) \leq \frac{n}{2} - a_0' n_1,$$

where $a_0' = \frac{7}{6} - \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{6}} \simeq 0.0513$, and with equality if and only if T is isomorphic to $U(n, n_1)$.

Proof. As in the proof of Theorem 4.5, we use expression (4.23) and relations (4.24) – (4.29). We will consider Equations (4.24) – (4.29) as a system of six linear equations in the unknowns x_{12}, x_{23}, x_{33}, n_2, n_3, n_4 and solve it. The solutions thus obtained will then depend on the remaining parameters, namely on : x_{13}, x_{14}, x_{22}, x_{24}, x_{34}, x_{44}, n_1 and n, because the graphs $U(n, n_1)$ have a majority of these latter parameters equal to zero.

The required solutions are :

$$\begin{aligned} x_{12} &= n_1 - x_{13} - x_{14}, && (4.61)\\ x_{23} &= 2n - 5n_1 + x_{13} + \frac{3}{2}x_{14} - 2x_{22} - \frac{1}{2}x_{24} + \frac{1}{2}x_{34} + x_{44} + 4, && (4.62)\\ x_{33} &= 4n_1 - n - x_{13} - \frac{3}{2}x_{14} + x_{22} - \frac{1}{2}x_{24} - \frac{3}{2}x_{34} - 2x_{44} - 5, && (4.63)\\ n_2 &= n - 2n_1 + \frac{1}{4}x_{14} + \frac{1}{4}x_{24} + \frac{1}{4}x_{34} + \frac{1}{2}x_{44} + 2, && (4.64) \end{aligned}$$

$$n_3 = n_1 - \frac{1}{2}x_{14} - \frac{1}{2}x_{24} - \frac{1}{2}x_{34} - x_{44} - 2, \tag{4.65}$$

and

$$n_4 = \frac{1}{4}x_{14} + \frac{1}{4}x_{24} + \frac{1}{4}x_{34} + \frac{1}{2}x_{44}. \tag{4.66}$$

Substitution of the relations (4.61) – (4.63) back into Equation (4.23) yields

$$Ra(T) = an + bn_1 + cx_{13} + dx_{14} + ex_{22} + fx_{24} + gx_{34} + hx_{44} + i, \tag{4.67}$$

where

$$\begin{aligned}
a &= -\frac{1}{3} + \frac{2}{\sqrt{6}} \simeq 0.4832, && (4.68)\\
b &= \frac{4}{3} + \frac{1}{\sqrt{2}} - \frac{5}{\sqrt{6}} \simeq -0.0008, && (4.69)\\
c &= -\frac{1}{3} - \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{6}} \simeq -0.0548, && (4.70)\\
d &= -\frac{1}{\sqrt{2}} + \frac{3}{2\sqrt{6}} \simeq -0.0947, && (4.71)\\
e &= \frac{5}{6} - \frac{2}{\sqrt{6}} \simeq 0.0168, && (4.72)\\
f &= -\frac{1}{6} + \frac{1}{2\sqrt{2}} - \frac{1}{2\sqrt{6}} \simeq -0.0172, && (4.73)\\
g &= -\frac{1}{2} + \frac{1}{2\sqrt{3}} + \frac{1}{2\sqrt{6}} \simeq -0.0072, && (4.74)\\
h &= -\frac{5}{12} + \frac{1}{\sqrt{6}} \simeq -0.0084, && (4.75)\\
i &= -\frac{5}{3} + \frac{4}{\sqrt{6}} \simeq -0.0337. && (4.76)
\end{aligned}$$

Suppose that $Ra(T)$ is maximal. We can thus apply Lemma 4.2, considering the two cases :

i) *If* $n_4 = 0$. In that case we have $x_{14} = x_{24} = x_{34} = x_{44} = n_4 = 0$ and $x_{22} = n - 3n_1 + 2 + x_{13}$ by Lemma 4.3. Substituting these values in Equation (4.67) gives, after simplification,

$$Ra(T) = \frac{n}{2} + n_1\left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{6}} - \frac{7}{6}\right) + x_{13}\left(\frac{1}{\sqrt{3}} - \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{6}} + \frac{1}{2}\right) \tag{4.77}$$

As $\left(\frac{1}{\sqrt{3}} - \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{6}} + \frac{1}{2}\right) \simeq -0.038005$ is negative, for fixed n, n_1, the values of $Ra(T)$ will be maximal if $x_{13} = 0$. We obtain from Equations (4.61) – (4.66) the solutions :

$$\begin{aligned}
x_{12} &= n_1, && (4.78)\\
x_{23} &= 2n - 5n_1 - 2x_{22} + 4, && (4.79)\\
x_{33} &= n_1 - 3, && (4.80)\\
n_2 &= n - 2n_1 + 2, && (4.81)\\
n_3 &= n_1 - 2, && (4.82)\\
n_4 &= 0. && (4.83)
\end{aligned}$$

It is easy to see that these values describe the graphs $U(n,n_1)$ and that the upper bound of the theorem is attained if and only if $x_{13}=0$.

ii) *If* $x_{22}=0$. In that case Lemma 4.3 gives

$$n_4 = 3n_1 - n - 2 - x_{13} - x_{14}. \tag{4.84}$$

We have that $\frac{h}{2}$ is greater than d, f and g. Thus, by Equation (4.67),

$$Ra(T) \leq an + bn_1 + cx_{13} + \frac{h}{2}(x_{14} + x_{24} + x_{34} + 2x_{44}) + i,$$

which is equivalent, by Equation (4.27), to

$$Ra(T) \leq an + bn_1 + cx_{13} + 2hn_4 + i.$$

Substituting n_4 by (4.84) leads, after simplification, to

$$Ra(T) \leq n(a-2h) + n_1(b+6h) + x_{13}(c-2h) + x_{14}(-2h) + (i-4h).$$

By Equations (4.68) – (4.76), we obtain the explicit form of this bound :

$$\begin{aligned} Ra(T) \leq \frac{n}{2} + n_1 &\left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{6}} - \frac{7}{6}\right) \\ &+ x_{13}\left(\frac{1}{2} + \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{6}}\right) + x_{14}\left(\frac{5}{6} - \frac{2}{\sqrt{6}}\right). \end{aligned} \tag{4.85}$$

We know by Lemma 4.4, and as $Ra(T)$ is maximal, that either $x_{12}=0$, or $x_{14}=0$.

a) If $x_{12}=0$, then one can check that T cannot contain any node of degree 2. Indeed, recall that $x_{22}=0$ and that Theorem 4.6 implies that edges counted in x_{23} or x_{24} have to be adjacent to pending paths. We can thus apply Lemma 4.5 and

$$Ra(T) \leq \frac{n}{2} + n_1\left(\frac{1}{\sqrt{3}} + \frac{1}{3}\right) < \frac{n}{2} + n_1\left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{6}} - \frac{7}{6}\right).$$

b) If $x_{14}=0$, then Equation (4.85) is maximal when $x_{13}=0$, as $\left(\frac{1}{2}+\frac{1}{\sqrt{3}}-\frac{1}{\sqrt{2}} \quad -\frac{1}{\sqrt{6}}\right)$ is negative. In this case, the bound of the theorem is attained. Remark that we get again $U(n,n_1)$ as extremal graphs (but in this case $x_{22}=0$). □

Figure 4.15 gives the results for $n=24$ where the circles represent the value of the maximal chemical trees obtained by AutoGraphiX, the dotted line is the upper bound of Theorem 4.2 and the black line is the new upper bound that we propose in Theorem 4.7. For this value of n the bound is attained for six values of n_1. Observe that the bound cannot be extended to paths without losing much of its precision (as measured by the number of values of n_1 for which it is tight). Such border effects appear to be common in graph theoretic problems studied by AutoGraphiX.

4.1.4 Correction of Theorem 4.3

Let G be a simple graph and $\kappa = \{v_1, v_2\}$ an edge in G. Then $G(\kappa)$ is the graph obtained from G by deleting κ. The following proposition is a correction of the Proposition 3.3 of [11].

Proposition 4.1. *Let G be a chemical graph and $\kappa = \{v_1, v_2\}$ be an edge of G with d_{v_1}, $d_{v_2} \geq 2$. Then*

$$Ra(G(\kappa)) - d_0' \leq Ra(G) \leq Ra(G(\kappa)) + b_0'$$

where $b_0' = (\sqrt{2}-1)/2 \simeq 0.2071$ and $d_0' = 2\sqrt{3} - \frac{13}{4} \simeq 0.2141$

Proof. By definition of the Randić index , we can check that, for all v_x, v_y such that $\{v_1, v_x\} \in E$, $v_x \neq v_2$ and $\{v_2, v_y\} \in E$, $v_y \neq v_1$,

$$\begin{aligned} Ra(G) - Ra(G(\kappa)) = \quad & \sum_{\{v_1,v_x\}} \frac{1}{\sqrt{d_{v_x}}} \left[\frac{1}{\sqrt{d_{v_1}}} - \frac{1}{\sqrt{d_{v_1}-1}} \right] + \\ & \sum_{\{v_2,v_y\}} \frac{1}{\sqrt{d_{v_y}}} \left[\frac{1}{\sqrt{d_{v_2}}} - \frac{1}{\sqrt{d_{v_2}-1}} \right] + \frac{1}{\sqrt{d_{v_1} d_{v_2}}}. \end{aligned}$$

The minimal value of this expression is reached when $d_{v_x} = 1$ for all $v_x \neq v_2$ adjacent to v_1 and $d_{v_y} = 1$ for all $v_y \neq v_1$ adjacent to v_2. The maximal value is reached when $d_{v_x} = 4$ for all $v_x \neq v_2$ adjacent to v_1 and $d_{v_y} = 4$ for all $v_y \neq v_1$ adjacent to v_2.

So,

$$\begin{aligned} & (d_{v_1}-1)\left[\frac{1}{\sqrt{d_{v_1}}} - \frac{1}{\sqrt{d_{v_1}-1}} \right] + (d_{v_2}-1)\left[\frac{1}{\sqrt{d_{v_2}}} - \frac{1}{\sqrt{d_{v_2}-1}} \right] + \frac{1}{\sqrt{d_{v_1} d_{v_2}}} \\ & \leq Ra(G) - Ra(G(\kappa)) \leq \end{aligned}$$

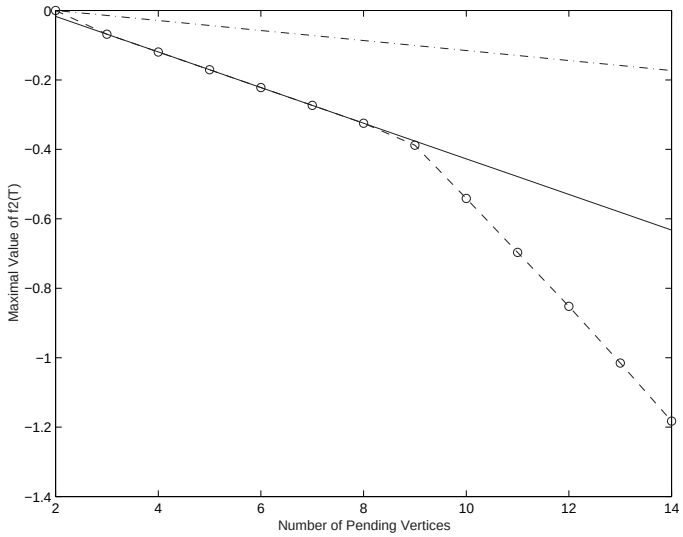


Figure 4.15: Maximal values for $f_2(T)$, $n = 24$

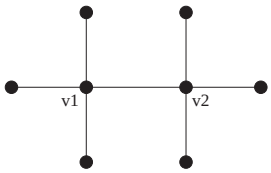


Figure 4.16: This graph reaches the Lower Bound of Proposition 4.1

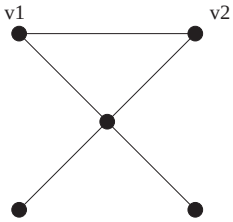


Figure 4.17: This graph reaches the Upper Bound of Proposition 4.1

$$\frac{1}{2}(d_{v_1}-1)\left[\frac{1}{\sqrt{d_{v_1}}}-\frac{1}{\sqrt{d_{v_1}-1}}\right]+\frac{1}{2}(d_{v_2}-1)\left[\frac{1}{\sqrt{d_{v_2}}}-\frac{1}{\sqrt{d_{v_2}-1}}\right]+\frac{1}{\sqrt{d_{v_1}d_{v_2}}}.$$

Next, we inspect all the possible cases for d_{v_1} and d_{v_2} ($\in\{2,3,4\}$ by assumption) for the lower bound and the upper bound. The minimum value for the difference $Ra(G)-Ra(G(\kappa))$ is reached when $d_{v_1}=d_{v_2}=4$ and is equal to $-d_0'$. The maximum value for this difference is reached when $d_{v_1}=d_{v_2}=2$ and is equal to b_0', and that shows the result. □

The constants b_0' and d_0' are not the same as the constants b_0 and d_0 of Theorem 4.3. We have constructed examples for which the new bounds are reached. Thus these bounds are best possible. The chemical graph of Figure 4.16 has a Randić index which reaches the lower bound of Proposition 4.1, when deleting the edge $\{v_1,v_2\}$.

In the same way, the graph of Figure 4.17 has a Randić index which reaches the upper bound.

Remark that these two graphs are counter-examples to Proposition 3.3 of [11].

We note that, by construction, the only possible graph which reaches the lower bound of Proposition 4.1 is the one described in Figure 4.16. This graph is a tree, and thus is not interesting for Theorem 4.3. This fact leads to a new proposition for chemicals graph with a cyclomatic number $\mu\geq 1$. The lower bound is better in this case.

Proposition 4.2. *Let G be a chemical graph for which the cyclomatic number $\mu(G)\geq 1$. Let $\kappa=\{v_1,v_2\}$ be an edge in a cycle of G. Then*

$$Ra(G(\kappa))-d_0^*\leq Ra(G)\leq Ra(G(\kappa))+b_0'$$

where $d_0^*=\frac{2}{\sqrt{6}}+\frac{4}{\sqrt{3}}-\frac{1}{\sqrt{2}}-\frac{9}{4}\simeq 0.1688$

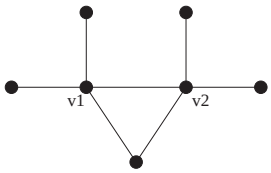


Figure 4.18: This graph reaches the Lower Bound of Proposition 4.2

Proof. Again, it is clear that

$$\begin{aligned} Ra(G) - Ra(G(\kappa)) = \; & \textstyle\sum_{v_1 - v_x \neq v_2} \frac{1}{\sqrt{d_{v_x}}} \left[\frac{1}{\sqrt{d_{v_1}}} - \frac{1}{\sqrt{d_{v_1}-1}} \right] + \\ & \textstyle\sum_{v_2 - v_y \neq v_1} \frac{1}{\sqrt{d_{v_y}}} \left[\frac{1}{\sqrt{d_{v_2}}} - \frac{1}{\sqrt{d_{v_2}-1}} \right] + \frac{1}{\sqrt{d_{v_1} d_{v_2}}} \end{aligned}$$

Like in the proof of the previous Proposition, the maximal value is reached when $d_{v_x} = 4$ ($v_1 - v_x \neq v_2$) and $d_{v_y} = 4$ ($v_2 - v_y \neq v_1$). But the minimal value is no longer reached when $d_{v_x} = d_{v_y} = 1$, because this leads to the special tree described in Figure 4.16. By hypothesis, G cannot be a tree because $\mu(G) \geq 1$. Moreover, the edge $\kappa = \{v_1, v_2\}$ must be on a cycle of G. So, there exist at least one node $v_x \neq v_2$, adjacent to v_1 and at least one node $v_y \neq v_1$, adjacent to v_2 for which $d_{v_x}, d_{v_y} \geq 2$, because v_x and v_y have to be connected by a path, different from $\{v_1, v_2\}$.

Hence

$$\begin{aligned} & (d_{v_1} - 2 + \frac{1}{\sqrt{2}}) \left[\frac{1}{\sqrt{d_{v_1}}} - \frac{1}{\sqrt{d_{v_1}-1}} \right] + \\ & (d_{v_2} - 2 + \frac{1}{\sqrt{2}}) \left[\frac{1}{\sqrt{d_{v_2}}} - \frac{1}{\sqrt{d_{v_2}-1}} \right] + \frac{1}{\sqrt{d_{v_1} d_{v_2}}} \\ & \leq Ra(G) - Ra(G(\kappa)) \end{aligned}$$

Again, by inspecting the possible values for d_{v_1} and d_{v_2} ($\in \{2,3,4\}$), we find d_0^* when $d_{v_1} = d_{v_2} = 4$. □

The chemical graph of Figure 4.18 (with $\mu(G) \geq 1$) has a Randić index which reaches the lower bound when deleting the edge $\{v_1, v_2\}$.

We can now give the following theorem, which is a correction of Theorem 4.3.

Theorem 4.8. *Let G be a chemical graph and T be a maximal subtree of G. Then*

$$Ra(T) - d_0^* \mu(G) \leq Ra(G) \leq Ra(T) + b_0' \mu(G). \tag{4.86}$$

In particular, if G has n nodes and T has n_1 pending nodes, then

$$\frac{n}{2} + c_0' n_1 + d_0'' + d_0^* \mu(G) \leq Ra(G) \leq \frac{n}{2} - a_0' n_1 + b_0' \mu(G). \tag{4.87}$$

where $c_0' = \frac{1}{2}(\frac{1}{\sqrt{2}} - 1) \simeq -0.1464$ and $d_0'' = \frac{3}{2} - \sqrt{2} \simeq 0.0858$.

Proof. If $\mu(G) = 0$, G is a tree, and assumption (4.86) is trivially verified because the only maximal subtree of a tree is the tree itself. If $\mu(G) = 1$, assumption (4.86) is correct by Proposition 4.2. By induction on $\mu(G)$ and by Proposition 4.2, assumption (4.86) still holds for $\mu(G) > 1$.

The second expression (4.87) is easily obtained when applying Theorem 4.5 to the lower bound of (4.86) and Theorem 4.7 to the upper bound of (4.86). □

4.2 Comparison of irregularity indices for chemical trees

A graph is said to be regular if all its nodes have the same degree. Otherwise, it is irregular. This, of course, also applies to chemical graphs. Regular graphs dealt with in chemical graph theory are mainly of degree 2 (molecular graphs of annulenes[1] and cycloalkanes[2]) and 3 (molecular graphs of fullerenes[3]). The interest towards regular molecular graphs has much increased after the experimental discovery of fullerenes and the elaboration of their theory [128].

In spite of this, the vast majority of graphs of interest to chemists is irregular. Clearly, some are more irregular than others. In this section, it is our intention to make this notion more precise, through some irregularity measure.

To this aim, we establish some properties of these irregularity measures, applicable to chemical graphs, chemical trees in particular. By doing this we reveal the close connection between the concept of chemical-graph-irregularity and branching. We show that some irregularity measures are, in fact, branching indices. Some, on the other hand, are unrelated to branching.

Next, according to Balaban [15], in order that a graph invariant be acceptable as a usable molecular structure descriptor (topological index) it must monotonically increase (or decrease) with the increase of the extent of branching, especially in the case of chemical trees. Thus, our analysis implies that some irregularity measures are, whereas some are not, expected to be of practical value in QSPR/QSAR studies (see Section 1.5).

Three main irregularity measures have been proposed in the literature:

(i) The *Collatz - Sinogowitz index* [67], defined for any graph $G = (V,E)$ with node set V and edge set E by

$$\mathrm{CS}(G) = \lambda_1 - \frac{2m}{n}, \tag{4.88}$$

where λ_1 is the index, m is the number of edges and n the number of nodes of G ($\frac{2m}{n}$ is the *average degree*);

[1] Annulenes are completely conjugated monocyclic hydrocarbons : they have the molecular formula C_nH_n (n is an even number) or C_nH_{n+1} (n is an odd number).

[2] The molecular formula of a cycloalkane is C_nH_{2n}.

[3] See definition p. 22.

(ii) the *variance of degrees* [27] defined for any graph $G = (V,E)$ by

$$\mathrm{VAR}(G) = \frac{1}{n}\sum_{i=1}^{n-1} n_i \left(i - \frac{2m}{n}\right)^2 \tag{4.89}$$

where n_i denotes the number of nodes of degree i for $i = 1,2,\ldots,n-1$;

(iii) the *irregularity* [1] (see Definition 1.8, p. 18).

For general graphs, it was conjectured in [67] that stars maximize the difference between λ_1 and $\frac{2m}{n}$. This conjecture was refuted in [85]. $\mathrm{CS}(G)$ and $\mathrm{VAR}(G)$ are compared in [27]. Tight upper bounds on $\mathrm{CS}(G)$ and $\mathrm{VAR}(G)$ expressed as functions of n and m for both connected and disconnected graphs are also given there. It is shown in [1] that

$$A(G) < \frac{4n^3}{27}$$

and that this bound can be approached arbitrarily closely. A tight upper bound on $A(G)$ expressed as a function of m and n is given below (see Theorem 5.1, Section 5.3, p. 125).

All three indices are non-negative and equal to 0 if and only if the graph G is regular.

More can be said if one restricts attention to some particular class of graphs, such as e.g., chemical graphs. In this section, we begin such a study by considering chemical trees, denoted here by T.

The system AutoGraphiX is first used to obtain extremal or near-extremal chemical trees for each of the three indices, taking n as parameter. This leads immediately to conjectures on the structure of such trees, from where analytical expressions for the maximum of $\mathrm{VAR}(T)$ and $A(T)$ are deduced. Then, these conjectures are proved.

The next three sections are devoted to results on $\mathrm{CS}(T)$, $\mathrm{VAR}(T)$, $A(T)$, together with their proofs (when available), respectively. Section 4.2.4 compares the indices.

4.2.1 Collatz-Sinogowitz Index

Extremal graphs

It is shown in [189] that $\mathrm{CS}(T)$ is minimal for paths. This automatically determines the chemical tree with minimal CS-value. Chemical trees with maximum $\mathrm{CS}(T)$ have recently been investigated, on the basis of a complete enumeration of all chemical trees with $n \leq 21$ [122]. The use of AutoGraphiX to find chemical trees with maximum and near maximum $\mathrm{CS}(T)$ for $n \leq 29$ (see Figure 4.19) fully corroborated these results.

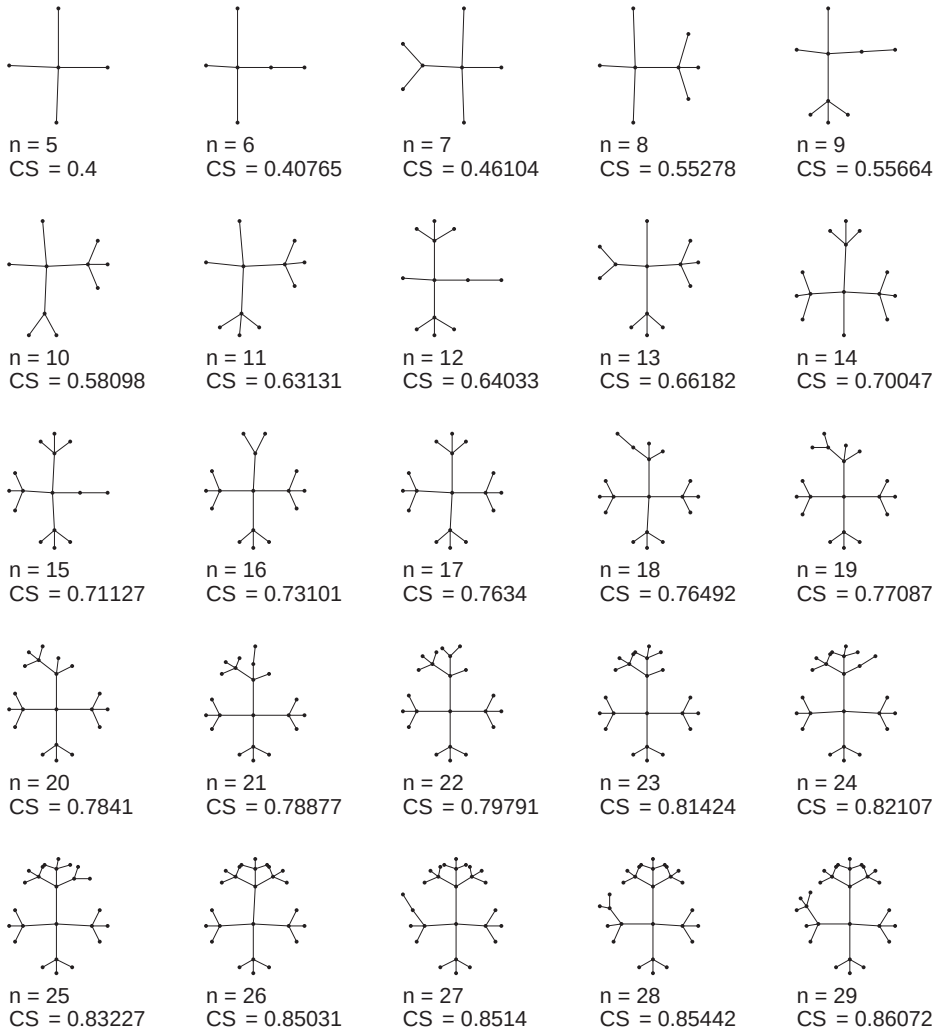


Figure 4.19: Extremal graphs found by AutoGraphiX for $\mathrm{CS}(T)$

Conjectures

Extremal chemical trees found for $n \leq 29$, belong to the class of *dendrimers*[4]. In what follows we refer to them as the D_n-dendrimers. These are defined as follows: a first node, 1, at level 0 is connected to up to 4 nodes 2, 3, 4, 5 at level 1 and exactly that number if $n \geq 5$. Then each node at level k in order of increasing indices is connected to 3 new nodes at level $k+1$, for $k = 1, 2, \ldots$ until nodes are exhausted (see Figure 4.20).

However, no mathematical proof that these trees are extremal with regard to $\mathrm{CS}(T)$ for all n is as yet available. So we reiterate

Conjecture 4.1. [122] *The chemical trees with maximum Collatz - Sinogowitz index are the D_n-dendrimers, and are unique for each n.*

[4]The pictures of the cover and the part's covers are representations of dendrimers.

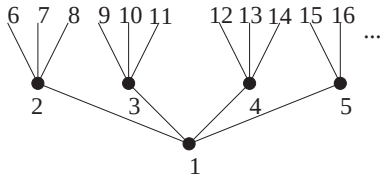


Figure 4.20: Construction of dendrimers, which are extremal graphs for $\mathrm{CS}(T)$

Several surprising results about D_n-dendrimers have been recently found [123] and are presented in [122]:

(i) they are chemical trees with minimum *Wiener index* or sum of distances between pairs of nodes (and hence with minimum average distance). For a review on the Wiener index of trees and an exhaustive bibliography see [98];

(ii) they are chemical trees with minimum *largest Laplacian eigenvalue* [154];

(iii) it follows from the characterization of chemical trees with minimum *Randic index* $Ra(T)$ (see Definition 1.15, p. 22), given in [52] that they have minimum Randic index if $n = 3k+2$ for $k = 1,2,\ldots$

Moreover, it is conjectured that for these latter values of n, dendrimers D_n also have minimum *energy* E (see Definition 1.16, p. 22) [140], as well as minimum *Hosoya index* Z, where

$$\mathrm{Z} = \mathrm{Z}(T) = \sum_{k\geq 0} m(T,k) \tag{4.90}$$

and the $m(T,k)$ are the numbers of k-matchings of T (see [151] pp. 127 – 134).

An analytical expression for $\mathrm{CS}(D_n)$ as a function of n is also missing at present. However, some insight may be obtained from its graphical representation, given in Figure 4.21. The curve of maximum $\mathrm{CS}(T)$ first increases rapidly, with kinks corresponding to successive sets of 3 values of n, then goes through a maximum and decreases slightly. A zoom on the values for $n = 45$ to 65, separated according to $n \pmod 3$, is presented in Figure 4.22. This leads to

Conjecture 4.2. *For all chemical trees T,*

$$CS(T) \leq CS(D_{56}) \simeq 0.748665.$$

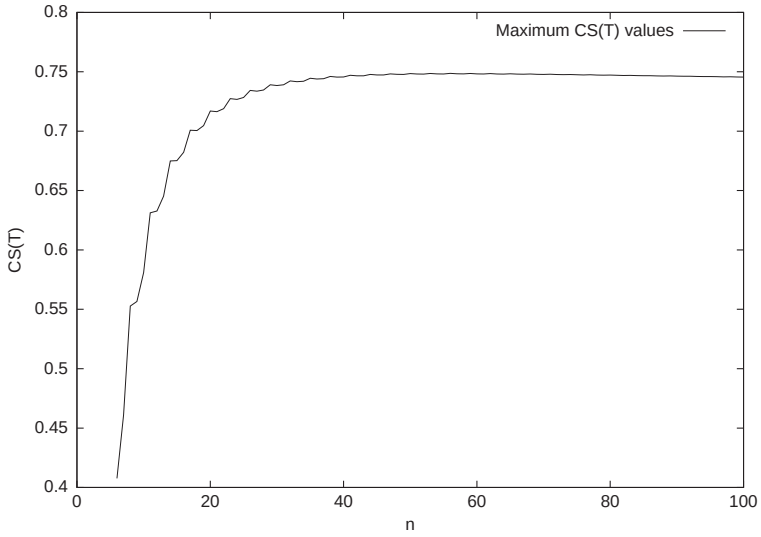


Figure 4.21: Bounds on $CS(T)$ for a chemical tree T of order n

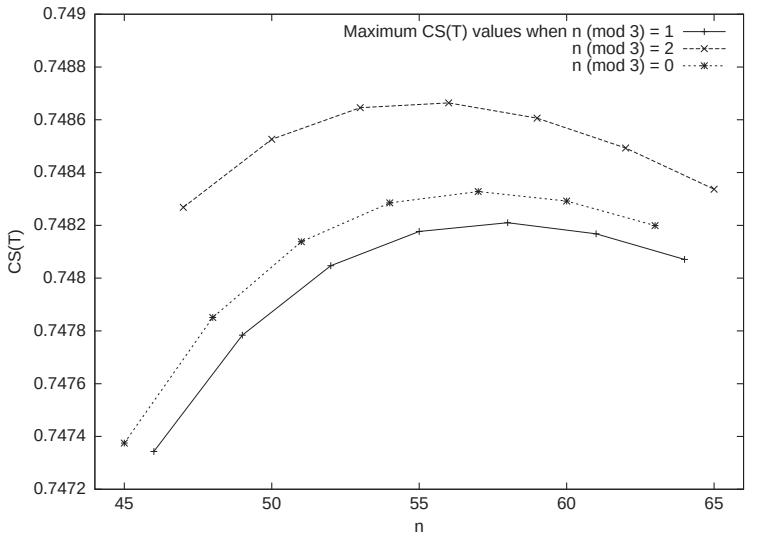


Figure 4.22: Bounds on $CS(T)$ for a chemical tree T of order $n \in [45,65]$

4.2.2 Variance of degrees

Extremal graphs

The extremal chemical trees found by AutoGraphiX for the variance of degrees are presented in Figure 4.23. These trees belong to 3 families and contain (as for the CS index) nodes of degree 1 and 4 as well as one node of degree 2 when $n(\mathrm{mod}\ 3) = 0$ and one node of degree 3 when $n(\mathrm{mod}\ 3) = 1$.

As $2n_4 + n_3 = n_1 - 2$ for all chemical trees, one can show that increasing n_4 and n_1 increases

the variance of degrees more than increasing n_3 or n_2. This shows that trees which are conjectured to be extremal for CS are also extremal for the variance of degrees.

However, they are a subset of them, because any tree having the same node degree distribution as D_n has the same VAR-value as D_n. Transformations which keep the distribution of degrees constant are easy to make. Recall that a *caterpillar* is a tree composed of a path and pending edges appended to some or all of its nodes. It thus appears that the following are representative elements of the class of n-node chemical trees with maximum VAR:

(i) caterpillars with all inner nodes of degree 4 when $n(\text{mod } 3) = 2$;

(ii) as in (i) with one edge added at an end node of a longest path when $n(\text{mod } 3) = 0$;

(iii) as in (i) with two edges added at an end node of a longest path when $n(\text{mod } 3) = 1$.

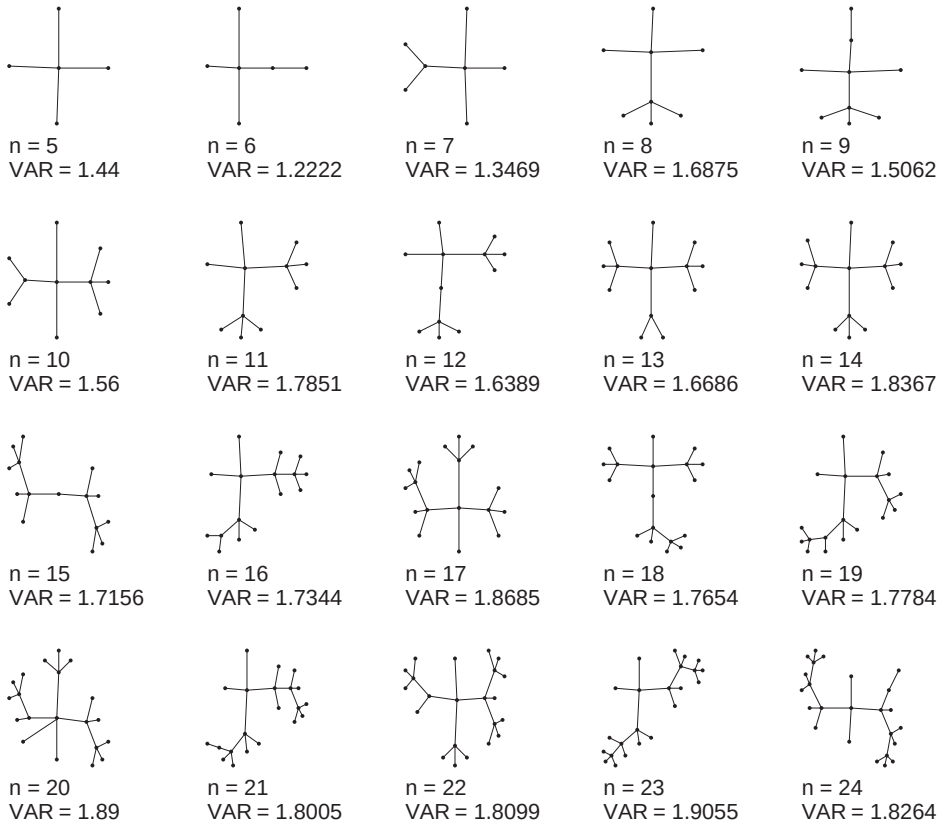


Figure 4.23: Extremal graphs found by AutoGraphiX for VAR(T)

In fact, in the trees just described, all nodes of degree 4 are on a path. Other extremal graphs for VAR(T) can be obtained easily: for family (i) we have to form a chemical tree with n_4 nodes and add the pending edges to obtain a degree 4 for each and every one of these nodes. A similar construction can be done for families (ii) and (iii) where we first have to choose a node which will

be of degree 2 or 3 respectively. This leads to the following characterization of extremal trees with maximum variance of degrees.

Theorem 4.9. ***Extremal graphs for VAR of family (i).*** *In this case $n(mod\ 3) = 2$. These are the trees with $n_1 = (2n+2)/3$, $n_2 = n_3 = 0$, $n_4 = (n-2)/3$. There are k different extremal chemical trees of order n where k is the number of chemical trees of order $\frac{n-2}{3}$ (the chemical trees with n_4 nodes for which we have to add the pending edges).*

Extremal graphs for VAR of family (ii). *In this case $n(mod\ 3) = 0$. These are the trees with $n_1 = 2n/3$, $n_2 = 1$, $n_3 = 0$, $n_4 = (n-3)/3$. There are $\sum_{i=1}^{l} m_i$ different extremal chemical trees of order n where l is the number of chemical tree of order $\frac{n}{3}$ and m_i the number of possibilities for choosing the node of degree 2, up to symmetry, for each of the l chemical trees (i.e., a node of degree 1 or 2 in the graph of order $\frac{n}{3}$).*

Extremal graphs for VAR of family (iii). *In this case $n(mod\ 3) = 1$. These are the trees with $n_1 = (2n+1)/3$, $n_2 = 0$, $n_3 = 1$, $n_4 = (n-4)/3$. A similar construction as for family (ii) applies, but here we choose a node of degree 3, among those of degree 1, 2 or 3 in a chemical tree of order $\frac{n-1}{3}$.*

The validity of this theorem lies in the fact that all extremal graphs of a family have the same variance of degrees and Theorem 4.10 will insure that these values are extremal.

For example if $n = 14$ (family (i)), the $k = 2$ different extremal graphs, based on the star S_4 and on the path P_4, are shown in Figure 4.24.

If $n = 12$ (family (ii)), $l = 2$ (S_4 and P_4), $m_1 = 1$ for the star and $m_2 = 2$ for the path. Such a construction is shown in Figure 4.25, and for $n = 13$ (family (iii)) the construction of the extremal graphs is represented in Figure 4.26.

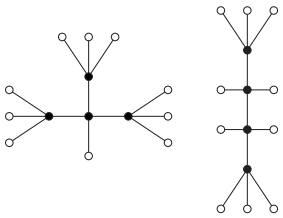


Figure 4.24: All extremal graphs for VAR when $n = 14$

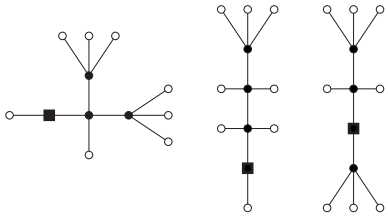


Figure 4.25: All extremal graphs for VAR when $n = 12$

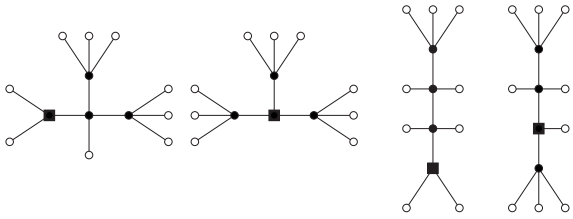


Figure 4.26: All extremal graphs for VAR when $n = 13$

Bounds

Theorem 4.10. *For all chemical trees T_n with $n \geq 2$ nodes,*

$$\begin{aligned} VAR(T_n) &\leq \tfrac{1}{n^2}\left[2n^2-2n-4\right] \quad \text{if } n(mod\ 3)=2, \\ &\leq \tfrac{1}{n^2}\left[2n^2-4n-4\right] \quad \text{otherwise,} \end{aligned}$$

and the bounds are sharp for all such n.

Proof. A simple rearrangement of the definition of VAR(G) in (4.89) yields

$$\mathrm{VAR}(G) = (M_1)/n - (2m/n)^2 \tag{4.91}$$

where $M_1 = \sum_{i=1}^{n-1} i^2 n_i$ is the so-called *first Zagreb index* [93, 145, 152, 153, 201, 226, 235]. In case of a chemical tree T,

$$\mathrm{VAR}(T) = (M_1)/n - [2(n-1)/n]^2 \tag{4.92}$$

and

$$M_1 = n_1 + 4n_2 + 9n_3 + 16n_4 \tag{4.93}$$

For all chemical trees,

$$n_1 + n_2 + n_3 + n_4 = n \tag{4.94}$$

$$n_1 + 2n_2 + 3n_3 + 4n_4 = 2(n-1) \tag{4.95}$$

Solve these equations in n_1 and n_4 and substitute into (4.92) and (4.93). This yields

$$\mathrm{VAR}(T) = (6n - 10 - 2n_2 - 2n_3)/n - [2(n-1)/n]^2 \tag{4.96}$$

Clearly, for a fixed value of n, VAR(T) will be maximal if $n_2 = n_3 = 0$ (if this is possible), or if n_2 and n_3 are as close to zero as possible.

If $n(\mathrm{mod}\ 3) = 2$ it is possible to construct chemical trees (that are not unique) with $n_2 = n_3 = 0$. If $n(\mathrm{mod}\ 3) = 0$ it is possible to construct chemical trees (that are not unique) with $n_2 = 1$, $n_3 = 0$. If $n(\mathrm{mod}\ 3) = 1$ it is possible to construct chemical trees (that are not unique) with $n_2 = 1$, $n_3 = 0$. Substituting these choices of n_2 and n_3 into (4.96) results in Theorem 4.10. □

Figure 4.27 shows the curve of values defined in Theorem 4.10.

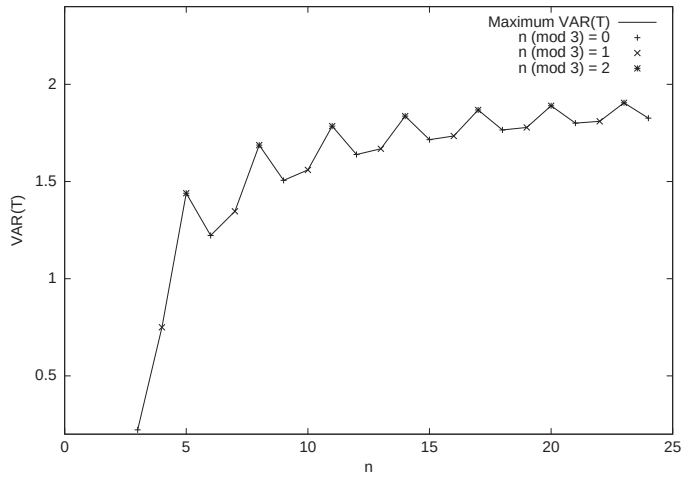


Figure 4.27: Bounds on $\mathrm{VAR}(T)$ for a chemical tree T of order n

4.2.3 Irregularity

Extremal graphs

The extremal chemical trees found by AutoGraphiX for the irregularity are presented in Figure 4.28. These trees have nodes of degree 1 and 4, a single node of degree 3 if $n(\mathrm{mod}\,4) = 3$, one node of degree 2 if $n(\mathrm{mod}\,4) = 2$ as well as, for all of them $\lceil \frac{n}{4} \rceil - 2$ additional nodes of degree 2. Note that there is an exception, when $n = 15$ and $A(T) = 34$, in the graphs presented in Figure 4.28. However, one can construct a chemical tree which respects the above description and for which $A(T) = 34$ (see below).

These trees are not unique. There are four classes, having the following representative elements which are caterpillars:

(i) caterpillars with nodes of degree $1,4,2,4,2,\ldots,4,1$ along the longest path;

(ii) – (iv) as in (i) with 1 (resp. 2,3) additional pending edges incident with an end node of the (or a) longest path.

Bounds

A sharp lower bound on $A(T)$ for all trees is readily obtained:

Theorem 4.11. *For any tree T,*

$$A(T) \geq \max_{v \in V} d_v(d_v - 1) \tag{4.97}$$

the bound being sharp if and only if T is homeomorphic to a star.

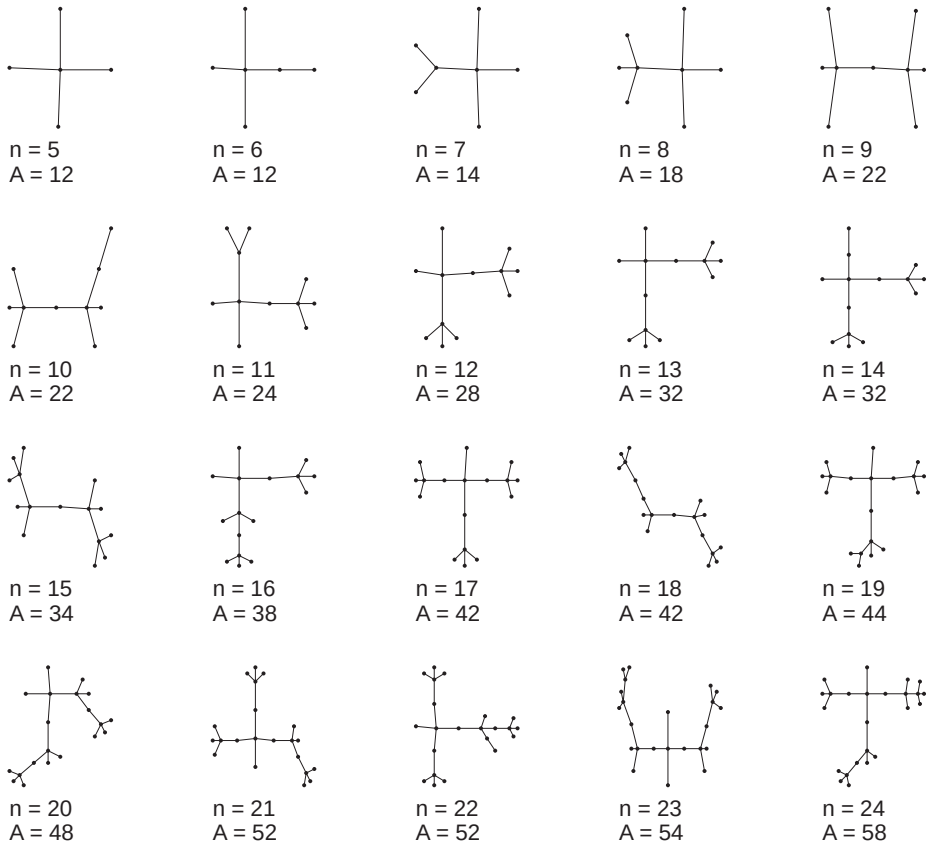


Figure 4.28: Extremal graphs found by AutoGraphiX for $A(T)$

Proof. Consider the node w with maximum degree; then there are d_w disjoint paths from w to a pendant node l_k for $k = 1,2,\ldots,d_w$. The sum of contributions to $A(T)$ of edges along each of these paths is at least $d_w - 1$ with equality if and only if the degrees of successive nodes along the path from w to l_k decrease monotonically (not necessarily strictly). This yields the lower bound. Moreover, if there is a node of degree > 2 different from w along one such path, there is at least one edge contributing positively to $A(T)$ and not in the union of these paths. Hence T is either a star or a tree homeomorphic to a star, i.e., such that removal of nodes v of degree $d_v = 2$ along any path, and merging the corresponding incident edges $\{r,v\}$ and $\{v,s\}$, gives a star. □

A sharp upper bound is given by:

Theorem 4.12. *For any chemical tree T with n nodes,*

$$A(T) \leq \begin{cases} \frac{5n-1}{2} & \text{if } n(mod\ 4) = 1, \text{ and} \\ \\ \frac{5n-n(mod\ 4)+4}{2} & \text{otherwise.} \end{cases}$$

Moreover this bound is sharp for all $n \geq 5$.

Proof. We use again a proof technique based on linear programming and first introduced in [52] (see also [150, 124] for other applications, and the proofs of Theorems 4.5 and 4.7).

Let x_{ij} be the number of edges with end nodes of degree i and j. Then,

$$A(T) = x_{12} + 2x_{13} + 3x_{14} + x_{23} + 2x_{24} + x_{34}. \tag{4.98}$$

We first solve the system of six linear equations (4.24) – (4.29) which are satisfied by all chemical trees, with unknowns $x_{14}, x_{24}, n_1, n_2, n_3, n_4$. That gives:

$$x_{14} = \frac{n+3}{2} - \frac{3x_{12}}{2} - \frac{7x_{13}}{6} - \frac{x_{22}}{2} - \frac{x_{23}}{6} + \frac{x_{33}}{6} + \frac{x_{34}}{3} + \frac{x_{44}}{2} \tag{4.99}$$

$$x_{24} = \frac{n-5}{2} + \frac{x_{12}}{2} + \frac{x_{13}}{6} - \frac{x_{22}}{2} - \frac{5x_{23}}{6} - \frac{7x_{33}}{6} - \frac{4x_{34}}{3} - \frac{3x_{44}}{2} \tag{4.100}$$

$$n_1 = \frac{n+3}{2} - \frac{x_{12}}{2} - \frac{x_{13}}{6} - \frac{x_{22}}{2} - \frac{x_{23}}{6} + \frac{x_{33}}{6} + \frac{x_{34}}{3} + \frac{x_{44}}{2} \tag{4.101}$$

$$n_2 = \frac{n-5}{4} + \frac{3x_{12}}{4} + \frac{x_{13}}{12} + \frac{3x_{22}}{4} + \frac{x_{23}}{12} - \frac{7x_{33}}{12} - \frac{2x_{34}}{3} - \frac{3x_{44}}{4} \tag{4.102}$$

$$n_3 = \frac{x_{13}}{3} + \frac{x_{23}}{3} + \frac{2x_{33}}{3} + \frac{x_{34}}{3} \tag{4.103}$$

$$n_4 = \frac{n-1}{4} - \frac{x_{12}}{4} - \frac{x_{13}}{4} - \frac{x_{22}}{4} - \frac{x_{23}}{4} - \frac{x_{33}}{4} + \frac{x_{44}}{4} \tag{4.104}$$

Replacing x_{14} and x_{24} by (4.99) and (4.100) respectively gives,

$$A(T) = \frac{5n-1}{2} - \frac{5x_{12}}{2} - \frac{7x_{13}}{6} - \frac{5x_{22}}{2} - \frac{7x_{23}}{6} - \frac{11x_{33}}{6} - \frac{2x_{34}}{3} - \frac{3x_{44}}{2}, \tag{4.105}$$

which is maximal for a fixed number of nodes when the values x_{12}, x_{13}, x_{22}, x_{23}, x_{33}, x_{34}, and x_{44} are equal to zero. However, in the case of n-node chemical trees the condition

$$x_{12} = x_{13} = x_{22} = x_{23} = x_{33} = x_{34} = x_{44} = 0 \tag{4.106}$$

can be satisfied only if $n \pmod 4 = 1$. Any chemical tree satisfying (4.106) has no nodes of degree 3, all its nodes of degree 2 are adjacent to two nodes of degree 4, and no two nodes of degree 4 are mutually adjacent. Denote the type of such n-node chemical trees by (i-1) (the type i when $n \pmod 4 = 1$).

Hence, if $n \pmod 4 = 1$, then for any n-node chemical tree, $A(T) \leq (5n-1)/2$. Equality is attained if and only if T is of type (i-1).

This completes the proof of Theorem 4.12 in the case $n \pmod 4 = 1$.

If $n \pmod 4 \neq 1$, then the conditions (4.106) cannot be satisfied by n-node chemical trees. Then, in order to find the chemical tree(s) with maximal A-values we have to find the values of

the parameters x_{12}, x_{13}, x_{22}, x_{23}, x_{33}, x_{34}, and x_{44} as close to zero as possible compatible to the existence of a chemical tree, i.e., for which the right hand sides of Eqs. (4.99) – (4.104) are integers and for which a graph exists. In the following discussion, we call x_{12}, x_{13}, x_{22}, x_{23}, x_{33}, x_{34}, and x_{44} the *observed parameters*. Looking at (4.105), we see that increasing by 1 any of the observed parameters subtracts a certain value from the maximal value of $A(T)$, expressed in sixths:

Observed parameter	x_{12}	x_{13}	x_{22}	x_{23}	x_{33}	x_{34}	x_{44}
Decrease on $A(T)$, in sixths	15	7	15	7	11	4	9

The gap between the maximum value of $A(T)$, which is $\frac{5n-1}{2}$ when all observed parameters are equal to zero, and the value of $A(T)$ when some of these parameters are positive will be called the *decrease*. The proof for the different cases is based on the examination of all possible increases of the values of the observed parameters, choosing those with minimal decrease. Fortunately, we can reduce the number of possibilities with some simple observations. AutoGraphiX gives us some (presumably) extremal graphs for each case. The decrease of these feasible solutions is a good upper bound on the minimal decrease. We can limit the enumeration taking only solutions with a decrease less or equal to this bound. Moreover Eq. (4.103) can considerably reduce the number of possibilities: if only one observed parameter is not equal to zero, the values of x_{13}, x_{23}, x_{33} or x_{34} must be multiples of 3; and in general $x_{13}+x_{23}+2x_{33}+x_{34}$ has to be a multiple of 3.

- Case $n \pmod 4 = 2$. A first lower bound on the maximal A-value is the graph obtained by AutoGraphiX and described above: in this case $x_{12} = 1$ and all other observed parameters are equal to zero. This solution has a decrease of 15 sixths. We will examine all possibilities for the observed parameters with a decrease less than or equal to 15 sixths, and for which Eq. (4.103) is not violated.

  First we consider the solutions where only one of the observed parameters is different from zero.

Non-zero param.	Decr. $(1/6)$	Comment
$x_{12} = 1$	15	Starting solution: type (i-2)
$x_{22} = 1$	15	Feasible: type (ii-2)
$x_{34} = 3$	12	Impossible in view of Eq. (4.99)
$x_{44} = 1$	9	Impossible in view of Eq. (4.104)

  We then examine the possible choices of two non-zero observed parameters with a decrease less than or equal to 15 sixths (and for which Eq. (4.103) is not violated). In the following table, RHS means the *right-hand sides* of the equations.

Non-zero param.	Decr. $(1/6)$	Comment
$x_{13} = 1$ and $x_{34} = 2$	15	Feasible: type (iii-2)
$x_{23} = 1$ and $x_{34} = 2$	15	Feasible: type (iv-2)
$x_{33} = 1$ and $x_{34} = 1$	15	RHS are integers but this solution is not feasible because by Eq. (4.103) we have that $n_3 = 1$ which is impossible because of $x_{33} = 1$

There are no possible choices of more than two non-zero observed parameters with decrease less than or equal to 15 sixths. Consequently, the four types (i-2), (ii-2), (iii-2) and (iv-2) of trees have the same A-value, equal to $(5n-1)/2-5/2$, and this is the greatest value that A can assume among n-node chemical trees.

This implies the validity of Theorem 4.12 for $n(\mathrm{mod}\,4) = 2$.

- Case $n(\mathrm{mod}\,4) = 3$. The analysis is analogous. We start again with a solution of AutoGraphiX where $x_{13} = 2$ and $x_{34} = 1$. This solution has a decrease of 18 sixths. We will examine the solutions with a decrease less than or equal to 18 sixths and for which Eq. (4.103) is not violated.

  First we consider the solutions where only one of the observed parameters is different from zero.

Non-zero param.	Decr. $(1/6)$	Comment
$x_{12} = 1$	15	Impossible in view of Eq. (4.99)
$x_{22} = 1$	15	Impossible in view of Eq. (4.99)
$x_{34} = 3$	12	Impossible in view of Eq. (4.102)
$x_{44} = 1$	9	Impossible in view of Eq. (4.99)
$x_{44} = 2$	18	Feasible: type (i-3)

We then examine the possible choices of two non-zero observed parameters.

Non-zero param.	Decr. $(1/6)$	Comment
$x_{13} = 1$ and $x_{33} = 1$	18	RHS are integers but this solution is not feasible because by Eq. (4.103) we have that $n_3 = 1$ which is impossible because of $x_{33} = 1$
$x_{13} = 1$ and $x_{34} = 2$	15	Impossible in view of Eq. (4.104)
$x_{13} = 2$ and $x_{34} = 1$	18	Feasible: type (ii-3) (Starting solution)
$x_{23} = 1$ and $x_{33} = 1$	18	RHS are integers but this solution is not feasible because by Eq. (4.103) we have that $n_3 = 1$ which is impossible because of $x_{33} = 1$
$x_{23} = 1$ and $x_{34} = 2$	15	Impossible in view of Eq. (4.104)
$x_{23} = 2$ and $x_{34} = 1$	18	Feasible: type (iii-3)
$x_{33} = 1$ and $x_{34} = 1$	15	Impossible in view of Eq. (4.104)

There is only one possible choice of exactly three non-zero observed parameters with a decrease less than or equal to 18 sixths: $x_{13} = 1$, $x_{23} = 1$ and $x_{34} = 1$. This solution is feasible and has a decrease of $18/6$ (type (iv-3)).

Consequently, the four types (i-3), (ii-3), (iii-3) and (iv-3) of trees, have the same A-value, equal to $(5n-1)/2-3$, and this is the greatest value that A can assume among n-node chemical trees.

This implies the validity of Theorem 4.12 for $n \pmod 4 = 3$.

- Case $n \pmod 4 = 0$. This case is easier because the starting solution (type (i-0)) given by AutoGraphiX has a decrease of 9 sixths ($x_{44} = 1$) and no other solution can be obtained with a decrease less than or equal to 9 sixths without violating Eq. (4.103). These trees have $A = (5n-1)/2-3/2$, verifying the respective part of Theorem 4.12. □

Figures 4.29 to 4.32 present all the types of extremal graphs presented in Theorem 4.12 when $n \in [12, 15]$. Figure 4.33 shows the curve of values defined in Theorem 4.12.

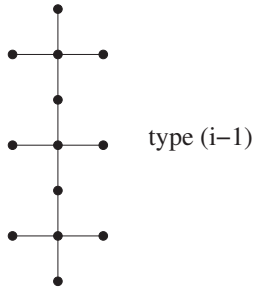


Figure 4.29: The unique extremal graph for A when $n = 13$

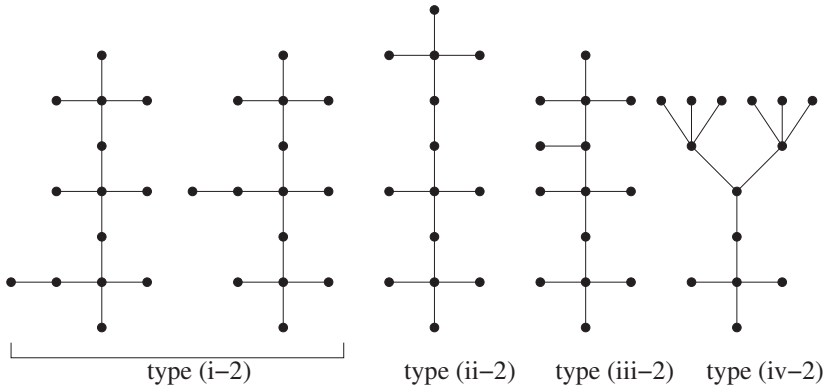


Figure 4.30: All extremal graphs for A when $n = 14$

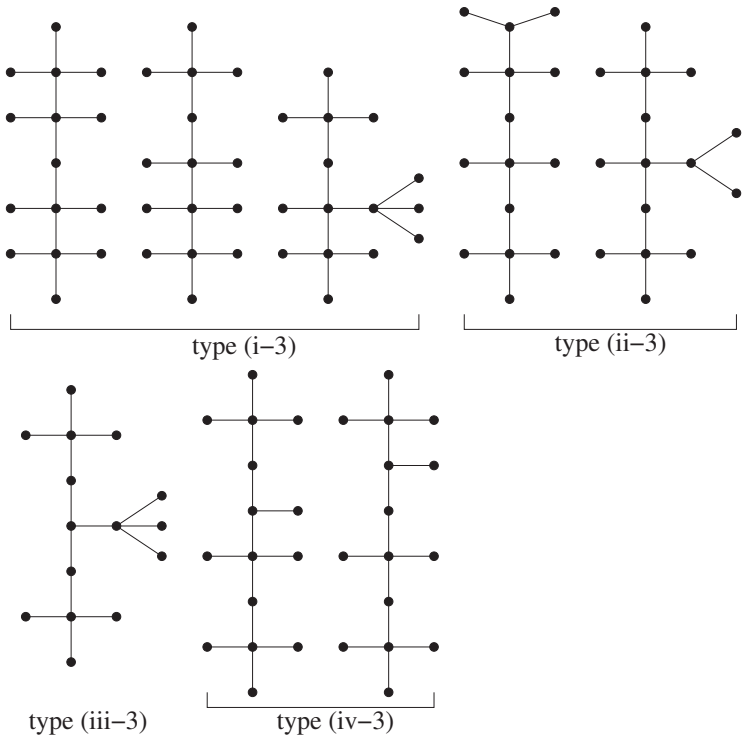


Figure 4.31: All extremal graphs for A when $n = 15$

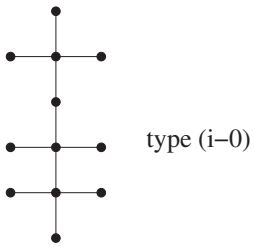


Figure 4.32: The unique extremal graph for A when $n = 12$

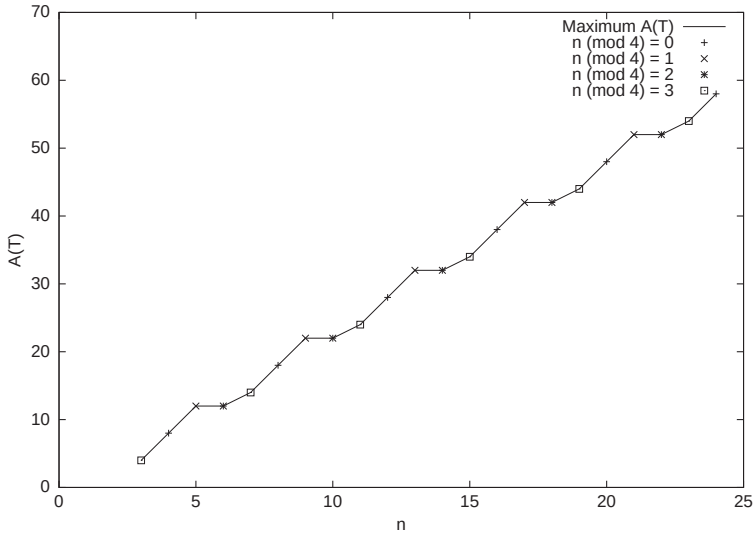


Figure 4.33: Bounds on $A(T)$ for a chemical tree T of order n

4.2.4 Comparison of the indices

In order to compare the irregularity indices, we studied how the extremal chemical trees for each index are evaluated according to the two other ones. The results of such observations are shown in Figures 4.34 to 4.39. In each figure we indicate the value of an irregularity index i_1 for chemical trees of order n (where $n \in [10, 18]$), extremal with regard to a different index i_2. The extremal graphs for given n and i_2 are indicated by +, connected by a line, and their i_2 values are represented on the y-axis. The maximum values for i_1 (computed from the previous results) are depicted by circles, linked by a dotted line. Some additional information is added, such as the location of the D_n-dendrimer, indicated by D (this does not imply that the dendrimers are the only extremal graphs with this value in the case of the variance of degrees).

Figures 4.34 and 4.35 show the A-values of the extremal graphs for $\mathrm{VAR}(T)$ and $\mathrm{CS}(T)$. One can see again that the D_n-dendrimers are extremal for $\mathrm{VAR}(T)$ and $\mathrm{CS}(T)$. An interesting fact is that the D_n-dendrimers are systematically on the minimum of each interval of values of $A(T)$ among the extremal graphs for $VAR(T)$ (Figure 4.34). This is due to the fact that degrees are non-increasing along any path from the root of a D_n-dendrimer to any of its pendant nodes. With some exceptions, maybe due to the fact that n is small, the extremal graphs do not attain the maximal A-value.

Figures 4.36 and 4.37 present the VAR-values of the chemical trees extremal with regard to $A(T)$ and $\mathrm{CS}(T)$. In Figure 4.36 the different types of extremal graphs for $A(T)$ are represented. Some of these families have the same variance of degrees, which is easy to compute from their characterization. No obvious link can be conjectured between $\mathrm{VAR}(T)$ and $A(T)$. Figure 4.37 reiterates that the D_n-dendrimers are extremal for $\mathrm{VAR}(T)$ and $\mathrm{CS}(T)$.

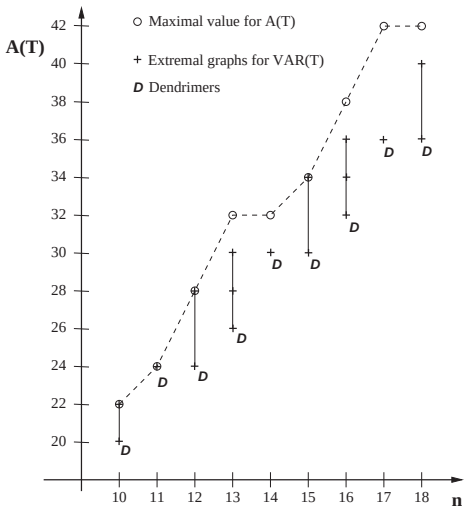


Figure 4.34: $A(T)$ values for the extremal graphs for VAR(T)

Figures 4.38 and 4.39 show the CS-values of the extremal graphs for $A(T)$ and VAR(T). With one exception (when $n = 11$), all the extremal graphs for VAR(T) have a different CS-value. It is also the case for $A(T)$ (for $n = 12$ and $n = 13$, extremal graphs for $A(T)$ are unique). Comparing differences between optimal CS(T) and CS-values for optimal graphs for $A(T)$ and for VAR(T), it appears that the differences are much larger in the former case (see Figure 4.38) than in the latter one (see Figure 4.39).

In summary, $A(T)$ (which has a period of 4 in the construction of the families) is apart from the two others (with a period of 3). It seems to express a different notion of irregularity.

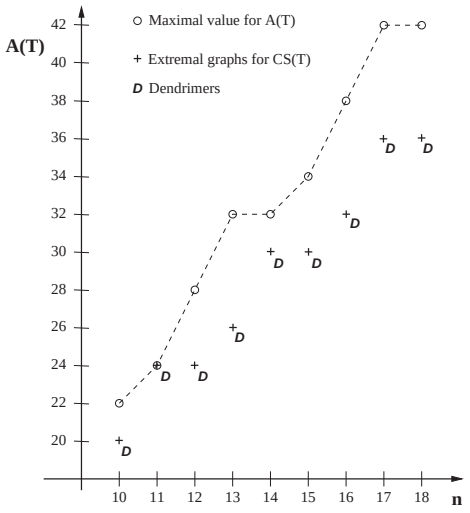


Figure 4.35: $A(T)$ values for the extremal graphs for CS(T)

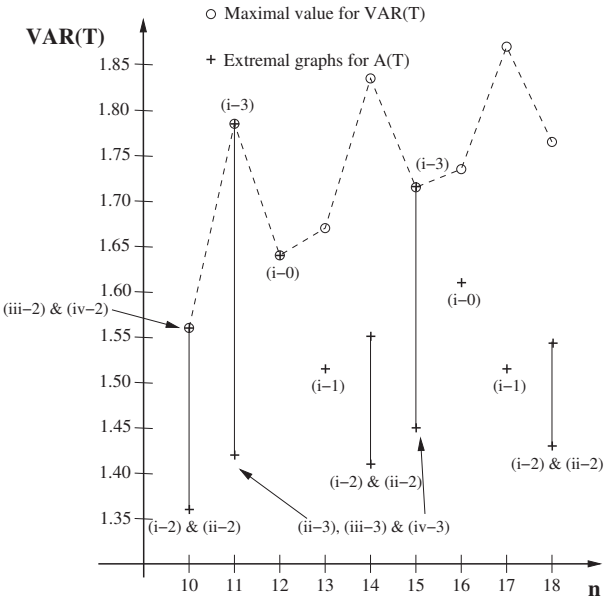


Figure 4.36: VAR(T) values for the extremal graphs for $A(T)$

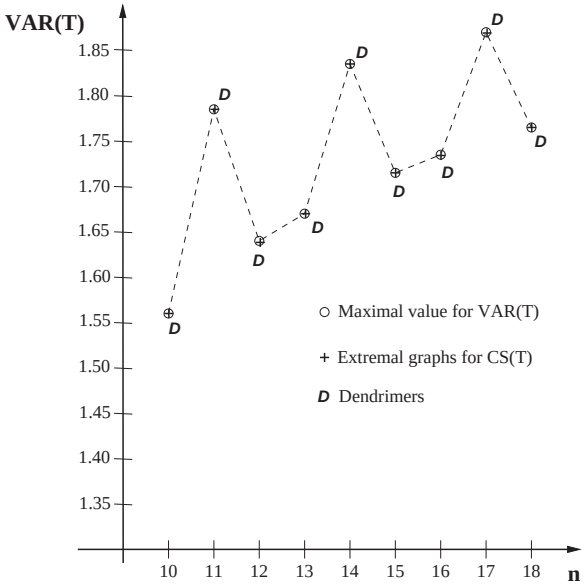


Figure 4.37: VAR(T) values for the extremal graphs for CS(T)

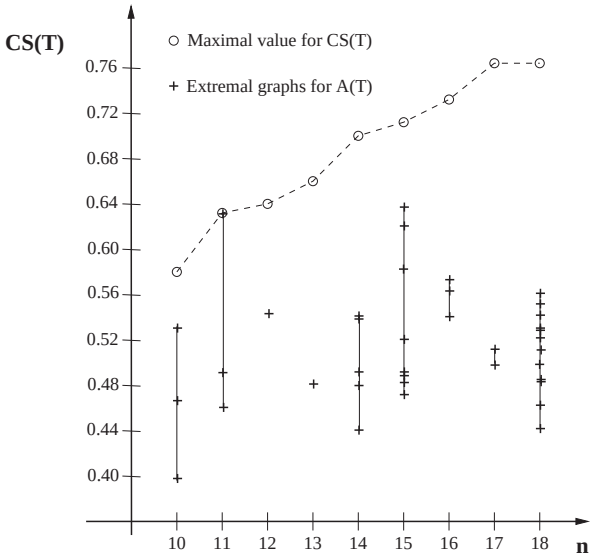


Figure 4.38: CS(T) values for the extremal graphs for $A(T)$

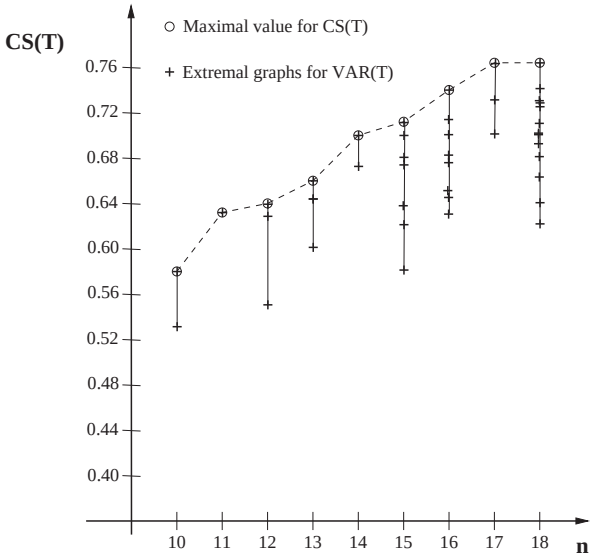


Figure 4.39: CS(T) values for the extremal graphs for VAR(T)

4.3 Variance of bounded degrees in graphs

We reconsider now the variance of degrees VAR in a more general context. A best possible bound and corresponding families of extremal graphs have been found by Bell [27] when the number of nodes n and the number of edges m are given and degrees are unbounded, i.e., $\Delta \leq n-1$. We consider here values of $\Delta < n-1$, and particularly $\Delta = 3$ and 4, which appear to be the most relevant to chemistry. Experiments with AutoGraphiX gave presumably optimal graphs, and are described in the next section. These graphs pointed the way to formulate and then prove the general result, given in Section 4.3.2. Special cases of chemical interest are considered in Section 4.3.3.

4.3.1 Experiments

We first used AutoGraphiX to find graphs with a fixed Δ and optimal or near-optimal values for the maximum variance of degrees. In order to get an idea of the dependence of the maximum degree Δ on these extremal graphs, we launched the system on different problems fixing the value of Δ and number of edges m. For example, Figure 4.40 shows the extremal graphs obtained by the system for $m = n-1$ (trees) and $\Delta = 3$ and Figure 4.41 when $m = n$ (unicyclic graphs) and $\Delta = 5$.

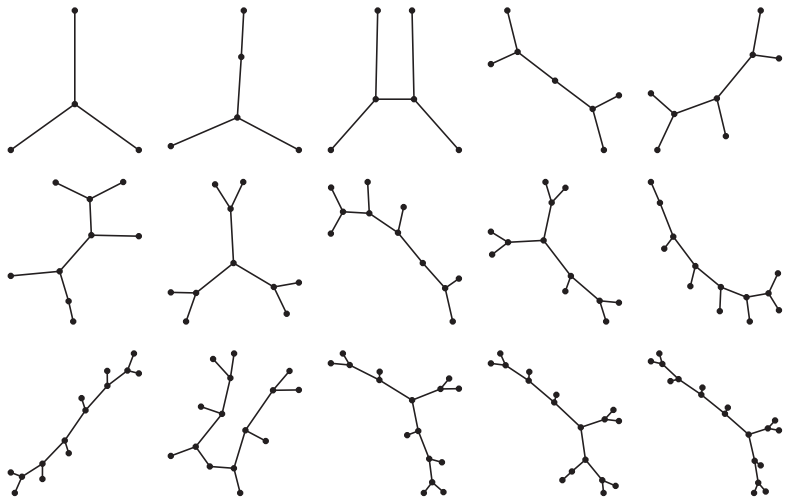


Figure 4.40: Some extremal trees with $\Delta = 3$ found by AutoGraphiX

As the variance depends only on the sequence of degrees, the observation of these extremal graphs leads quickly to the following conjecture : the extremal graphs have only nodes of degree Δ and 1 if this is compatible with the existence of a graph. We also observe that all extremal graphs, found by the system for different values of m and Δ, have at most one node which has a degree different from 1 or Δ.

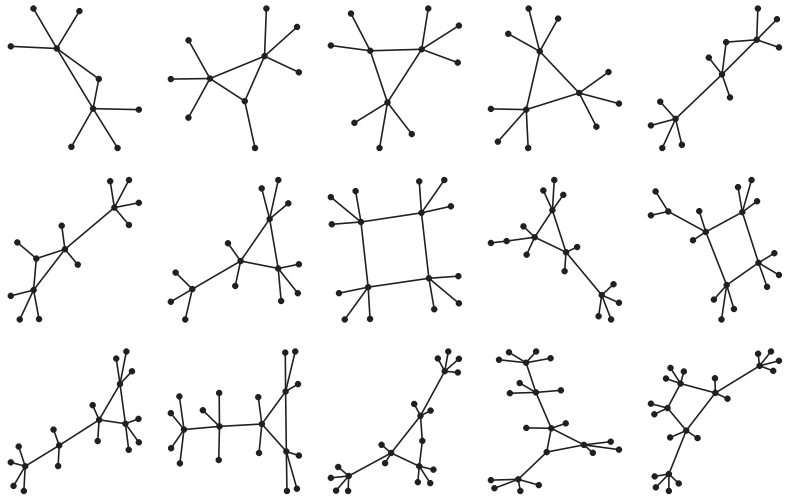


Figure 4.41: Some extremal unicyclic graphs with $\Delta = 5$ found by AutoGraphiX

4.3.2 Results

For a fixed value of n, there are only two connected graphs with $\Delta = 2$, the path P_n and the cycle C_n; the variance of their degrees is readily computed as $\mathrm{VAR}(P_n) = (2n-4)/n^2$ and $\mathrm{VAR}(C_n) = 0$. Hence, it is not restrictive to assume that $\Delta > 2$ in the following theorem.

Theorem 4.13. *For all connected graphs G with maximum degree $\Delta \geq 3$, n nodes and m edges,*

$$VAR(G) \leq \frac{2m(\Delta+1) - n\Delta + (1-k)(\Delta-k)}{n} - \left(\frac{2m}{n}\right)^2, \tag{4.107}$$

where

$$k = \Big[(2m-n)(mod\ \Delta - 1)\Big] + 1.$$

with equality in (4.107) attained if and only if at most one node of G has degree different from 1 *and* Δ.

Proof. Eq. (4.91) shows that for fixed values of Δ, m and n, VAR will be maximum if the first Zagreb index M_1 is maximum.

By definition of Δ and n_i,

$$M_1 = \sum_{i=1}^{\Delta} i^2\, n_i. \tag{4.108}$$

For all connected graphs with n nodes, m edges and bounded degree Δ, summing numbers of nodes of all degrees yields

$$n_1 + n_\Delta + \sum_{i=2}^{\Delta-1} n_i = n, \tag{4.109}$$

and summing degrees :

$$n_1 + \Delta n_\Delta + \sum_{i=2}^{\Delta-1} i\, n_i = 2m. \tag{4.110}$$

Solving (4.109) and (4.110) in n_1 and n_Δ gives

$$n_1 = \frac{1}{\Delta-1}\left[n\Delta - 2m + \sum_{i=2}^{\Delta-1}(i-\Delta)n_i\right], \tag{4.111}$$

and

$$n_\Delta = \frac{1}{\Delta-1}\left[2m - n + \sum_{i=2}^{\Delta-1}(1-i)n_i\right]. \tag{4.112}$$

Then, substituting (4.111) and (4.112) into (4.108) yields

$$\begin{aligned} M_1 &= \tfrac{1}{\Delta-1}\Bigg[\; n\Delta - 2m + 2m\Delta^2 - n\Delta^2 + \\ &\qquad \sum_{i=2}^{\Delta-1}\left(i - \Delta + \Delta^2 - i\Delta^2 + i^2\Delta - i^2\right) n_i \Bigg] \\ &= \tfrac{1}{\Delta-1}\Bigg[\; \Big(\Delta-1\Big)\Big(2m(\Delta+1) - n\Delta\Big) + \\ &\qquad \sum_{i=2}^{\Delta-1}\Big((\Delta-1)(1-i)(\Delta-i)\Big) n_i \Bigg] \end{aligned}$$

As $\Delta > 1$,

$$M_1 = 2m(\Delta+1) - n\Delta + \sum_{i=2}^{\Delta-1} f(i) n_i, \tag{4.113}$$

where $f(i)$ is the quadratic function

$$f(i) = (1-i)(\Delta - i). \tag{4.114}$$

Observe that $f(i)$ is strictly negative for $2 \le i \le \Delta - 1$. This implies that, for fixed values of n, m and Δ, the first Zagreb index M_1 (and VAR also) will be maximum if $n_i = 0$ for $i = 2, 3, \ldots, \Delta - 1$.

In this case, Eqs (4.112) and (4.111) lead to

$$n_1 = \frac{n\Delta - 2m}{\Delta - 1}, \tag{4.115}$$

and

$$n_\Delta = \frac{2m - n}{\Delta - 1}. \tag{4.116}$$

However, n_1 and n_Δ should have integer values which, as $n\Delta - 2m = n - 2m + n(\Delta - 1)$, is true in (4.115) and (4.116) if $2m - n$ is a multiple of $\Delta - 1$, i.e., if

$$(2m - n) (\mathrm{mod}\ \Delta - 1) = 0. \tag{4.117}$$

Suppose now that condition (4.117) is not respected. In this case, we have to maximize $\sum_{i=2}^{\Delta-1} f(i)n_i$. Let k denote the value of the left-hand side of (4.117) plus 1; then k is an integer between 2 and $\Delta-1$. It is always possible to choose the n_i such that

$$\begin{aligned} n_k &= 1, \\ n_i &= 0 \qquad \forall i(\neq k) = 2,3,\ldots,\Delta-1. \end{aligned}$$

and Eqs. (4.109) and (4.110) are satisfied. Indeed, as

$$k = \Big[(2m-n)(\mathrm{mod}\ \Delta-1)\Big] + 1, \tag{4.118}$$

Eqs. (4.111) and (4.112) become

$$n_1 = \frac{n\Delta - 2m + k - \Delta}{\Delta - 1}, \tag{4.119}$$

and

$$n_\Delta = \frac{2m - n + 1 - k}{\Delta - 1}, \tag{4.120}$$

which then have integer values.

Observe that k is the degree of the unique node of degree different from 1 and Δ. Indeed, the sum of all degrees is $2m$. Then consider a graph with all degrees equal to 1; this reduces this sum to n. Furthermore increase degrees one at a time up to Δ, as long as this is possible; this reduces the sum to $(2m-n)(\mathrm{mod}\ \Delta-1)$, which is equal to $k-1$, so the degree of one more node can be increased, up to k.

Now assume there are at least two nodes of degree i and j, with i not greater than j and both i and j in the interval $[2,\Delta-1]$. Reducing the degree of the first node by 1 changes (4.113) by

$$(1-i)(\Delta-1) - (1-i-1)(\Delta-i-1) = \Delta - 2i;$$

augmenting the degree of the second node by 1 changes (4.113) by

$$(1-j)(\Delta-j) - (1-j+1)(\Delta-j+1) = -\Delta + 2j - 2.$$

The net effect of both operations leaves the sum of degrees unchanged and changes (4.113) by

$$\Delta - 2i - \Delta + 2j - 2 = 2(j-i+1)$$

which is positive.

This means that the optimal choice for the n_i with $i \in \{2,\ldots,\Delta-1\}$, if condition (4.117) is not respected, is to choose only one n_k positive and equal to 1. In this case, by Eq. (4.113),

$$M_1 \leq 2m(\Delta+1) - n\Delta + (1-k)(\Delta-k). \tag{4.121}$$

Observe that if condition (4.117) is respected, Eq. (4.118) gives $k=1$, which expresses the fact that all n_i with $2 \leq i \leq \Delta-1$ are then equal to zero. Hence, the bound (4.121) is still valid when $k=1$.

Substituting (4.121) in (4.91) gives the result. □

4.3.3 Particular cases

Theorem 4.13 is valid for all connected graphs. However, the bound is sharp only if the values of n, m and Δ are compatible with the existence of a graph which has the specific sequence of degrees described in the proof of this theorem. Given a general condition on n, analyzing the dependence of the values of m and Δ requires a long discussion which breaks down in many particular cases. In this section, we restrict the graphs considered to some classes of chemical interest, and show how the theorem can be applied for such classes. In these cases, the bound is often sharp for reasonable values of n.

Again, we assume that $\Delta > 2$.

Trees

If T is a tree, as $m = n-1$, Theorem 4.13 gives after simplifications :

$$\mathrm{VAR}(T) \leq \frac{1}{n^2}\Big[n^2\Big(\Delta-2\Big)+n\Big(k^2-k(\Delta+1)-\Delta+6\Big)-4\Big], \tag{4.122}$$

where

$$k = \Big[(n-2)(\mathrm{mod}\,\Delta-1)\Big]+1. \tag{4.123}$$

If $n \geq 4$, this bound is sharp because the smallest tree with $\Delta \geq 3$ is the star S_4. For fixed values of $n \geq 4$ and $\Delta \leq n-1$, one can always construct a tree for which the equality holds. Starting from the star $S_{\Delta+1}$, add $\Delta-1$ pending edges to a node of degree 1, until nodes are exhausted.

Trees with $\Delta \leq 3$

If T has degrees bounded by 3, Eqs. (4.122) and (4.123) give

$$\mathrm{VAR}(T) \leq \frac{1}{n^2}\Big[n^2+n\Big(k^2-4k+3)-4\Big], \tag{4.124}$$

where

$$k = \Big[n(\mathrm{mod}\,2)\Big]+1. \tag{4.125}$$

Figure 4.42 shows some extremal trees with $\Delta \leq 3$ that can be obtained by applying the construction method explained above. Of course, for a given n, there can be more than one extremal graph, but they all share the following properties :

$$n_1 = \left\lfloor \frac{n+2}{2} \right\rfloor,\ n_2 = n(\mathrm{mod}\,2)\ \text{ and }\ n_3 = \left\lfloor \frac{n-2}{2} \right\rfloor,$$

where $\lfloor a \rfloor$ denotes the largest integer not larger than a.

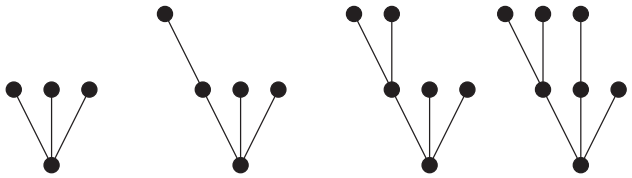


Figure 4.42: Some extremal trees with $\Delta \leq 3$

Chemical trees

If T is a chemical tree, i.e., $\Delta \leq 4$, Eqs. (4.122) and (4.123) gives

$$\mathrm{VAR}(T) \leq \frac{1}{n^2}\Big[2n^2 + n(k^2 - 5k + 2) - 4\Big], \tag{4.126}$$

where

$$k = \Big[(n-2)(\mathrm{mod}\ 3)\Big] + 1. \tag{4.127}$$

This problem was already solved in Section 4.2 in which a characterization of extremal chemical trees was also given[5].

Unicyclic graphs

If U is a unicyclic graph, as $m = n$, Theorem 4.13 leads to :

$$\mathrm{VAR}(U) \leq \frac{1}{n}\Big[n\Big(\Delta - 2\Big) + k^2 - k(\Delta + 1) + \Delta\Big], \tag{4.128}$$

where

$$k = \Big[n(\mathrm{mod}\ \Delta - 1)\Big] + 1. \tag{4.129}$$

If $n \geq 2\Delta - 1$, this bound is sharp because the smallest unicyclic graph with degrees bounded by Δ for which equality holds is the graph $U^*(\Delta)$ next described. This graph consists of a triangle of nodes $\{u, v, w\}$ with $\Delta - 2$ pending edges adjacent to v and $\Delta - 2$ pending edges adjacent to w. For example, the graph $U^*(5)$ is depicted in Figure 4.43. These graphs have $2\Delta - 1$ nodes.

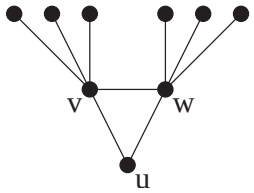


Figure 4.43: The graph $U^*(5)$

[5] See Theorem 4.10, p. 91 which is equivalent to Equations (4.126) and (4.127).

Starting from $U^*(\Delta)$, one can always construct a larger unicyclic graph for which the bound is sharp. It suffices to add $\Delta-2$ pending edges to the node u, and then $\Delta-1$ pending edges to a node of degree 1 until nodes are exhausted.

Unicyclic graphs with given girth

Unicyclic graphs with a triangle are not frequent in chemistry. Observe that it is possible to construct a unicyclic graph with maximum variance and given girth g (the length of the shortest cycle), if $n \geq g(\Delta-1)-\Delta+2$. One starts with a graph similar to U^* : a cycle of length g with all nodes (except one denoted by u) adjacent to $\Delta-2$ pending edges. At this point $n = g(\Delta-1)-\Delta+2$ and this unicyclic graph with given girth is the smallest one for which the bound is sharp. The same reasoning than before can be applied to construct a larger unicyclic graph for which the girth will always be g : add $\Delta-2$ pending edges to the node u, and then $\Delta-1$ pending edges to a node of degree 1 until nodes are exhausted.

Unicyclic graphs with $\Delta \leq 3$

If U is a unicyclic graph with $\Delta \leq 3$,

$$\mathrm{VAR}(U) \leq \frac{1}{n}\Big[n+k^2-4k+3\Big], \tag{4.130}$$

where

$$k = \Big[n(\mathrm{mod}\ 2)\Big]+1. \tag{4.131}$$

From the discussion on the unicyclic graphs, if $n \geq 5$ (or $n \geq 2g-1$ if the girth is given), the bound is sharp and all the extremal graphs have the following sequence of degrees :

$$n_1 = n_3 = \left\lfloor \frac{n}{2} \right\rfloor \ \text{ and } \ n_2 = n(\mathrm{mod}\ 2).$$

Figure 4.44 shows some unicyclic graphs with $\Delta \leq 3$ and a girth equal to 6 for which the bound is sharp.

Unicyclic chemical graphs

If U is a unicyclic chemical graph,

$$\mathrm{VAR}(U) \leq \frac{1}{n}\Big[2n+k^2-5k+4\Big], \tag{4.132}$$

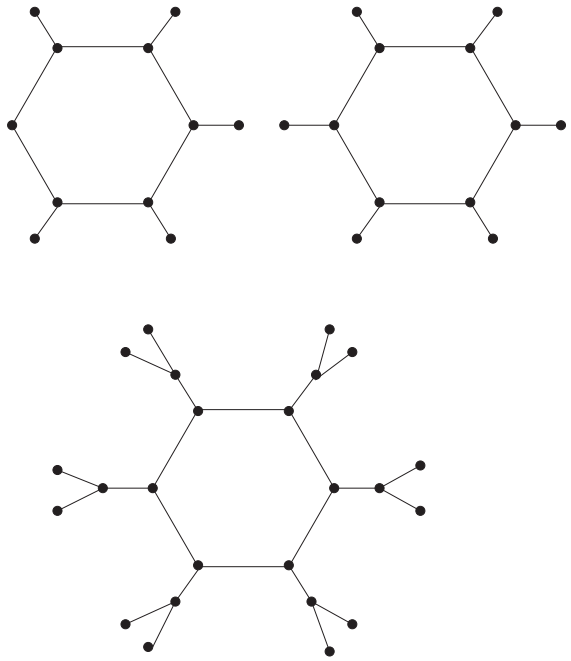


Figure 4.44: Some extremal unicyclic graphs with $\Delta \leq 3$ and $g = 6$

where

$$k = \Big[n(\text{mod } 3)\Big] + 1. \tag{4.133}$$

If $n \geq 7$ (or $n \geq 3g-2$ for a fixed girth), the bound is sharp and the extremal graphs have :

$$n_1 = \left\lfloor \frac{2n}{3} \right\rfloor,\ n_4 = \left\lfloor \frac{n}{3} \right\rfloor,$$

$n_2 = n_3 = 0$ if $n(\text{mod } 3) = 0$, $n_2 = 1, n_3 = 0$ if $n(\text{mod } 3) = 1$ and $n_2 = 0, n_3 = 1$ if $n(\text{mod } 3) = 2$.

Bicyclic graphs

If B is a bicyclic graph, then $m = n+1$. In this case, Theorem 4.13 leads to :

$$\text{VAR}(B) \leq \frac{1}{n^2}\Big[n^2\Big(\Delta-2\Big) + n\Big(k^2 - k(\Delta+1) + 3\Delta - 6\Big) - 4\Big], \tag{4.134}$$

where

$$k = \Big[(n+2)(\text{mod } \Delta-1)\Big] + 1. \tag{4.135}$$

The smallest bicyclic graph with degrees bounded by Δ for which equality holds is the graph $B^*(\Delta)$ which consists of a cycle of successive nodes (u,v,w,x) with a diagonal edge (v,x), $\Delta-2$ pending edges adjacent to w and $\Delta-3$ pending edges adjacent to v and x. Figure 4.45 shows $B^*(4)$ as an example. These graphs have $3\Delta-4$ nodes.

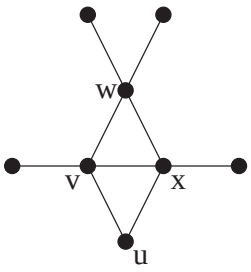


Figure 4.45: The graph $B^*(4)$

Again, one can construct a larger bicyclic graph for which the bound is sharp, starting from $B^*(\Delta)$ and applying the same method as for the unicyclic graphs : add $\Delta-2$ pending edges to the node u, and then $\Delta-1$ pending edges to a node of degree 1 until nodes are exhausted.

Bicyclic graphs with given girth

If the girth g is fixed, one can starts with two cycles of length g (in place of triangles) such that these cycle have an edge in common, and construct a larger bicyclic graph. Hence, the smallest bicyclic graph with girth g and maximum degree Δ for which the bound (4.134) is sharp has $g(2\Delta-2)-3\Delta+2$ nodes.

Figure 4.46 shows a smallest bicyclic chemical graph with girth 6 for which the bound is sharp.

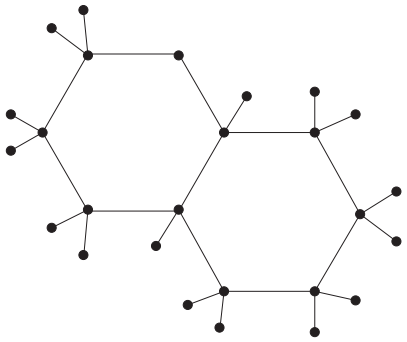


Figure 4.46: Extremal bicyclic chemical graph with $g=6$

Bicyclic graphs with $\Delta \leq 3$

If B is a bicyclic graph with $\Delta \leq 3$,

$$\mathrm{VAR}(B) \leq \frac{1}{n^2}\Big[n^2+n\Big(k^2-k(\Delta+1)+3\Delta-6\Big)-4\Big], \tag{4.136}$$

where

$$k = \Big[(n+2)(\mathrm{mod}\ 2)\Big] + 1, \tag{4.137}$$

and the bound is sharp if $n \geq 5$ (or $n \geq 4g-7$ for a fixed girth).

Bicyclic chemical graphs

If B is a bicyclic chemical graph,

$$\mathrm{VAR}(B) \leq \frac{1}{n^2}\Big[2n^2 + n\Big(k^2 - 5k + 6\Big) - 4\Big], \tag{4.138}$$

where

$$k = \Big[(n+2)(\mathrm{mod}\ 3)\Big] + 1, \tag{4.139}$$

and the bound is sharp if $n \geq 8$ (or $n \geq 6g-10$ for a fixed girth).

4.4 Summary

In Section 4.1, the usefulness of the AutoGraphiX system for computer-aided study of graphs is illustrated by a discussion of its use in the analysis, correction and improvement of bounds on the Randić index of chemical trees and graphs, recently proposed by Araujo and De la Peña [11]. The ramification index $ram(T)$ of a tree T is shown to be equal to the number of pending nodes of T minus 2. General trees with minimum Randić index, given order and number of pending nodes are characterized, extending a result of Bollobás and Erdös [36] : they are comets.

In view of a series of presumably extremal graphs obtained with AutoGraphiX, all bounds proposed by Araujo and De la Peña [11] are improved. Sharp lower and upper bounds on the Randić index of trees as function of this index for general trees and the ramification index (or, which is equivalent, the number of pending nodes) are obtained. Moreover, chemical trees of given order and number of pending nodes with minimum and with maximum Randić index are characterized. Lower and upper bounds on the Randić index of chemical graphs as a function of this index for induced maximal subgraphs which are trees are corrected and tight bounds obtained.

The theorems proved were first obtained as conjectures, which follow easily from the arrays of presumably extremal graphs given by AutoGraphiX. These graphs can be recognized to belong to some known family, as, e.g., comets in the case leading to Theorem 4.4. Moreover, figures representing values of corresponding indices, or functions of indices, often lead to conjectures in a straightforward way. This is the case for Theorem 4.5 and especially for Theorem 4.7.

These results show that the computer can be used in many more ways than number crunching in the study of chemical graphs.

The vast majority of chemical graphs is irregular, i.e., not all degrees of their nodes are the same. Clearly, some of these graphs are more irregular than others, hence the interest in irregularity measures. We studied and compared three of them in Section 4.2 : the Collatz-Sinogowitz index $CS(T)$, the variance of degrees $VAR(T)$ and the irregularity $A(T)$, applied to the class of chemical trees. To this effect, extremal graphs were generated with the system AutoGraphiX, conjectures derived and most of them proven by mathematical means.

The families of extremal chemical trees for $VAR(T)$ and for $A(T)$ are characterized and best possible bounds on these indices for all $n \geq 5$ are given.

It appears that $CS(T)$ and $VAR(T)$ are similar, extremal graphs having a periodicity of 3 in their construction. In contrast extremal trees for $A(T)$ have a periodicity of 4 and this index appears to express a different notion of irregularity than the two other ones. It also appears that CS and VAR measure not only irregularity, but also the extent of branching. This shows that "branching", a concept which has been extensively studied in the chemical literature, and the much less familiar concept of "irregularity" are not fully independent. This coincidence of the two concepts seems to hold as long as "irregularity" is defined globally, so as to depend on the number of certain structural features of the (molecular) graph, but not on their distribution within the graph. The irregularity $A(T)$ is of a different kind: it depends not so much on the node degree distribution, but on the differences of the node degrees of neighbors - a local structural feature. Thus, the irregularity $A(T)$ reflects properties of (molecular) graphs fully different from "branching", and would therefore deserve due attention in the future.

While a best upper bound for the variance of degrees of graphs has been available since 1992 [27], such a bound may provide values far from those observed for chemical graphs, where the maximum degree Δ is usually 3 or 4. We therefore provided, in Section 4.3, a formula explicitly taking into account the fact that degrees are bounded. We also examined when it is sharp for trees, unicyclic and bicyclic graphs.

CHAPTER 5

Bounding the irregularity of graphs : a simulated algorithm type of proof

Experiments and automated conjectures • Interpretation and new conjectures • Proofs • Summary

In Section 4.2.3, we proposed a sharp upper bound on the irregularity of chemical trees[1]. In this chapter we go further in the study of the irregularity, when the graphs are no more limited to chemical trees.

Albertson [1] defined the *imbalance* of an edge $\{v,w\}$ as

$$\mathrm{imb}_{vw} = |d_v - d_w|,$$

i.e., the difference of the degrees of that edge's end nodes, in absolute value, and the irregularity $A(G)$ of a graph G is thus the sum of all edge imbalances :

$$A(G) = \sum_{\{v,w\}\in E} \mathrm{imb}_{vw} = \sum_{\{v,w\}\in E} |d_v - d_w|.$$

It is shown in [1] that

$$A(G) < \frac{4n^3}{27},$$

and that this bound may be approached arbitrarily closely. However, no bound in function of n and m is given. It is the purpose of the present chapter to fill in this gap by providing a formula which is best possible in the strong sense, i.e., which is tight for all values of n and m compatible with the existence of a graph. Moreover, extremal graphs are characterized and belong to a single well-defined family of split graphs.

[1] See Theorem 4.12, p. 93.

These results [164] were again obtained using the system AutoGraphiX. Recall that AutoGraphiX uses the VNS metaheuristic to find extremal graphs (see Section 2.5). To find ideas of a proof, the program was run with severe restrictions on the moves allowed within the heuristic (e.g., only rotation, or displacement, of a single edge at a time). If such moves suffice for obtaining again the extremal graphs previously found, the effort can be focused on proving that they suffice in general.

First experiments with AutoGraphiX and their results, in terms of extremal graphs and 4 conjectures automatically obtained are described in the next section. They lead to conjecture a characterization of the family of extremal graphs and an upper bound on $A(G)$ in terms of n and m, presented in Section 5.2. Proofs of these conjectures are given in Section 5.3.

5.1 Experiments and automated conjectures

We run AutoGraphiX to find graphs with maximal irregularity for $3 \leq n \leq 12$ and $n-1 \leq m \leq \frac{n(n-1)}{2}$, as well as diagrams of these graphs, curves of values for a given n, and whatever conjectures might follow automatically. To find extremal graphs, the descent routine of AutoGraphiX was used with the following neighborhoods: move of an edge, rotation of an edge, split and insertion; and a number of neighborhoods $k_{max} = 10$ (see [55] for details). To find conjectures on extremal graphs, the invariants ω (clique number), α (stability number), χ (chromatic number), r (radius), Δ (maximum degree), and a few others were computed.

Then the numerical method of AutoGraphiX, described in [54, 56] (see Section 2.5), was applied to find automatically a basis of affine relations between these invariants, satisfied by all extremal (or near-extremal) graphs found.

A subset of the 230 graphs obtained is presented in Figure 5.1; the corresponding curves of $A(G)$ for $9 \leq n \leq 12$, are drawn in Figure 5.2. Moreover, AutoGraphiX provided the following 4 conjectures, which we group for conciseness :

Conjecture 5.1. *If G is a graph with n nodes, m edges, clique number ω, stability number α, chromatic number χ, maximum degree Δ, radius r and maximum irregularity A, then*

$$\omega = \chi, \tag{5.1}$$

$$n = \Delta + 1, \tag{5.2}$$

$$r = 1, \tag{5.3}$$

$$\omega + \alpha = \Delta + 2. \tag{5.4}$$

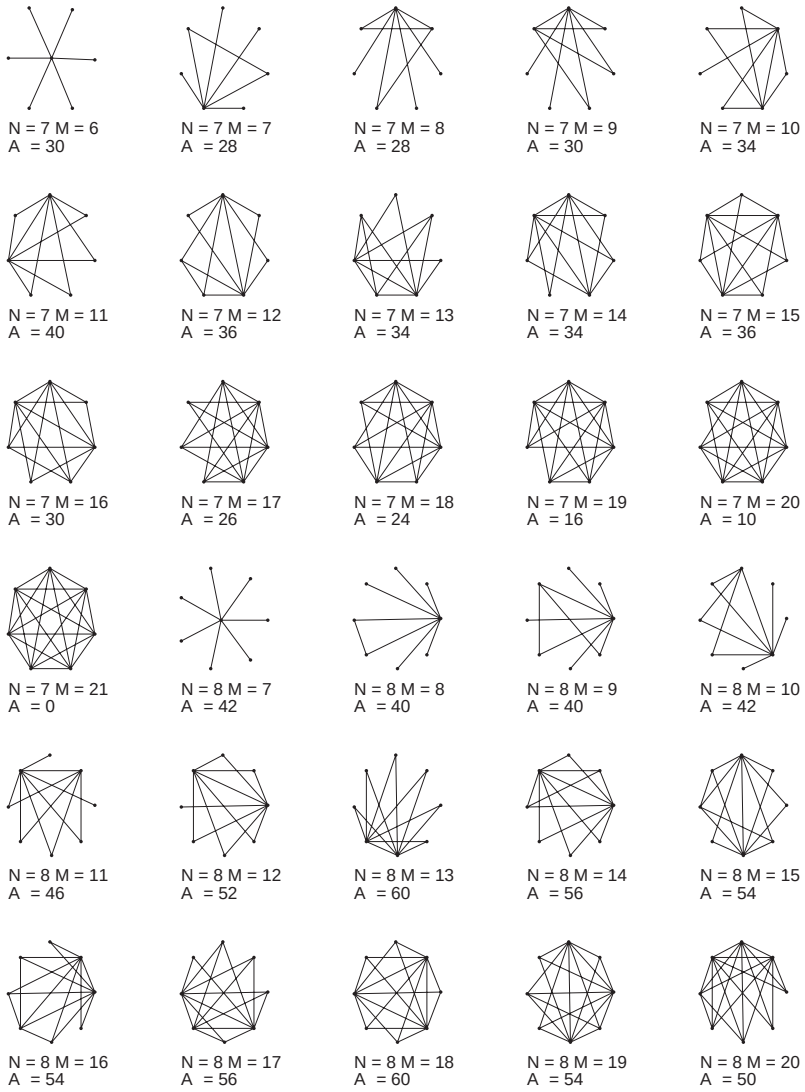


Figure 5.1: Some extremal graphs found by AutoGraphiX

5.2 Interpretation and new conjectures

Characterization of extremal graphs

From the basis of affine relations (5.1) – (5.4) given in Conjecture 5.1 one derives :

$$\omega + \alpha = n + 1; \tag{5.5}$$

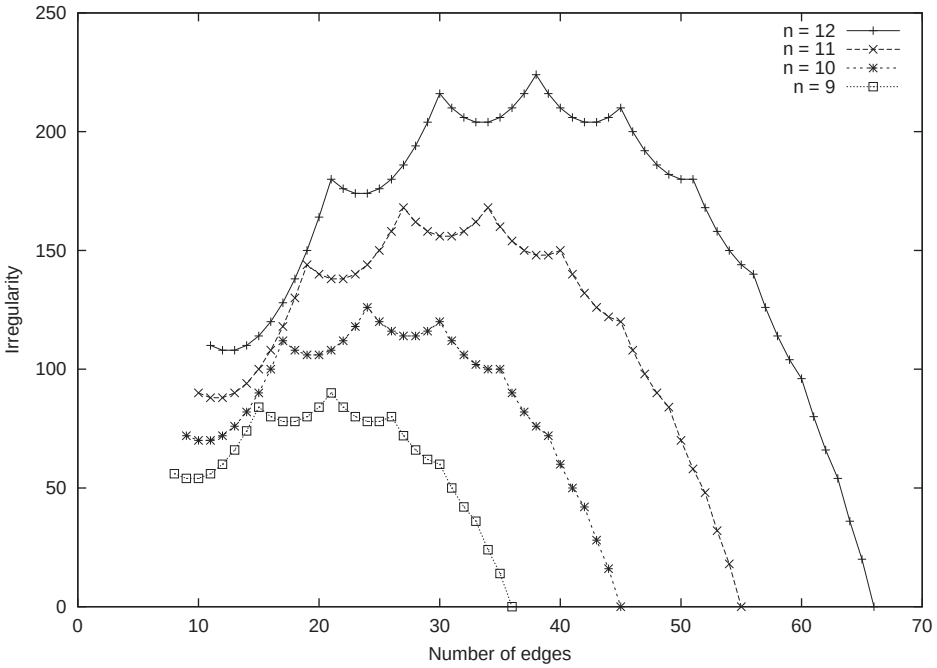


Figure 5.2: Values of Irregularity for some extremal graphs found by AutoGraphiX

so G comprises a clique C with ω nodes and an independent set I with α nodes which have a node in common. This implies they are specific split graphs. Indeed, split graphs consist of a clique on n_1 nodes, a disjoint independent set on $n_2 = n - n_1$ nodes and possibly some edges joining a node of one set to one of the other. If all such edges are present G is a complete split graph. Replacing n by $n_1 + n_2$ in (5.5) and as $\alpha = n_2$ by definition of a split graph,

$$\omega = n_1 + 1.$$

In other words, one node of the independent set is adjacent to all nodes of the clique in the split graph.

Moreover, at least one node is joined to all others, as $\Delta = n - 1$; this last relation implies $r = 1$ and conversely. Finally, the structure described implies $\chi = \omega$ as $\chi \geq \omega$ and a coloring in ω colors can be obtained by giving the same color to all of the n_2 nodes of I and a different color to each remaining node.

This gives fairly good information on the structure of extremal graphs. However, a glance at these graphs themselves (see again Figure 5.1), and particularly at successive ones with the same n and increasing m, shows one can say more. Indeed, they appear to be complete split graphs with a clique of d nodes, an independent set of $n - d$ nodes and possibly $1 \leq t \leq n - d - 1$ additional edges joining a node of the (previously) independent set to others of this set. We call such graphs *fanned split graphs* (as the addition of successive edges at a node is reminiscent of opening a fan). An example of a fanned split graph is presented in Figure 5.3 where $n = 8$, $m = 20$, $d = 3$ and $t = 2$. The three nodes of the clique are the black squares, and the dotted lines are the edges added

from a node not in the clique to two other ones noted by triangles. So we can formulate the next conjecture (which is computer-aided, not automated, but for which the largest part of the job was clearly done by computer) :

Conjecture 5.2. *A graph G with n nodes and m edges has maximum irregularity if and only if it is a fanned split graph.*

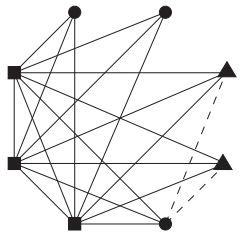


Figure 5.3: An example of fanned split graph with $n = 8$, $m = 20$, $d = 3$ (■) and $t = 2$ (▲)

Irregularity of fanned split graphs

A fanned split graph has

d nodes of degree $n-1$;
1 node of degree $d+t$;
t nodes of degree $d+1$;
$n-d-t-1$ nodes of degree d.

Moreover, G has $\frac{d(d-1)}{2}$ edges joining pairs of nodes of the clique, with imbalance 0; d edges joining nodes of the clique to the (first, if $t = 1$) node of degree $d+t$, with imbalance $n-d-t-1$; dt edges joining nodes of the clique to nodes of degree $d+1$, with imbalance $n-d-2$; $d(n-d-t-1)$ edges joining a node of the clique to a node of degree d, with imbalance $n-d-1$; and t edges joining a node of degree $d+t$ to a node of degree $d+1$ with imbalance $t-1$.

Summing, one obtains for a fanned split graph $G = FS_{dt}$ an irregularity of

$$A(FS_{dt}) = d(n-d-t-1) + dt(n-d-2) + d(n-d-1)(n-d-t-1) + t(t-1).$$

A few algebraic manipulations then lead to

$$A(FS_{dt}) = d(n-d)(n-d-1) + t(t-2d-1). \tag{5.6}$$

So, if $t = 0$ one gets the irregularity of a complete split graph with a clique on d nodes and an independent set on $n-d$ nodes, i.e., $d(n-d)(n-d-1)$. The effect of the additional t edges will reduce $A(G)$ if $t < 2d+1$, let it unchanged if $t = 2d+1$ and increase it if $t > 2d+1$. The local maxima for irregularity (see Figure 5.2) are obtained for complete split graphs.

From the definition of fanned split graph, one may easily compute d and t : indeed summing edges in the clique and between the clique and the independent set gives that d is the largest integer such that

$$\frac{d(d-1)}{2} + d(n-d) \leq m$$

from where it follows that

$$d = \lfloor n - \frac{1}{2} - \sqrt{(n-\frac{1}{2})^2 - 2m} \rfloor \tag{5.7}$$

where $\lfloor b \rfloor$ denotes the largest integer not larger than b and

$$t = m - (n-d)d - d(d-1)/2. \tag{5.8}$$

Note that these graphs are unique for fixed numbers n of nodes and m of edges.

We summarize these results in the following conjecture :

Conjecture 5.3. *For all graphs G with n nodes and m edges the irregularity*

$$A(G) \leq d(n-d)(n-d-1) + t(t-2d-1) \tag{5.9}$$

where d and t are given by (5.7) and (5.8). Moreover, the bound is attained for all n and $0 \leq m \leq \frac{n(n-1)}{2}$.

5.3 Proofs

FannedSplitGraph algorithm

Before looking for a proof of the conjectures, we ran AutoGraphiX with only one move allowed within the heuristic : we chose the simplest one for which the number of nodes and edges are not altered, i.e., rotation of a single edge. Let G be a graph and u, v, w three different nodes of G such that $\{u,v\} \in E(G)$ and $\{u,w\} \notin E(G)$; one can define the graph G' obtained after rotation $\mathrm{rot}(u,v,w)$ is applied to G as follows :

$$G' = G - \{u,v\} + \{u,w\}.$$

Such a move is represented in Figure 5.4.

With this restriction, AutoGraphiX found again systematically the fanned split graphs as extremal graphs. This observation led us to use the graphical interface of AutoGraphiX, which permits to modify the graphs manually and see how the invariants change in consequence. Such manipulation, together with the information on the structure of the fanned split graph collected before, led us to write the following algorithm which transforms any graph G with n nodes and m edges into a fanned split graph with the same number of nodes and edges, using only rotation.

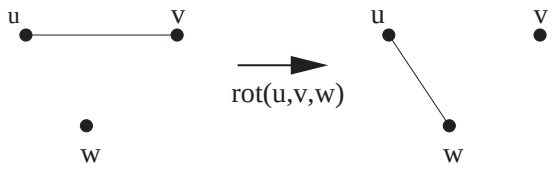


Figure 5.4: A rotation rot(u,v,w)

FANNEDSPLITGRAPH(G):
Input: a graph G with n nodes and m edges.
Output: a fanned split graph with n nodes and m edges.

1. *Initialization.*
   (a) Let F be the set of nodes $v \in V(G)$ with $d_v = n-1$. We call them the *fixed* nodes in the next steps.
   (b) $Stop \leftarrow FALSE$.
   (c) *Choose w.* Choose a non-fixed node w with maximum degree.
2. *Make a move.* While $Stop$ is $FALSE$, do
   (a) *Choose u.* Choose a node $u \notin F \cup \{w\}$ such that $\{u,w\} \notin E(G)$.
      i. If $d_u > |F|$, choose a node $v \notin F \cup \{u,w\}$ such that $\{u,v\} \in E(G)$ and do rot(u,v,w).
      ii. Else ($d_u = |F|$), try to find an edge $\{v_1,v_2\}$ where $v_1, v_2 \notin F \cup \{u,w\}$ and $d_{v_1} \geq d_{v_2}$.
         A. If such an edge exists, do
            rot(v_1,v_2,u) followed by rot(u,v_1,w).
         B. Else, $Stop \leftarrow TRUE$.
   (b) *Update F.* If $d_w = n-1$, $F \leftarrow F \cup w$ and choose a new w as a non-fixed node of maximum degree.

Lemma 5.1. *Algorithm **FannedSplitGraph** terminates if and only if G has been transformed into a fanned split graph.*

Proof. Let $x_1, x_2, \ldots x_f$ be the fixed nodes of G (with degree equal to $n-1$). Let w be the node of maximum degree $< n-1$. We see that in step **2**, a move will always increase the degree of w. Indeed, this algorithm will stop only if we can no more add an edge to w, i.e., if $f = d$ and if there is no more edge $\{v_1,v_2\}$ between nodes of the set $V \setminus F \setminus w$. This situation occurs only when G is a fanned split graph with the value d and t defined above.

Note that if $m < n-1$ this algorithm will construct a graph composed of a star with $m+1$ nodes and $n-m-1$ isolated nodes which is a fanned split graph as well. Moreover if G is not

connected and if $m \geq n-1$, this algorithm will choose w as the node of maximum degree until $d_w = n-1$. At this step G will be (and remain) connected.

Remark also that u, a node non-adjacent to w, can always be chosen in step **2a** because $d_w < n-1$ (otherwise w would be fixed) and if $d_u > |F|$, a node v can always be found because u is adjacent to the f fixed nodes and has at least one another edge to a node $\neq w$. □

Edge Rotation Lemma

We now introduce a lemma which expresses the change of the irregularity of a graph, when a rotation is applied on it. To do this, let us recall some notations introduced by Albertson [1]. If $u \in V$,

$$d_u^{>} = |\{x : \{x,u\} \in E \text{ and } d_u > d_{v_x}\}|,$$

$$d_u^{=} = |\{x : \{x,u\} \in E \text{ and } d_u = d_{v_x}\}|,$$

and

$$d_u^{<} = |\{x : \{x,u\} \in E \text{ and } d_u < d_{v_x}\}|.$$

Remark that $d_u = d_u^{>} + d_u^{=} + d_u^{<}$.

Lemma 5.2 (Edge Rotation Lemma)**.** *Let u,v,w be three different nodes of G. If $\{u,v\} \in E$ and $\{u,w\} \notin E$, set $G' = G - \{u,v\} + \{u,w\}$. Then we have the following results:*

Case 1	*If $d_u \geq d_v$ and $d_u > d_w$*	$A(G') = A(G) + 2[d_v^{<} + d_v^{=} - d_w^{<} - 1] + k$
Case 2	*If $d_u \geq d_v$ and $d_u \leq d_w$*	$A(G') = A(G) + 2[d_v^{<} + d_v^{=} + d_w^{>} + d_w^{=} - d_u] + k$
Case 3	*If $d_u < d_v$ and $d_u > d_w$*	$A(G') = A(G) + 2[d_u - d_v^{>} - d_w^{<}] + k$
Case 4	*If $d_u < d_v$ and $d_u \leq d_w$*	$A(G') = A(G) + 2[d_w^{>} + d_w^{=} - d_v^{>} + 1] + k$

where $k = 2$ if $(v,w) \in E(G)$ and $d_v = d_w + 1$, and $k = 0$ otherwise.

Proof. Suppose first that there is no edge between v and w. The imbalance of the edges of G' will change only on the edges adjacent with the nodes v and w in G. We will compute the contribution of the moving edge $\{u,v\} \to \{u,w\}$, of the edges adjacent with w in G and of the edges adjacent to v in G different from $\{u,v\}$.

- The contribution of the moving edge will be

$$|d_u - (d_w + 1)| - |d_u - d_v|. \tag{5.10}$$

- The total contribution of the edges adjacent with w in G will be

$$d_w^> + d_w^= - d_w^<. \tag{5.11}$$

- Finally, the contribution of the edges adjacent with v different from (u,v) will be

$$d_v^< + d_v^= - d_v^> - 1 \quad \text{if} \quad d_u \geq d_v \tag{5.12}$$
$$d_v^< + d_v^= - d_v^> + 1 \quad \text{if} \quad d_u < d_v \tag{5.13}$$

  because the moving edge, for which the imbalance has already been computed, counts for one in $d_v^< + d_v^=$ if $d_u \geq d_v$ or in $d_v^>$ otherwise.

Summing these contributions leads to the four cases (with $k = 0$), where we compare d_u and d_w as well to remove the absolute value in expression (5.10).

Suppose now that there exist an edge (v,w). The contributions will be similar of the previous one but we count the edge (v,w) twice : once in (5.11) and once in (5.12) or (5.13). In fact, the contribution of this edge will be

$$|(d_v - 1) - (d_w + 1)|. \tag{5.14}$$

Depending of d_v and d_w, this leads to several cases where we look how this edge was counted twice, the contribution already counted in the previous sum, and the value of imb_{vw} induced by (5.14) :

Assumption	Counted in	previous count	imb_{vw}
$d_v > d_w + 1$	$d_v^>$ and $d_w^<$	-2	-2
$d_v = d_w + 1$	$d_v^>$ and $d_w^<$	-2	0
$d_v = d_w$	$d_v^=$ and $d_w^=$	2	2
$d_v < d_w$	$d_v^<$ and $d_w^>$	2	2

There is only a difference when $d_v = d_w + 1$, which justifies the values of k in Lemma 5.2. □

Proof of Conjectures 5.2 and 5.3

We can now verify what AutoGraphiX suggest to us. These suggestions were summarized in Conjectures 5.2 and 5.3.

Lemma 5.3. *Running the **FannedSplitGraph** algorithm on any graph G which is not a fanned split graph will strictly increase irregularity in each iteration.*

Proof. If G is a fanned split graph, the algorithm will not do any move. Otherwise, there will be at least one move. So we only have to prove that steps **2(a)i** and **2(a)ii** will strictly increase irregularity of G.

Let $f = |F|$ be the number of nodes of degree $n-1$. As in steps **1c** or **2b** w is chosen such as d_w is maximum among the non-fixed nodes :

$$\begin{aligned} d_w^< &= f; && (5.15)\\ d_y^< &\geq f; && (5.16)\\ d_y^> + d_y^= + d_y^< &\leq d_w^> + d_w^= + d_w^<; && (5.17) \end{aligned}$$

for any non-fixed node y. Expressions (5.15) and (5.16) give

$$d_w^< \leq d_y^< \tag{5.18}$$

and, subtracting in both sides of (5.17)

$$d_y^> + d_y^= \leq d_w^> + d_w^=. \tag{5.19}$$

Let u be the node chosen in step **2a**.

- Suppose first that $d_u > f$. In this case, a node v is selected in step **2(a)i** and rotation $\mathrm{rot}(u,v,w)$ applied on G.

  By the choice of w and u, it is easy to check that $d_u < d_w$, and we have two cases :

  - If $d_u < d_v$, we are in Case 4 of Lemma 5.2 ,

    $$A(G') = A(G) + 2[d_w^> + d_w^= - d_v^> + 1] + k,$$

    and as v is non-fixed (5.19) implies that $A(G') > A(G)$.
  - If $d_u \geq d_v$, we are in Case 2 :

    $$A(G') = A(G) + 2[d_v^< + d_v^= + d_w^> + d_w^= - d_u] + k,$$

    and as $d_u < d_w$,

    $$d_v^< + d_v^= + d_w^> + d_w^= - d_u > d_v^< + d_v^= + d_w^> + d_w^= - d_w = d_v^< + d_v^= - d_w^<$$

    which is positive as $d_v^< \geq d_w^<$ by (5.18). Again, $A(G') > A(G)$.

- Suppose now that $d_u = f$. In this case, one tries to find an edge (v_1, v_2) in step **2(a)ii** such that $\{v_1, v_2\} \notin F \cup w$ and $d_{v_1} \geq d_{v_2}$. Then two rotations are applied on G. Let G' be the graph obtained after rotation $\mathrm{rot}(v_1, v_2, u)$ is applied on G and G'' the graph obtained after rotation $\mathrm{rot}(u, v_1, w)$ is applied on G'. As some degrees will change after the first rotation, we will note them d_i' in G'. By construction $d_{v_1} \geq d_{v_2} > d_u = f$ which means that we are in Case 1 of Lemma 5.2 for the first rotation :

  $$A(G') = A(G) + 2[d_{v_2}^< + d_{v_2}^= - d_u^< - 1] + k$$

where $k = 0$ because there is no edge (v_2, u) and where $d_u^< = d_u = f$. So,

$$A(G') = A(G) + 2[d_{v_2}^< + d_{v_2}^= - f - 1]. \tag{5.20}$$

Let us see now how the degrees have changed before the second rotation. It is clear that

$$d'_u = d_u + 1, d'_{v_1} = d_{v_1} \text{ and } d'_w = d_w \tag{5.21}$$

but some other changes can happen in the decomposition of these degrees in the sets $d^<, d^>$ and $d^=$. There are three different possible configurations :

1. $d_{v_1} > d_{v_2}$. As $d_{v_2} \geq d_u + 1$ and by (5.21) we have $d'_{v_1} > d'_u$. Moreover, $d'_w \geq d'_u$ and we are in Case 4 of Lemma 5.2:

   $$A(G'') = A(G') + 2[d_w^{>'} + d_w^{='} - d_{v_1}^{>'} + 1] + k$$

   where $k \geq 0$ and where $d_{v_1}^{>'} = d_{v_1}^>, d_w^{>'} = d_w^>$ and $d_w^{='} = d_w^=$. Finally, replacing $A(G')$ by (5.20) gives

   $$A(G'') = A(G) + 2[d_{v_2}^< + d_{v_2}^= - f - 1 + d_w^> + d_w^= - d_{v_1}^> + 1] + k.$$

   which is strictly greater than zero because $d_{v_2}^< \geq f + 1$ (the f edges from v_2 to the fixed nodes, plus the edge $\{v_1, v_2\}$) and $d_w^> + d_w^= \geq d_{v_1}^>$ by (5.19).

2. $d_{v_1} = d_{v_2}$ and $d_{v_2} > d_u + 1$. We are again in Case 4 but the difference here is that $d_{v_1}^{>'} = d_{v_1}^> + 1$:

   $$A(G'') = A(G) + 2[d_{v_2}^< + d_{v_2}^= - f - 1 + d_w^> + d_w^= - d_{v_1}^>] + k.$$

   which is strictly greater than zero because $d_{v_2}^< \geq f + 1$, $d_w^> + d_w^= \geq d_{v_1}^>$ as before and $d_{v_2}^= \geq 1$ due to the edge (v_1, v_2).

3. $d_{v_1} = d_{v_2}$ and $d_{v_2} = d_u + 1$. In this last case, $d'_{v_1} = d'_u$ and $d'_w \geq d'_u$, so we are in Case 2 :

   $$A(G'') = A(G') + 2[d_{v_1}^{<'} + d_{v_1}^{='} + d_w^{>'} + d_w^{='} - d'_u] + k.$$

   One can check that $k = 0$, $d'_u = f + 1$, $d_{v_1}^{='} = 1$, $d_{v_1}^{<'} = f$, $d_w^{>'} = d_w^>$ and $d_w^{='} = d_w^=$. Putting these values in the previous expression and replacing $A(G')$ by (5.20) leads to

   $$A(G'') = A(G) + 2[d_{v_2}^< + d_{v_2}^= + d_w^> + d_w^= - f - 1],$$

   which is strictly positive because $d_{v_2}^= = 1$, $d_{v_2}^{<'} = f$ and $d_w^> > 0$. □

Theorem 5.1. *(Conjectures 5.2 and 5.3) For any graph G with n nodes, m edges and irregularity $A(G)$,*

$$A(G) \leq d(n-d)(n-d+1) + t(t-2d-1)$$

where

$$d = \lfloor n - \frac{1}{2} - \sqrt{(n-\frac{1}{2})^2 - 2m} \rfloor$$

and

$$t = m - (n-d)d - d(d-1)/2.$$

Moreover, this value is attained if and only if G is a fanned split graph.

Proof. From Lemma 5.1, algorithm **FannedSplitGraph** applied to any graph G ends with the unique fanned split graph with n nodes and m edges. From Lemma 5.3 all rotations or pairs of rotations applied increase strictly the irregularity of the graph. It follows that fanned split graphs are extremal (which is Conjecture 5.2). Then the bound follows from the computations preceding Conjecture 5.3. The fact that it is the best possible bound and that extremal graphs can be characterized as fanned split graphs also follows from Lemmas 5.1 and 5.3. □

5.4 Summary

The problem of finding a best possible bound on the irregularity of graphs G with n nodes and m edges, as defined by Albertson [1], is entirely solved. Moreover, extremal graphs are characterized.

These results have been obtained through three of the main capacities of the system AutoGraphiX :

1. Finding extremal or near extremal graphs. In this case AutoGraphiX obtained 230 graphs which were extremal, without one exception.

2. Finding automatically conjectures. Four such conjectures were obtained, from which it follows that extremal graphs are split graphs with one node of largest possible degree.

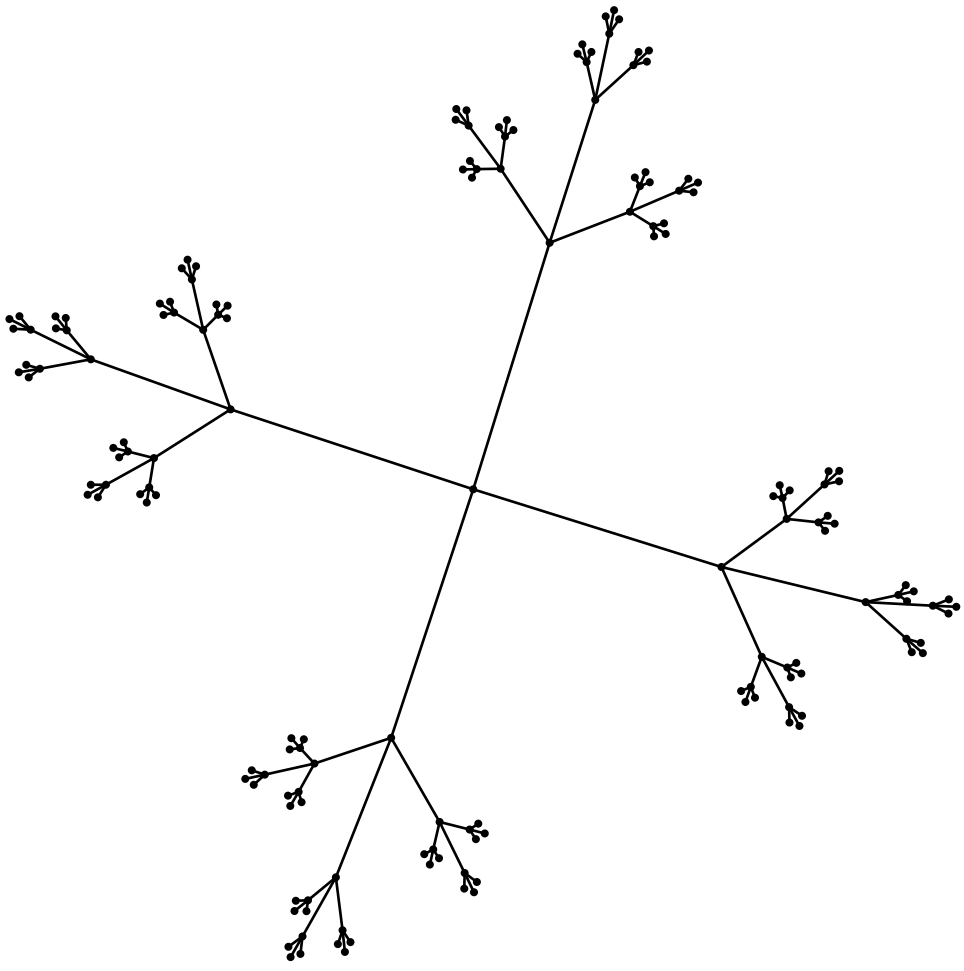
   Using these results and the representation of graphs found, a new family of graphs was identified : *fanned split graphs*, which are complete split graphs with possibly some additional edges all incident with a same node. The 230 graphs obtained all belong to this family, which has one and only one member for each pair of numbers n of nodes and m of edges compatible with the existence of a graph. The conjectures that extremal graphs always belong to this family, as well as a numerical bound on irregularity, follow.

3. Suggesting proof strategies. The extremal graphs obtained could be found once again using only the edge rotation move, which is the simplest one leaving n and m unchanged. This suggested an algorithm to go from any graph G to a fanned split graph with the same number of nodes and edges, using only moves which increase irregularity. Such an algorithm using one or two rotations at each step could be obtained, with help of the interactive component of AutoGraphiX. The two conjectures were thus proved.

We believe the “simulated algorithm” type of proof used here to be worthy of further study.

Part III

The system GraPHedron — principles and results



Introduction

Several conjecture-making systems have been presented in Chapter 2. It appears that each system explores specific types of conjectures and that some theorems, discovered without computer assistance, have forms that are not considered by them at this time [160]. The forms of conjectures found with the help of these systems heavily depend on the principles and approach they use, but are very often algebraic expressions of invariants.

We present here a new computer-assisted system, called GraPHedron [62, 58, 196], that we developed in the context of this PhD dissertation. It can, in some cases, give conjectures in a fully automated way. Its specificity lies in the fact that it tries to answer the following question :

> *What are all the best inequalities among a selected set of invariants, valid for graphs of a given class ?*

The arguments used to answer this question are taken from polyhedral theory. Actually, the name GraPHedron is the contraction of the words *graph* and *polyhedron*.

The system GraPHedron is outlined in Chapter 6. In this chapter, we present its principles, applications and future perspectives. The forms of the conjectures that can be derived with the help of GraPHedron are also studied. A few words are said about automatically derived conjectures, but this feature is still under development. This presentation chapter is illustrated with a complete example : all optimal relations, between the diameter D and the number of edges m of connected graphs with n nodes, are suggested by GraPHedron and then proved.

Chapter 7 is devoted to the first results obtained with GraPHedron. Even if the system is recent[2], we obtained several interesting results and its strategy appears to be very fruitful.

[2]The first prototype version (written in *C*) was started in December 2002 but was only usable by the author itself. However, as this experiment led already to some conjectures, we completely rewrote this prototype (in C++) and added an important part of its current features. This first real version of GraPHedron was ready in November 2003. It is currently used by a dozen of computer scientists and mathematicians.

CHAPTER 6

A new conjecture-making system

Principles • Outline • Applications • Forms of conjectures • Automated conjectures • Summary

We present now GraPHedron, a computer system which helps to find optimal conjectures in graph theory.

This chapter is organized as follows. In Section 6.1, we explain the principles of GraPHedron, and give our answer to the question stated in the introduction of this part (p. 129). The outline of GraPHedron is detailed in Section 6.2. We next present several possible applications in Section 6.3. The forms of the conjectures that can be derived with GraPHedron are studied in Section 6.4. Some conjectures are given automatically but this feature has to be improved. This is briefly discussed in Section 6.5.

6.1 Principles

Some new notations and definitions are introduced in Section 6.1.1. Then, in Section 6.1.2, the polyhedral approach of GraPHedron is explained. This approach is illustrated in Section 6.1.3 by an example. Finally, we explain in Section 6.1.4 the differences between this approach and the geometric procedure used in the system AutoGraphiX (see Section 2.5 and Part II).

6.1.1 Notations

Classical notions of polyhedral theory are used in this chapter. Readers that are not familiar with these notions are referred to the standard textbook of Ziegler [236]. However, we recall and fix some notations and definitions in this section.

Definition 6.1. *The* set of invariants $I = \{i_1, i_2, \ldots, i_p\}$ *is a set of p algebraic expressions involving one or more graph invariants, excluding the number of nodes n.*

Remark 6.1. This definition is not ambiguous because an algebraic expression involving several invariants is also an invariant.

Definition 6.2. *For a given set I of p invariants, we define the p-dimensional* space of invariants $\mathbb{S}^p$. *Each coordinate* x_k *of* $\mathbb{S}^p$, *for* $k = 1, 2, \ldots p$, *is associated with the invariant* i_k.

Remark 6.2. Depending on the type of the invariant values, $\mathbb{S}^p$ can be $\mathbb{Z}^p$, $\mathbb{Q}^p$ or $\mathbb{R}^p$. If the invariants i_k have boolean or integer values, $\mathbb{S}^p$ is $\mathbb{Z}^p$ and boolean values are considered as integers (where true is 1 and false is 0). If at least one invariant $i_k \in \mathbb{R}$, then $\mathbb{S}^p$ is the real space $\mathbb{R}^p$. Finally, if at least one invariant $i_k \in \mathbb{Q}$ and no other invariant has real values, then $\mathbb{S}^p$ is $\mathbb{Q}^p$.

Definition 6.3. *Let S be a set of points in* $\mathbb{S}^p$. *The p-dimensional* polytope $\mathcal{P} = \text{conv}(S)$ *is the convex hull of S.*

From a geometrical point of view, a polytope is an intersection of half-spaces. A p-dimensional polytope can thus be defined as a set of solutions $x \in \mathbb{S}^p$ of a system of k linear inequalities

$$Ax \leq b \tag{6.1}$$

where $A \in \mathbb{S}^{k \times p}$ is a $k \times p$ matrix and $b \in \mathbb{S}^k$ is a k-vector.

Definition 6.4. *A* face *of* $\mathcal{P}$ *is the intersection of* $\mathcal{P}$ *with a tangent hyperplane. Zero-dimensional faces are* vertices*, one-dimensional faces are* polytope edges *and* $(p-1)$*-dimensional faces are* facets.

Remark 6.3. We recall that to avoid confusion we use the terms *node* and *edge* for graphs and *vertex* and *polytope edge* for polytopes.

Definition 6.5. *An inequality of the system (6.1) is called a* facet defining inequality.

Definition 6.6. $\mathcal{F}(\mathcal{P})$ *is the system of facets defining inequalities describing* $\mathcal{P}$.

6.1.2 A polyhedral approach

As already stressed in the introduction of this part, and as often seen in the previous chapters, theorems in graph theory are often expressions (equalities or inequalities) involving a set of invariants, under some conditions that graphs should respect. These conditions, or hypotheses, are generally a specific *class of graphs* $\mathcal{C}$. For example, a theorem may be valid only for connected, bipartite or planar graphs.

Definition 6.7. *$\mathcal{C}_n$ is the set of all non-isomorphic graphs with n nodes, belonging to a given class $\mathcal{C}$.*

Questions

We ask the following question :

> *What are all the best linear inequalities among invariants of I, valid for all graphs of $\mathcal{C}_n$?*

Actually, to answer this question, one needs to answer some related ones :

a) *How to define a "best" or "optimal" linear inequality when n is given ?*

b) *What does it mean "all" inequalities for a given problem ?*

c) *How to find these inequalities ?*

Answers

We derive answers from a polyhedral approach, considering graphs as points in a space of invariants $\mathbb{S}^p$. For a fixed value of n, each graph G of $\mathcal{C}_n$ is represented by a point $(x_1, x_2, \ldots, x_p)$ where $x_k = i_k(G)$.

Definition 6.8. *For a given class of graphs $\mathcal{C}$, a given set of invariants I and a fixed integer n, we define the* invariants polytope $\mathcal{P}_n^{\mathcal{C},I}$ *as*

$$\mathcal{P}_n^{\mathcal{C},I} = \text{conv}\left(\mathcal{S}_n^{\mathcal{C},I}\right),$$

where

$$\mathcal{S}_n = \{(x_1, x_2, \ldots, x_p) \in \mathbb{S}^p \mid \exists G \in \mathcal{C}_n \ : i_1(G) = x_1, i_2(G) = x_2, \ldots, i_p(G) = x_p\}.$$

Remark 6.4. When no confusion is possible, $\mathcal{P}_n^{\mathcal{C},I}$ is simply denoted by $\mathcal{P}_n$ and $\mathcal{S}_n^{\mathcal{C},I}$ is denoted by $\mathcal{S}_n$.

The polytope $\mathcal{P}_n$ can be described by a system $\mathcal{F}(\mathcal{P}_n)$ of facets defining inequalities. Any such linear inequality can be considered as best possible :

- any valid linear inequality among the invariants of I is dominated by a positive combination of facets defining inequalities $\in \mathcal{F}(\mathcal{P}_n)$;
- $\mathcal{F}(\mathcal{P}_n)$ constitutes a minimal system describing $\mathcal{P}_n$, i.e., no facet defining inequality $\in \mathcal{F}(\mathcal{P}_n)$ can be a logical consequence of any other valid inequalities.

This definition of optimality is thus stronger than a "tight" inequality – a classical argument of quality – which means only that the inequality defines a supporting hyperplane of $\mathcal{P}_n$.

Conjectures finding

From the computation of $\mathcal{P}_n$, it follows a fruitful strategy to formulate conjectures. If n is small, $\mathcal{P}_n$ can be computed effectively with a computer. The idea of the computer system GraPHedron is to compute the polytopes $\mathcal{P}_n$ for some reasonable values of n and to display detailed information about them. If similarities between a facet of each $\mathcal{P}_n$ can be pointed out, one often obtains a conjectured generalization of this facet for all n. Sometimes, the complete system $\mathcal{F}(\mathcal{P}_n)$ can be conjectured and generalized for each n. The characterization of the *vertex graphs* is also very helpful to derive conjectures.

Definition 6.9. *Let $G \in \mathcal{C}_n$ be a graph and $\mathcal{P}_n^{\mathcal{C},I}$ an invariants polytope. If the point $(i_1(G), i_2(G), \ldots, i_p(G))$ is a vertex of $\mathcal{P}_n^{\mathcal{C},I}$, then G is called a* vertex graph *of $\mathcal{P}_n^{\mathcal{C},I}$.*

6.1.3 Illustration

To illustrate the polyhedral approach, we introduce a simple but beautiful example.

Problem 6.1. What are all the best linear inequalities between the diameter D and the number of edges m of connected graphs with n nodes ?

Applying the polyhedral approach

Thus $I = \{D, m\}$ and $\mathcal{C}$ is the class of connected graphs. Figure 6.1 shows all graphs in $\mathcal{C}_4$, the associated vectors of coordinates (D, m) and the corresponding polytope $\mathcal{P}_4$.

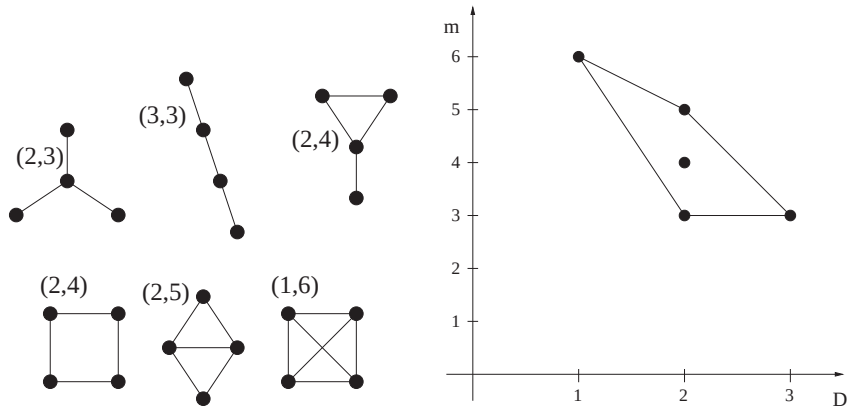


Figure 6.1: Graphs of $\mathcal{C}_4$, their coordinates (D,m) and the polytope $\mathcal{P}_4$

This polytope is defined by the following set of facets defining inequalities :

$$\begin{aligned} m &\geq 3, \\ 3D+m &\geq 9, \\ D+m &\leq 7, \\ 2D+m &\leq 9. \end{aligned}$$

Polytopes $\mathcal{P}_n$ with larger values of n can then be computed. Figure 6.2 shows representations of $\mathcal{P}_9$ and $\mathcal{P}_{10}$ produced by GraPHedron.

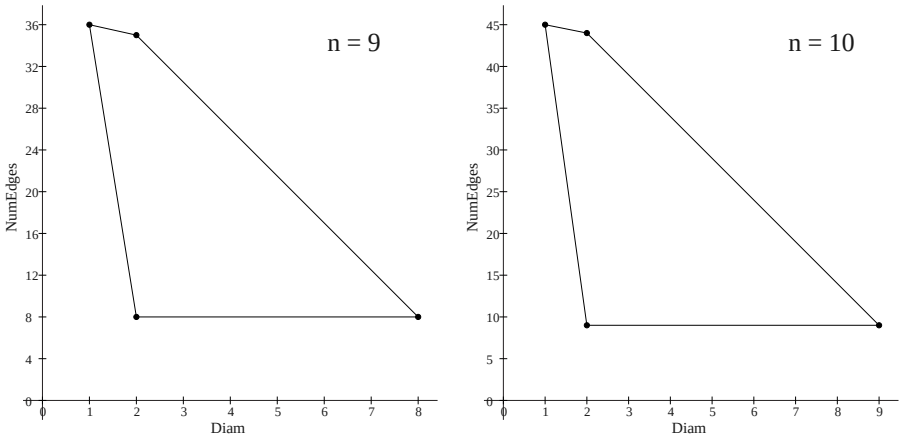


Figure 6.2: The polytopes $\mathcal{P}_9$ and $\mathcal{P}_{10}$

Figures 6.3 and 6.4 show drawings of the vertex graphs (see Definition 6.9, p. 134) of $\mathcal{P}_9$ and $\mathcal{P}_{10}$, as displayed by GraPHedron.

Derivation of conjectures

A look at the information displayed by GraPHedron for the polytopes $\mathcal{P}_n$ $(n = 4,5,\ldots,11)$ leads to the following observations. The polytopes have always four facets and a very similar shape.

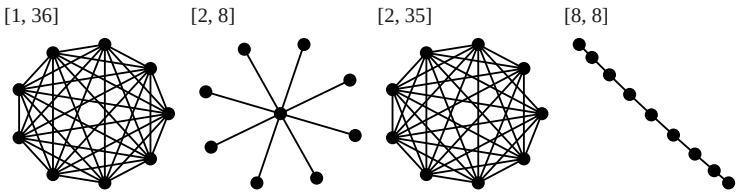


Figure 6.3: The vertex graphs of $\mathcal{P}_9$

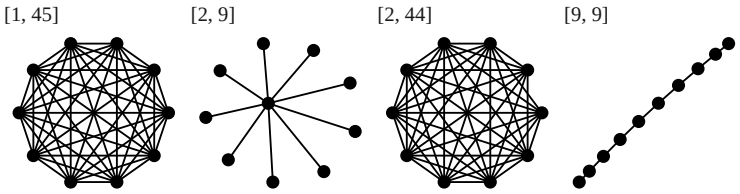


Figure 6.4: The vertex graphs of $\mathcal{P}_{10}$

Each vertex always corresponds to only one graph. These vertex graphs are easily characterized : the star S_n, the path P_n, the complete graph K_n and the complete graph with a removed edge $K_n \backslash e$.

Remark 6.5. The report created by GraPHedron for this problem is presented in Appendix B.1 (pages 199 to 217).

The coordinates (D,m) of these vertex graphs are trivially generalized in terms of n :

$$S_n \quad : \quad (2, n-1), \tag{6.2}$$

$$P_n \quad : \quad (n-1, n-1), \tag{6.3}$$

$$K_n \quad : \quad \left(1, \binom{n}{2}\right), \tag{6.4}$$

$$K_n \backslash e \quad : \quad \left(2, \binom{n}{2} - 1\right). \tag{6.5}$$

If one conjectures that the coordinates associated with these graphs are always vertices of $\mathcal{P}_n$ (and that no other graph corresponds to a vertex), the facets defining inequalities are also easily generalized. It suffices to compute the equations of the lines through each pair of respective points. The polytopes description $\mathcal{F}(\mathcal{P}_n)$ can now be conjectured for each n. When n is too small, some border effects arise. For instance, $S_3 \simeq P_3$. To avoid such discussion, we assume that $n \geq 4$.

Conjecture 6.1. *For each connected graph G with $n \geq 4$ nodes, m edges and a diameter D,*

$$\begin{aligned} m &\geq n-1, && (6.6)\\ \frac{(n-1)(n-2)}{2}D + m &\geq (n-1)^2, && (6.7)\\ D + m &\leq \frac{n(n-1)}{2} + 1, && (6.8)\\ nD + 2m &\leq (n+2)(n-1), && (6.9) \end{aligned}$$

and these inequalities are the only possible facets defining inequalities for each n.

This conjecture is proven in Section 6.3.1.

Remark 6.6. As shown in Section 9 of the report (see Appendix B.1, p. 215), inequalities (6.6) – (6.9) are automatically found by GraPHedron. This feature is discussed in Section 6.5.

6.1.4 The geometric procedure of AutoGraphiX

This polyhedral approach is simple and natural but has not yet been exploited systematically. However, there are similarities between this approach and one of the features of AutoGraphiX.

Recall that the system AutoGraphiX (see Section 2.5 and Part II) uses a metaheuristic to obtain a set of graphs G_k, where $k = k_{min}, \ldots, k_{max}$ is a parameter such as the number of nodes. These graphs are extremal or near-extremal for a given objective function $f(G)$. The set of graphs considered can be extended using two parameters together and G can be restricted to a specific class by adding constraints to the optimization problem.

We have seen (see again Section 2.5) that AutoGraphiX applies three different methods to obtain conjectures automatically from the set G_k. One of theses methods is the *geometric procedure*, which consists in considering the graphs G_k "as points in a space of characteristics, then uses a convex-hull (or gift-wrapping) algorithm to find facets, which correspond to conjectures" ([56], p. 83). This procedure is thus similar to our approach but there are important differences.

Caporossi and Hansen restrict the points to a set of graphs which are extremal or near-extremal for $f(G)$ and they study only the facets relevant for the type of the optimization problem (minimization or maximization) : the coordinates are thus always $(f(G_k), k)$. Of course, the order of graphs G_k is larger than what can be obtained by enumeration.

On the contrary, we consider all non-isomorphic graphs of $\mathcal{C}$ and try to characterize all the facets of the convex hull of a free dimensional set of points. It allows to identify a finite number of optimal relations from which all other relations follow, i.e., a minimal system of optimal relations. Therefore, the forms of the conjectures derived with the two systems are different. Finally, we output the polyhedral information about the problem which can be useful to understand the relations between invariants of I and to prove the derived conjectures.

6.2 Outline

GraPHedron is written in *C++* and is developed to run on Unix-like systems. The current version is a command-line application.

Input

The input of the program is the definition of a given *problem* written in a text file.

Example 6.1. An example of a problem file written to solve Problem 6.1 is given in Appendix B.1, p. 199.

A problem is defined by

i) a set of invariants I;

ii) a class of graphs $\mathcal{C}$;

iii) a value's range for $n = n_{min}, \dots n_{max}$;

iv) and several options.

In Section 6.2.1, we explain which expressions of invariants are recognized by GraPHedron. These expressions can be used in the problem statement, reviewed in detail in Section 6.2.2.

Output

The output is a set of various files. Among them is a report written in LaTeX with

- a complete description of the systems $\mathcal{F}(\mathcal{P}_n)$ (for $n = n_{min}, \dots, n_{max}$);
- representations of $\mathcal{P}_n$ in $\mathbb{S}^2$, if the number of invariants is 2;
- drawings of the vertex graphs;
- automated conjectures, if available;
- optional other information.

Process

To produce this output, GraPHedron is made of the following 3 stages :

1. Data generation (graphs and invariants);
2. Polytopes computation;
3. Creation of the report and derivation of conjectures.

Details of each step are explained in Sections 6.2.3 to 6.2.5.

Finally, in Section 6.2.6, we give the current limitations of GraPHedron in time and storage. We also explain the choices made in the implementation, in order to overcome these limitations.

6.2.1 Algebraic expressions of invariants in GraPHedron

An invariant of I can be defined as an algebraic expression of other invariants. One can use three types of operators to build an expression. Arithmetic operators are used to construct *arithmetic expressions* and relational and logical operators allow the construction of *boolean expressions*. Boolean expressions are generally used to define $\mathcal{C}$ (see Section 6.2.2). Tables B.1 – B.3 in Appendix B.3 list the operators currently available in GraPHedron.

If an invariant is not constructed from any available operator, it is called a *basic invariant*.

If an expression E uses an invariant with an irrational value, or a constant written in floating notation, then E is called an *approximated expression*. If all expressions involved in a given problem are non-approximated, all computations are made in exact arithmetic using the *GMP* library [135]. Otherwise, the user has to be aware of possible round-off errors : all values are converted to floating point numbers (in double precision).

Example 6.2. The diameter D is a basic invariant, and the invariant defined by $\frac{D}{2}$ is an exact arithmetic expression.

6.2.2 Problem statement

A *problem* is defined from this information :

1. *Set of invariants* I. Each element of I is an algebraic expression of graph invariants[1], also called a *coordinate* of the problem. In the case of a boolean expression, its value is

[1] We recall that an algebraic expression of several graph invariants is also a graph invariant.

considered as an integer (true is 1 and false is 0). There exist currently about 50 invariants implemented in GraPHedron (see the list of invariants in Appendix B.4). The code is written in such a way that new invariants can be easily added.

2. *Class of graphs $\mathcal{C}$.* Graphs considered can be restricted to a specific class $\mathcal{C}$ in two ways :

   (a) *Selective generation.* GraPHedron uses, as a sub-routine, the graph generator *geng* of Mc Kay [192] (see Section 2.1). The program *geng* allows to restrict the graphs generated by fixing some parameters (e.g., minimum and maximum degree, minimum and maximum number of edges, ...). These parameters are fixed internally by GraPHedron when the user asks to generate graphs from one of the following class : trees, bipartite graphs, k-regular graphs, triangle-free graphs (graphs without K_3 subgraphs) or C_4-free graphs (graphs without cycles with 4 nodes). These classes can be combined. For instance, one can generate only bipartite cubic graphs, i.e. bipartite 3-regular graphs. For a given problem, we note $\mathcal{C}_n^{sel}$ the set of all non-isomorphic graphs with n nodes generated by selective generation.

   (b) *Graphs filtering.* The user can define a set of *conditions*, i.e., boolean expressions of invariants (see Section 6.2.1), that graphs should respect. Each graph enumerated by selective generation is tested and accepted only if all conditions are true for it. We note $\mathcal{C}_n^{fil}$ the set of all non-isomorphic graphs with n nodes which respect the conditions defined for a given problem.

   Remark 6.7. Each method has its advantage : selective generation allows to reach larger values for n and, as explained in Section 6.2.6, graphs filtering can be used to avoid recomputation of graphs and invariants values.

   Selective generation and graphs filtering are combined to define $\mathcal{C}$. For a given class $\mathcal{C}$ and a value of n, the set of graphs $\mathcal{C}_n$ is defined by

   $$\mathcal{C}_n = \mathcal{C}_n^{sel} \cap \mathcal{C}_n^{fil}.$$

3. *Order of graphs.* The user has to give the values n_{min} and n_{max}. The polytopes $\mathcal{P}_n$ for $n = n_{min}, \ldots, n_{max}$ are then computed in a following stage. An optional value n_{step} can be defined to go from n_{min} to n_{max} by steps of n_{step}. Acceptable values of n_{max} depend on the choice made in the selective generation. Some current limits are given in Section 6.2.6.

4. *Options.* Several options are available allowing to adapt, or to add information to, the output generated by the software. For instance, one can get statistics about the coordinate's values, as illustrated in Section 6.3.2.

6.2.3 Data generation

The data are generated as follows, for $n = n_{min}, \ldots, n_{max}$,

1. Let q be the number of basic invariants[2] used in the definition of I and $\mathcal{C}_n^{fil}$. We note these basic invariants $bi_1, bi_2, \ldots, bi_q$.

2. GraPHedron will create $1+q$ files.

   (a) The graphs of $\mathcal{C}_n^{sel}$ for $n = n_{min}, \ldots, n_{max}$ are generated by *geng* and stored in a binary format.

   Remark 6.8. For a given set $\mathcal{C}_n^{sel}$, a corresponding file `gra_C_n_sel` is created.

   Remark 6.9. A graph G is stored as a set of bytes in which a bit represents a boolean value of the upper triangle of the adjacency matrix. The number of bytes used for G is an integer and the bits which are not used are set to 0. Therefore , if G has n nodes, we need

   $$\left\lceil \frac{n(n-1)}{16} \right\rceil$$

   bytes to store G.

   (b) For each basic invariant bi_k for $k = 1, 2, \ldots q$, its values $bi_k(G)$ is computed for all $G \in \mathcal{C}_n^{sel}$ and stored in a binary file.

   Remark 6.10. For a given set $\mathcal{C}_n^{sel}$ and a basic invariant bi_k, a file `inv_bi_k_C_n_sel` is created.

3. These files are organized as follows. To each graph G, stored in the file `gra_C_n_sel`, is associated an *implicit reference*, which is simply its position *pos* in the file. In files `inv_bi_k_C_n_sel`, the values of the invariants $bi_k(G)$ are stored at the same position *pos*. As one knows the amount of space used to store a graph or a value of an invariant, it allows a direct access to a graph G and the values $i_k(G)$ if one knows the position *pos*. Moreover, in the subsequent steps, one can work with only the values bi_k and "forget" completely that a graph is associated with them.

4. These files are stored after the execution of GraPHedron for a given problem. If another problem uses again i_k and the same set $\mathcal{C}_n^{sel}$, the data can be directly reused from `gra_C_n_sel` and `inv_bi_k_C_n_sel`. This is very useful when computing bi_k takes a long time, as for NP-hard invariants. Moreover, experiences show that, in the whole process, the computation of invariants is the most time-consuming step when n grows, even for invariants with a polynomial complexity (see Section 6.2.6).

Example 6.3. If a first problem studies $I = (D, m)$, the files `inv_D_C_n_sel` and `inv_m_C_n_sel` are created. When studying a second problem $I = (\frac{D}{m}, \alpha)$, only the file `inv_alpha_C_n_sel` will be created.

[2] Recall that a basic invariant is an invariant which is not constructed with any arithmetic, relational or logical operator.

6.2.4 Polytopes computation

For each $n = n_{min}, \ldots, n_{max}$, the polytope $\mathcal{P}_n$ is computed as follows.

1. *Determine a set $\mathcal{S}_n$ of points.* Recall that $\mathcal{S}_n$ is defined by

$$\mathcal{S}_n = \{(x_1, x_2, \ldots, x_p) \in \mathbb{S}^p \mid \exists G \in \mathcal{C}_n \ : i_1(G) = x_1, i_2(G) = x_2, \ldots, i_p(G) = x_p\}.$$

   It is possible that several graphs are associated to the same point $(x_1, x_2, \ldots, x_p)$. To each point $(x_1, x_2, \ldots, x_p)$, GraPHedron associates

   (a) the set $\mathcal{G}$ of the positions (*pos*, see previous section) of the graphs G such that

$$x_1 = i_1(G),\ x_2 = i_2(G),\ \ldots,\ x_p = i_p(G),$$

   (b) the number of such graphs.

   Remark 6.11. The number of references stored in a point can be limited. These references are used later, e.g., to draw graphs corresponding to a given point. The number of graphs in $\mathcal{G}$ associated with a point is always computed.

   The construction of $\mathcal{S}_n$ is made as follows.

   (a) $\mathcal{S}_n$ is empty

   (b) For each position *pos*, the coordinate x_k is computed from the files containing the values of the basic invariants (see Section 6.2.3) used in i_k. The conditions (graph filtering) are also computed, again reading values of basic invariants at position *pos*, in corresponding files. If $v' = (x_1, x_2, \ldots, x_p)$ is accepted,

      i. If v' is already in $\mathcal{S}_n$, the set $\mathcal{G}$ of v' is updated and the number of graphs incremented;

      ii. otherwise, a new point is added in $\mathcal{S}_n$.

   Example 6.4. If $I = (\frac{D}{2}, m)$, the values of D and m are read in the files `inv_D_C_n_sel` and `inv_m_C_n_sel`. Then, x_1 is computed by dividing D by 2.

   Remark 6.12. When the dimension p is small and if the vectors often have same values, e.g., integers between 0 and n, the number of points in $\mathcal{S}_n$ is generally much smaller than the number of graphs in $\mathcal{C}_n$ (see Section 6.2.6).

2. *Compute the convex hull of $\mathcal{S}_n$.* This is done using *libcdd* of Fukuda [130]. Several algorithms exist to compute the convex hull of a set of points. See [13] for a survey on these algorithms and an efficiency comparison of the main existing tools. Avis et al. [13] conclude that there is no algorithm which is the best in all cases. Their efficiencies depend on the type of polytopes. Unfortunately, as we want to design a system which can be applied with any

set of invariants, we cannot know *a priori* the characteristics of the polytopes. However, our requirements are that the convex hull computation software should be able to manage a lot of points, which are not necessarily vertices of the polytope, and to deal with both exact arithmetic and floating arithmetic. This is the case of *libcdd*. Moreover, *libcdd* is a C/C++ library and thus avoids the call of an external software system.

Remark 6.13. In the same spirit as for the previous stage (data generation), if the system detects that a polytope has already been computed, it does not perform the computation again.

6.2.5 Report and conjectures

The output of GraPHedron is a report created with LaTeX. This report contains all the information computed. The polytopes $\mathcal{P}_n$ are described in both representations :

a) a system $\mathcal{F}(\mathcal{P}_n)$ of facets defining inequalities;

b) a set of vertices.

Example 6.5. See again Appendix B.1 or Appendix B.2 for examples of reports.

To derive conjectures, the user has to study the inequalities by hand or by interacting with the system where she/he can see and/or print the vertex graphs. In the case of two coordinates, the report contains a drawing of the polytopes. For three coordinates, the polytopes can be exported to be visualized and manipulated with *polymake* [133, 134]. In some cases, automated conjectures are found and presented in a specific section (see Section 6.5). The last section of the report contains tables with detailed information about CPU times needed to solve the problem.

Remark 6.14. All the files generated to create the report are reusable and written in classical formats, e.g., EPS files for figures.

6.2.6 Limits and performance

Table 6.1 records (estimations of) the resources needed in the case of general graphs (for which the limitation today is $n = 11$). These results were obtained on a Linux computer with a Pentium (R) IV (3 GHz) processor. Of course, for more restricted classes, the manageable number n of nodes can reach larger values. For instance, Table 6.2 refers to the class of trees.

Table 6.3 shows the CPU times – and the relative percentage of total time for each n – spent for each step in the computation of the problem posed in Problem 6.1. The class of connected graphs was specified here using graph filtering. There are 3 invariants which have polynomial complexities : $m(G)$, $D(G)$ and the boolean invariant $c(G)$ which is true only if G is connected.

Table 6.1: Needed resources for general graphs

n	# of graphs	CPU time	File size (in Kb)
5	34	0.00 sec	0.17
6	156	0.00 sec	0.41
7	1044	0.00 sec	3.16
8	12 346	0.02 sec	48.33
9	274 668	0.45 sec	1341.26
10	12 005 168	21.38 sec	7.034×10^4
11	1 018 997 864	27 min 14.68 sec	6.966×10^6
12	1.65×10^{11}	(estimated) 3.4 days	(computed) 1.451×10^9
13	5.05×10^{13}	(estimated) 3.3 years	(computed) 4.932×10^{11}

Table 6.2: Needed resources for trees

n	# of trees	CPU time	File size (Kb)
5	3	0.00 sec	0.11
10	106	0.00 sec	0.72
15	7 741	1.79 sec	105.94
20	823 065	60 min 29.74 sec	1.93×10^4

It appears often that, when n grows, the relative percentage of time that can be saved by precomputation (graph generation and invariants computation) is increasing. It is thus efficient in time to keep the data computed in files. This observation is reinforced with invariants hardest to compute. Of course, this feature has a cost in terms of storage. Fortunately, the compression of invariants data files greatly reduces the space needed for them.

Table 6.3: CPU Times (in sec.) for each step when solving Problem 6.1

Value of n	**7**	**8**	**9**	**10**
Graph generation	0.00 (0%)	0.02 (5.4%)	0.45 (5.1%)	21.38 (4.9%)
Computation of c	0.00	0.04	1.07	56.99
Computation of D	0.01	0.07	2.10	118.98
Computation of m	0.00	0.02	0.55	32.02
Total invariants	0.01 (25.0%)	0.13 (35.1%)	3.72 (42.0%)	207.99 (47.8%)
Computation of $\mathcal{P}_n$	0.03 (75.0%)	0.22 (59.5%)	4.69 (52.9%)	205.95 (47.3%)
Total time	0.04	0.37	8.86	435.32

Remark 6.15. Generation of graphs and computation of invariants can be separated in a finite number of parts : either separating the computation for different values of n, or even splitting computation for a given n. It allows to make computation in parallel and to keep the size of files under the limits of operating systems.

6.3 Applications

The main application of GraPHedron is to find optimal conjectures in graph theory, as illustrated with Problem 6.1. This example has led to Conjecture 6.1, which is proven in Section 6.3.1. Chapter 7 is devoted to other theoretical results obtained with GraPHedron.

Other possible applications are presented in Section 6.3.2.

6.3.1 Discover new and optimal conjectures

Proof of Conjecture 6.1

In order to prove Conjecture 6.1 (p. 136), we introduce the following lemma.

Lemma 6.1. *For each connected graph G with $n \geq 3$ nodes, m edges and a diameter $2 \leq D \leq n-1$,*

$$2m \leq D^2 - D(2n+1) + (n-1)(n+4) \tag{6.10}$$

Proof. Let $G^*(n,D)$ be a graph with a fixed number of nodes $n \geq 3$, a fixed diameter $2 \leq D \leq n-1$ and constructed from two disjoint subgraphs : a path P_{D+1} and a clique K_{n-D-1}. Three *consecutive* nodes of P_{D+1} are adjacent to all nodes of K_{n-D-1}. There are no other edges between the path and the clique. Such a graph is depicted in Figure 6.5. Note that, for given values n and D, $G^*(n,D)$ is generally not unique since one can move, through the path, the 3 consecutive nodes adjacent to the clique. However, when D is equal to 2 or $n-1$, $G^*(n,D)$ is unique since $G^*(n,2) \simeq K_n \backslash e$ and $G^*(n,n-1) \simeq P_n$.

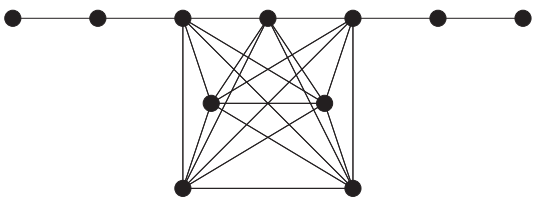


Figure 6.5: A graph $G^*(11,6)$: $n = 11$, $D = 6$, $K_{n-D-1} = K_4$ and $P_{D+1} = P_7$

Graphs $G^*(n,D)$ have a maximum number of edges among all graphs with n nodes and a diameter D. Indeed, adding one edge in $G^*(n,D)$ will inevitably reduce D by creating a shortcut in the path P_{D+1}.

It follows that

$$m(G) \leq m(G^*(n,D)), \tag{6.11}$$

for all graphs G with n nodes and a diameter D. By construction the number of edges of $G^*(n,D)$ is equal to

$$m(G^*(n,D)) = D + \binom{n-D-1}{2} + 3(n-D-1),$$

which can be simplified by collecting D :

$$m(G^*(n,D)) = \frac{1}{2}\left[D^2 - D(2n+1) + (n-1)(n+4)\right]. \tag{6.12}$$

Eqs. (6.11) and (6.12) give the result. □

Remark 6.16. Lemma 6.1 expresses the values of the points which are *undominated* but do not lie on a facet. Recently, we added to GraPHedron a detection of such points, also called *Pareto points* (see Appendix B.1)

Conjecture 6.1 can now be reformulated as a theorem.

Theorem 6.1. *For each connected graph G with $n \geq 4$ nodes, m edges and a diameter D,*

$$\begin{aligned} m &\geq n-1, && (6.13)\\ \frac{(n-1)(n-2)}{2}D + m &\geq (n-1)^2, && (6.14)\\ D + m &\leq \frac{n(n-1)}{2} + 1, && (6.15)\\ nD + 2m &\leq (n+2)(n-1). && (6.16) \end{aligned}$$

and these inequalities are the only possible facets defining inequalities for each n.

Proof. Inequality (6.13) is valid since we consider only connected graphs of order n. The minimum number of edges is thus $n-1$ which arises only for trees. The diameter of trees varies from 2 for the star S_n to $n-1$ for the path P_n. The corresponding vectors (D,m), namely $(2,n-1)$ and $(n-1,n-1)$ are linearly independent when $n \geq 4$. It follows that (6.13) is facet defining.

If $D(G) = 1$, then $G \simeq K_n$. As $m(K_n) = \binom{n}{2}$, the two sides of (6.14) are equal and this inequality is valid. If $D(G) \geq 2$, and because $m \geq n-1$, we have

$$\frac{(n-1)(n-2)}{2}D + m \geq \frac{(n-1)(n-2)}{2}2 + n - 1 = (n-1)^2,$$

and (6.14) remains valid. When $n \geq 3$, vectors (D,m) corresponding to K_n and S_n are linearly independent and satisfy (6.14) with equality. Inequality (6.14) is thus facet defining.

Remark that if $D < 2$, Inequality (6.16) is dominated by Inequality (6.15). It is the contrary when $D > 2$ and these two inequalities are equal when $D = 2$. We can thus assume that $D \leq 2$ (resp. $D \geq 2$) to prove that Inequality (6.15) (resp. (6.16)) is facet defining. To prove their validity, one has to answer the following question : what is the maximum number of edges in connected graphs with fixed number of nodes and diameter ?

If $D \leq 2$, the candidate graphs are trivially K_n and $K_n \backslash e$ which have linearly independent vectors (D,m) if only $n \geq 3$. Inequality (6.15) is thus facet defining.

Lemma 6.1 gives the answer when $D \geq 2$. It is straightforward to check that, for a fixed number of nodes n, Inequality (6.10) is dominated by Inequality (6.16) when $2 < D < n-1$. These two inequalities intersect when $D = 2$ and $D = n-1$, at points $\left(2, \binom{n}{2} - 1\right)$ and $(n-1, n-1)$. These points correspond to $K_n \backslash e$ and P_n respectively and are again linearly independent when $n \geq 4$. Inequality (6.16) is thus also a inequality defining a facet.

Finally, remark that, for all values of n and D, we have characterized all the graphs maximizing (graphs $G^*(n,D)$) and minimizing (trees) the number of edges. We have shown that their coordinates either satisfy one or two inequalities of Theorem 6.1, or are dominated by them. It follows that the description of the polytopes is complete for each n. □

Other suggested conjectures

Chapter 7 is entirely devoted to results obtained with the help of GraPHedron. Among these results, we give an answer to the following question.

Problem 6.2. What is the minimum number of edges m in connected graphs with a fixed number of nodes n and a fixed stability number α ?

This question was listed as an open problem in Ore [202] since 1962. It constitutes a variant of a famous theorem of Turán [228], which is applied when graphs are not necessarily connected. A complete answer to Problem 6.2, obtained with the help of GraPHedron is given in Theorem 7.2, which is presented and proven in Section 7.1.

Remark 6.17. In Problem 6.2, $I = \{\alpha, m\}$ and $\mathcal{C}$ is the class of connected graphs. The report created by GraPHedron for this problem is presented in Appendix B.2 (pages 217 to 240). The solution for this problem is studied in Section 7.1.

Another illustration of this strategy to obtain new results in graph theory is also given in Section 7.2. In this section, our methodology is applied to find relations, on connected graphs with n nodes, between three invariants : the maximum degree $\Delta(G)$, the irregularity $A(G)$ and the diameter $D(G)$.

6.3.2 Other applications

Here are some other applications of GraPHedron :

a) *Check existing conjectures and theorems.* One can submit to GraPHedron an existing conjecture C or theorem T expressed as an inequality (or an equality) among p invariants. The system can then be used in two ways :

   1) *Check for validity and optimality.* This can be done without computing the polytopes. Browsing the graphs of $\mathcal{C}_n$, one can check if :

      - C is *rejected* since a counter-example exists;
      - C is *valid*;
      - C or T is *tight* by exhibiting a graph satisfying the conjecture or theorem with equality;
      - C or T is *facet defining* by exhibiting a set of p graphs, satisfying the conjecture or theorem with equality, and corresponding to p linearly independent vectors of coordinates.

   2) *Extract invariants and compute associated polytopes.* A conjecture or a theorem which is not facet defining in all computed polytopes means that one can maybe improve it. Applying the strategy described in Section 6.1 can lead to a better formulation. Otherwise, showing that an existing theorem is facet defining is a strong criterion of quality.

b) *Help for theorem-proving.* Because the vertex graphs are extremal for a given problem, their knowledge is of great help when one has to prove the conjectures derived with or by the system. Moreover, if their coordinates are characterized, one directly obtains a set of linearly independent vectors. This is intensively used in the proofs, see for instance the proof of Theorem 6.1 and the proofs of Chapter 7.

   The representation of the polytopes is useful too when one would like to detect if an inequality is dominated by another.

c) *Compare values of heuristics and exact algorithms.* Let $h^*(G)$ be the worst-case value obtained when a heuristic H is applied to G. Define an invariant by dividing $h^*(G)$ by the value obtained by an exact algorithm. This invariant represents a factor of approximation. This approach is illustrated in Section 7.3 : the maximal-matching heuristic provides a 2-approximation of two classical NP-hard problems (*minimum node cover* and *minimum maximal matching*) [132]. Using GraPHedron, finer worst-case approximation factors were obtained, under some assumptions on the density of the graphs.

d) *Education.* The system is easy to use and some well-chosen problems, such as Problem 6.1, can be used by students. Information is displayed in a way which makes basic notions of graph theory and polyhedral theory handy to illustrate. Of course, the proofs of conjectures – which can appear quite easy – is another challenge. This provides an exciting way to be initiated to the world of research in graph theory.

e) *Get information about the points distribution.* One can ask, as an option, that the system adds the distribution of the points of S, corresponding to graphs, *inside* the polytopes. It can be

of interest as sometimes a vast majority of graphs are far away from some facets. Figure 6.6 shows the distribution inside the polytope $\mathcal{P}_9$ derived from Problem 6.2. A white point means that there is only one graph corresponding to this coordinate. An increasing level of gray shows the growing frequency of graphs sharing the same coordinates. See also, in Appendix B.2, the report created to solve Problem 6.2.

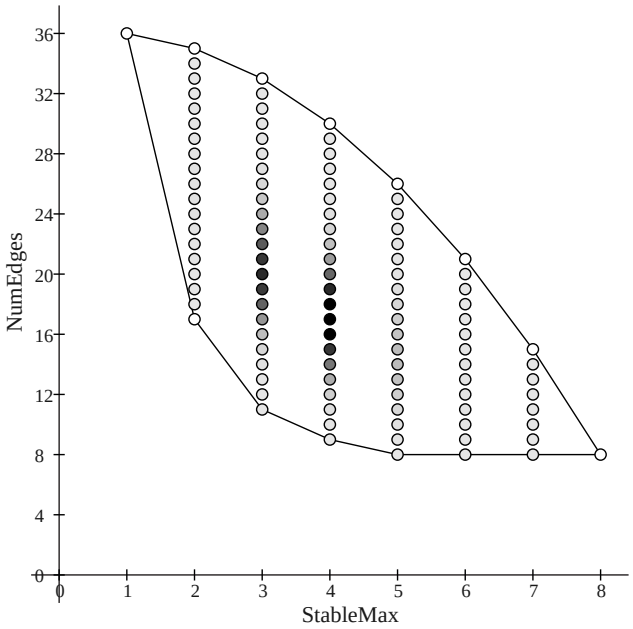


Figure 6.6: Distribution of points in $\mathcal{P}_9$ for Problem 6.2

f) *Get information about invariants.* One can ask, as an option, that the system computes, during its process, various statistics about the coordinates of the problem. See for instance Appendix B.2, in which a GraPHedron's report contains statistics about coordinates.

6.4 Forms of conjectures

Basic form of conjectures

When n is fixed, facets are linear inequalities among p invariants. A generalization of these inequalities in terms of n leads to a first basic form of conjectures :

$$f_1(n)i_1 + f_2(n)i_2 + \cdots + f_p(n)i_p \leq f_0(n) \tag{6.17}$$

where $f_i(n)$ are functions, linear or not, of n. Theorem 6.1 is an example of such conjectures where coefficients $f_i(n)$ are linear or quadratic functions. It follows that (6.17) is linear for each n, i.e., when n is fixed, but can be non linear when n is taken as a parameter.

Families of inequalities

In Section 7.1, we apply GraPHedron to the set of invariants $I = \{\alpha, m\}$ for connected graphs with n nodes (see also the report created by GraPHedron for this problem in Appendix B.2). In this case, it appears that the number of facets of the polytopes is no longer constant. However, one can point out *families of inequalities*. For instance, the rightmost polytope edges (see Eq. (7.4) in Section 7.1.2) can be expressed as $n-2$ inequalities of the form (6.17) :

$$m \leq \binom{n-k}{2} + k(n-\alpha) \qquad \text{for } k = 1, 2, \ldots, n-2. \tag{6.18}$$

When a family of inequalities occurs to describe (partially) the polytopes $\mathcal{P}_n$, the corresponding vertex graphs of $\mathcal{P}_n$ are often also of a specific family. For instance, the $n-1$ graphs which are vertex graphs of the $n-2$ facets described by (6.18) are complete split graphs (see again Appendix B.2). Recall that a complete split graph $CS^{\alpha}_{n-\alpha}$ with $1 \leq \alpha \leq n$ is constructed from an independent set of size α and a clique of size $n-\alpha$. Each node in the independent set is adjacent to each node in the clique. By construction,

$$m(CS^{\alpha}_{n-\alpha}) = \binom{n-\alpha}{2} + (n-\alpha)\alpha = \frac{1}{2}\left[n(n-1) + \alpha(\alpha-1)\right]. \tag{6.19}$$

There is exactly one such graph for each possible value of α, i.e., for the integers $\alpha = 1, 2, \ldots, n$. It follows that the family of inequalities (6.18) can be reformulated, without loss of generality, as

$$m \leq \frac{1}{2}\left[n(n-1) + \alpha(\alpha-1)\right], \tag{6.20}$$

which is non-linear in α, even if n is fixed. Figure 6.7 shows the representations in the plane of (6.18), the black lines, and (6.20), the dotted curve, when $n = 5$.

We present other families of inequalities in the next chapter. Their formulations are not always polynomial. For instance, another family of inequalities is given later (see Eqs. (7.10) and (7.19)). In this case these functions are not polynomial as the ceil operator is used.

Therefore, forms of conjectures that can be derived, when families of inequalities appear, are not limited to the form (6.17).

6.5 Automated conjectures

In some cases, for instance when

i) the number of facets of $\mathcal{P}_n$ is constant independently of n;

ii) and the number p of invariants in I is 2;

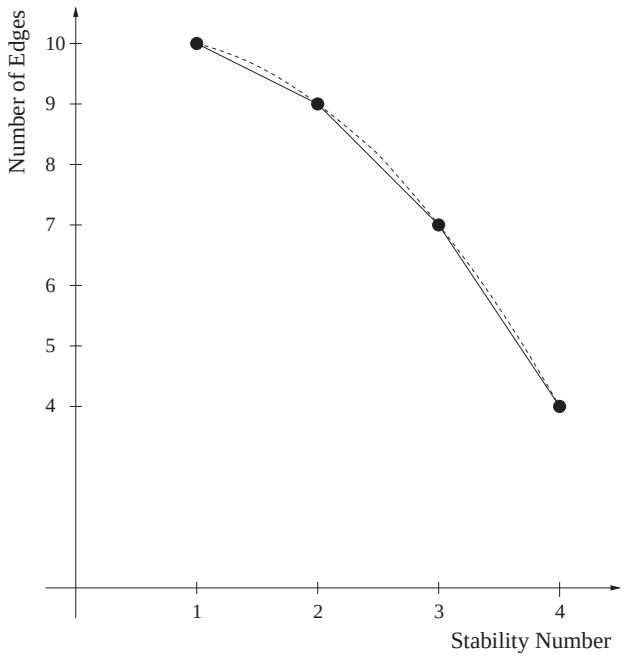


Figure 6.7: Family of inequalities (6.18) and its quadratic generalization (6.20), when $n = 5$

the system is able to generate conjectures in a fully automated way.

For instance, Conjecture 6.1 is automatically detected and added to the report (see Appendix B.1). In Section 6.1.3, we used a characterization of the vertex graphs to derive Conjecture 6.1. GraPHedron uses a different method. It tries to detect similarities between facets, analyzing the values of their coefficients. Then, it generalizes similar facets in terms of n. This recognition works effectively if the coefficients $f_k(n)$ are polynomials in n with a degree ≤ 6. Other functions $f_k(n)$ can also be detected (see [161]).

Automation in more general cases, e.g., when the number of facets is increasing when n grows, are under study [161]. First types of automated conjectures were presented, very recently, in the communication [195].

6.6 Summary

We have presented a formal framework, supported by a computer system, allowing to identify facet defining inequalities among graph invariants. These linear inequalities can be considered as optimal when n is given. It appears that this strategy is efficient and leads to unexplored forms of conjectures. From detailed information given in the report created by GraPHedron, it is often easy to derive conjectures by observation. However, there is much work to do to get conjectures in a fully automated way, but first results seem to be promising.

The proposed approach can also be used to check existing conjectures and theorems, to help in proving conjectures, to study approximation ratios, in education, to get the spatial distribution of the graphs inside the polytopes and to get statistics about invariants.

To avoid re-computation of graphs and invariants, the system is quite greedy in storage. Therefore, we think that a web portal, on a dedicated server, allowing to use GraPHedron with much pre-computed data, is preferable to a downloadable version of the software. This web site is under construction [136].

CHAPTER 7

First results

Stability number and number of edges of connected graphs • Irregularity, diameter and maximum degree of connected graphs • Tight analysis of the maximal matching heuristic • Summary

In the previous chapter we have introduced the new computer system GraPHedron, based on a polyhedral approach. In Section 6.3.1, we have given all optimal linear inequalities among the diameter D and the number of edges m, for connected graphs with a fixed number of nodes n. In this chapter, we present other results, surely more important than the study of (D,m).

In Section 7.1, all optimal linear inequalities among the stability number α and the number of edges – for connected graphs with a fixed number of nodes n – are given. This study gives an answer to Problem 6.2 quoted in the previous chapter (p. 147).

A similar study is made for a set of three invariants in Section 7.2 : the irregularity A, the maximum degree Δ and the diameter D of connected graphs. We have established some of the optimal inequalities among this set of invariants. However, the problem of finding all optimal linear inequalities for the diameter, the irregularity and the maximum degree is left open.

Finally, in Section 7.3, we use GraPHedron to compare an heuristic and exact algorithms. Indeed, we study the worst-case approximation ratio of the maximal matching heuristic for the minimum node cover and minimum maximal matching problems, as function of density parameters.

7.1 Stability number and number of edges of connected graphs

The general methodology outlined in Chapters 6.1 and 6.2 is applied here to the case of two invariants, the stability number and the number of edges. Fixing the number n of nodes, we associate to any connected graph $G = (V,E)$ on n nodes the pair (α, m) where $\alpha = \alpha(G)$ and $m = m(G)$. Our goal is to determine, for any n, all facet defining inequalities for the convex polygon

$$\mathcal{P}_n^{\mathcal{C},\{\alpha,m\}} = \text{conv}\big\{(\alpha,m) | \exists G \in \mathcal{C}_n,\ \alpha(G) = \alpha,\ m(G) = m\big\}, \tag{7.1}$$

where $\mathcal{C}$ is the class of connected graphs.

Remark 7.1. In the following $\mathcal{P}_n^{\mathcal{C},\{\alpha,m\}}$ is denoted by $\mathcal{P}_n^{\alpha,m}$

The polygon $\mathcal{P}_n^{\alpha,m}$ lies in a plane $\mathbb{R}^2$ having two coordinates x_α and x_m. For $n = 10$, it is illustrated in Figure 7.1.

Remark 7.2. The report created by GraPHedron to solve this problem is presented in Appendix B.2.

Figure 7.1, together with similar figures for other values of n (see Appendix B.2), suggests to sort out the polytope edges of $\mathcal{P}_n^{\alpha,m}$ in three families. These families respectively consist of: (i) one horizontal polytope edge lying on the line $x_m = n-1$; (ii) polytope edges to the right, forming a path from vertex $(n-1, n-1)$ to vertex $\left(1, \binom{n}{2}\right)$; (iii) polytope edges to the left, connecting vertex $(\lfloor \frac{n+1}{2} \rfloor, n-1)$ to vertex $\left(1, \binom{n}{2}\right)$. Our analysis treats these three families one after the other, and establishes in each case the complete list of polytope edges together with the corresponding facet defining inequalities. To avoid trivialities, we assume $n \geq 4$ in this section.

7.1.1 The horizontal polytope edge

Since we consider only connected graphs on n nodes, the minimum number of edges in such a graph equals $n-1$, a value which arises exactly for trees. The stability number of such trees varies from $\lfloor \frac{n+1}{2} \rfloor$ (for a path on n nodes) to $n-1$ (for a star $K_{1,n-1}$). There results a horizontal polytope edge for $\mathcal{P}_n^{\alpha,m}$, with vertices $(\lfloor \frac{n+1}{2} \rfloor, n-1)$ and $(n-1, n-1)$. The corresponding facet defining inequality is of course $x_m \geq n-1$.

7.1.2 The rightmost polytope edges

The following proposition handles the second family of polytope edges.

Proposition 7.1. *For $k = 1, 2, \ldots, n-2$, the inequality*

$$k x_\alpha + x_m \leq \binom{n-k}{2} + kn \tag{7.2}$$

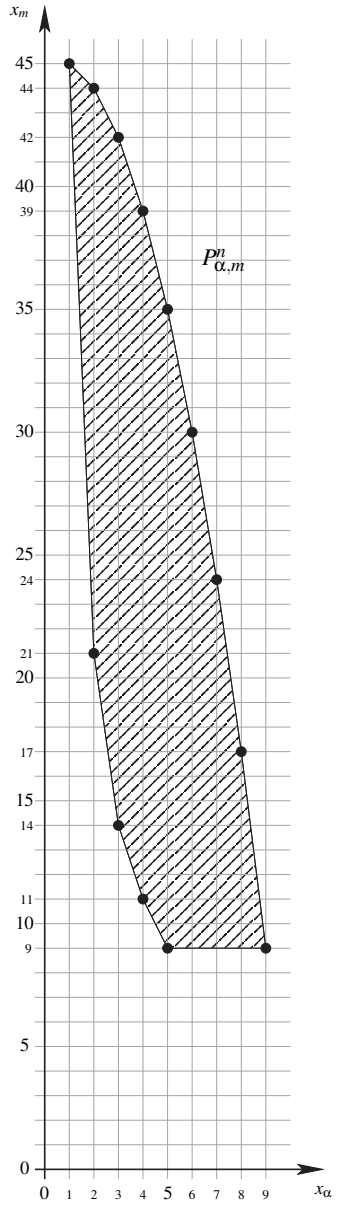


Figure 7.1: The polygon $\mathcal{P}^n_{\alpha,m}$ for $n = 10$.

defines a polytope edge of $\mathcal{P}^{\alpha,m}_n$ *with vertices*

$$\left(k, \binom{n-k}{2} + k(n-k)\right) \text{ and } \left(k+1, \binom{n-k}{2} + k(n-k-1)\right). \tag{7.3}$$

All polytope edges and vertices of $\mathcal{P}^{\alpha,m}_n$ *to the right of some point of* $\mathcal{P}^{\alpha,m}_n$ *are of these types.*

There results the sequence of vertices $\left(k, \binom{n-k}{2} + k(n-k)\right)$ for $k = 1, 2, \ldots, n-1$, which starts at $\left(1, \binom{n}{2}\right)$ and ends at $(n-1, n-1)$. Two successive vertices in the sequence have their abscissas differing by 1. Hence $\mathcal{P}^{\alpha,m}_n$ has no other vertex on the rightmost part of its boundary.

Remark 7.3. Note that these vertex graphs are complete split graphs (see Appendix B.2). This family of graphs appears often in this dissertation (see Chapter 3 and Sections 5.2 and 7.3).

Proof. We first establish that Inequality (7.2) is valid for each vertex of $\mathcal{P}_n^{\alpha,m}$. With α denoting the stability number of a graph $G=(V,E)$ for which $|V|=n$ and $|E|=m$, this amounts to prove

$$m \leq \binom{n-k}{2} + k(n-\alpha). \tag{7.4}$$

We proceed by considering two cases.

a) Case $k \leq \alpha$. In some stable set S of maximum size in G, select some subset T of k nodes. The number of edges disjoint from T is at most $\binom{n-k}{2}$. Any other edge has exactly one node in T; its other node lies outside S. The number of such edges is at most $k(n-\alpha)$. Summing up, we get Inequality (7.4).

b) Case $k > \alpha$. Select a set U of k nodes which contains a maximum-size stable set S. Any edge is either (i) disjoint from U, or (ii) formed by one node in S and the other one in $V \setminus S$, or (iii) formed by one node in $U \setminus S$ and the other one in $V \setminus S$. Thus the total number of edges is at most $\binom{n-k}{2} + \alpha(n-\alpha) + (k-\alpha)(n-\alpha)$, from which Inequality (7.4) follows.

We thus have proved that Inequality (7.2) holds for $\mathcal{P}_n^{\alpha,m}$.

Each of the two points given in (7.3) comes from at least one connected graph, namely the graph having as edges all those pairs of nodes not included in a fixed subset of k, resp. $k+1$, nodes. As easily checked, these two points satisfy Inequality (7.2) with equality, and the same for a second linear inequality also valid for the polytope. Consequently, these two points are the vertices lying on the polytope edge defined by Inequality (7.2). Moreover, there can be no other polytope edge because we have found vertices with abscissas increasing by step of 1 from 1 to $n-1$. □

We point out that another proof of Proposition 7.1 can be forged by first finding out vertices, which means: (i) determining the maximum number of edges in a connected graph with given number of nodes and given stability number; (ii) showing that the resulting points in $\mathbb{R}^2$ are convexly independent. The easy solution to (i) states that the maximum number of edges equals $(n-\alpha)(n+\alpha-1)/2$. As this function of α is strictly concave, (ii) becomes trivial. Inequality (7.2) can then be derived. This line of argument produces also a nonlinear inequality attributed in [158] to Tomescu :

$$\alpha \leq \left\lfloor \frac{1}{2} + \sqrt{\frac{1}{4} + n(n-1) - 2m} \right\rfloor$$

7.1.3 The leftmost polytope edges

For the third and last family of polytope edges, we first infer the corresponding vertices from the answer to the following question: what is the minimum number of edges in connected graphs with fixed number of nodes and fixed stability number? This question is listed as an open problem in

Ore [202]. A complete answer is contained in Theorem 7.2 below. It constitutes the variant for connected graphs of a celebrated result of Turán [228], which we now recall. For given integer numbers n and α satisfying $n \geq \alpha \geq 1$, the Turán graph $T(n,\alpha)$ has n nodes and is the disjoint union of α cliques with balanced sizes (i.e., sizes equal to $\lfloor \frac{n}{\alpha} \rfloor$ or $\lceil \frac{n}{\alpha} \rceil$; the last two expressions give the same value if and only if α divides n). We let $t(n,\alpha)$ denote the number of edges in the Turán graph $T(n,\alpha)$.

Theorem 7.1. [228] *Any graph on n nodes with stability number α has at least $t(n,\alpha)$ edges. Moreover, this graph has exactly $t(n,\alpha)$ edges iff it is (isomorphic to) the Turán graph $T(n,\alpha)$.*

Thus the only graph with minimum number of edges in Theorem 7.1 has α connected components. Of course, by adding $\alpha - 1$ carefully selected edges, we obtain various connected graphs on n nodes and stability number α. Notice that in case $n = 2\alpha + 1$, odd cycles on n nodes still provide other examples with the same values of α and m . Despite their diversity (as illustrated by GraPHedron, see again Appendix B.2), we are able to prove that all these connected graphs have the minimum possible number of edges for given n and α. The proof is more involved than the simple ones known for Turán result (and exposed e.g. in Bollobás [35]). Also, it covers more cases than the connected one (which is obtained in Theorem 7.2 below for $c = 1$, see Corollary 7.1).

Theorem 7.2. *Any graph G on n nodes with stability number α and with c connected components has at least $t(n,\alpha) + \alpha - c$ edges. The lower bound is tight in all cases.*

For $1 \leq c \leq \alpha \leq n$, let $f_c(n,\alpha)$ be the minimum number of edges for graphs as in Theorem 7.2. To prove $f_c(n,\alpha) = t(n,\alpha) + \alpha - c$, we first establish two lemmas (a third one comes later).

Lemma 7.1. *Suppose $n \geq 2\alpha$. If $f_1(n,\alpha) = t(n,\alpha) + \alpha - 1$, then for $c = 1,2,\ldots,\alpha$, we have $f_c(n,\alpha) = t(n,\alpha) + \alpha - c$.*

Proof. The assumption $n \geq 2\alpha$ implies that each maximal clique of the Turán graph $T(n,\alpha)$ has more than one node. By adding $\alpha - c$ carefully selected edges to $T(n,\alpha)$, we see

$$t(n,\alpha) + \alpha - c \geq f_c(n,\alpha),$$

and similarly by adding $c-1$ carefully selected edges to a graph realizing $f_c(n,\alpha)$,

$$f_c(n,\alpha) + c - 1 \geq f_1(n,\alpha).$$

Thus

$$t(n,\alpha) + \alpha - c \geq f_c(n,\alpha) \geq f_1(n,\alpha) - (c-1). \tag{7.5}$$

By our assumption, the first and last expressions in Equation (7.5) are equal. The thesis follows. □

Lemma 7.2. *For $1 \leq \tilde{\alpha} \leq \alpha \leq n$, we have*

$$t(n,\tilde{\alpha}) \geq t(n,\alpha) + \alpha - \tilde{\alpha}. \tag{7.6}$$

Proof. We need consider only the case $\alpha = \tilde{\alpha}+1$. To transform $T(n,\tilde{\alpha}+1)$ into $T(n,\tilde{\alpha})$, we pick a clique of size $\lfloor \frac{n}{\tilde{\alpha}+1} \rfloor$ in the first graph, delete its nodes and add $\lfloor \frac{n}{\tilde{\alpha}+1} \rfloor$ nodes in all to the other cliques. Clearly, the number of edges increases at each deletion/addition of a single node. □

Proof of Theorem 7.2. By adding $\alpha - c$ selected edges to $T(n,\alpha)$, we see that inequality $f_c(n,\alpha) \leq t(n,\alpha)+\alpha-c$ holds. Hence, there only remains to prove the following inequality:

$$f_c(n,\alpha) \geq t(n,\alpha)+\alpha-c.$$

Assume first $n < 2\alpha$. Then $t(n,\alpha) = n-\alpha$ because the Turán graph then consists of $2\alpha - n$ isolated nodes plus $n-\alpha$ parallel edges. On the other hand, any graph on n nodes has at least $n-c$ edges, thus $f_c(n,\alpha) \geq n-c = t(n,\alpha)+\alpha-c$, which gives the result.

Assume now $n \geq 2\alpha$. By Lemma 7.1, we need to establish the thesis only for G connected. There remains to prove

$$f_1(n,\alpha) \geq t(n,\alpha)+\alpha-1, \tag{7.7}$$

by recurrence on the number n of nodes.

Let S be a maximum stable set of G, and let b be the number of connected components of $G \setminus S$ (thus $b \leq \alpha$).

By Lemma 7.3 below, the number of edges in the cut $\delta(S)$ is at least $n-\alpha+b-1$. On the other hand, G induces on $V \setminus S$ a graph with b components and stability number $\tilde{\alpha}$ satisfying $\tilde{\alpha} \leq \alpha < n-\alpha$. By the induction assumption together with Lemma 7.1, the graph induced on $V \setminus S$ has at least $f_b(n-\alpha,\tilde{\alpha}) = t(n-\alpha,\tilde{\alpha})+\tilde{\alpha}-b$ edges. Summing up, we get

$$m \geq n-\alpha+b-1+t(n-\alpha,\tilde{\alpha})+\tilde{\alpha}-b$$

and then by Lemma 7.2

$$\begin{aligned} m &\geq n-\alpha+b-1+t(n-\alpha,\alpha)+\alpha-\tilde{\alpha}+\tilde{\alpha}-b \\ &= n-\alpha+t(n-\alpha,\alpha)+\alpha-1 \\ &= t(n,\alpha)+\alpha-1 \end{aligned}$$

(the last equality directly derives from the structure of Turán graphs: the adjunction of a node to each of the α maximal cliques of $T(n-\alpha,\alpha)$ produces $T(n,\alpha)$). Inequality (7.7) is thus proved.

The last assertion in Theorem (7.2) is correct in view of the examples provided just before the statement. □

Lemma 7.3. *As in the proof of Theorem 7.2, let $G=(V,E)$ be a connected graph with n nodes, m edges and stability number α. Take a maximum stable set S in G, and let C_1, C_2, ..., C_b be the b*

connected components of $G \setminus S$. Then

$$|\delta(S)| \geq n - \alpha + b - 1. \tag{7.8}$$

Proof. For $i = 1, 2, \ldots, b$, build a spanning tree of C_i. Then extend the union of these b (sub)trees into a spanning tree T of G. Clearly, $E(T) \cap \delta(S)$ is an acyclic set of edges of G which covers S and at least one node in each C_i (maybe several nodes in a given C_i). Denote by G' the graph $(V, E(T) \cap \delta(S))$ and let $A_1, A_2, \ldots, A_\ell$ be the connected components with more than one node of G'. For $j = 1, 2, \ldots, \ell$, denote by s_j (resp. r_j) the number of nodes of A_j in S (resp. not in S). Thus, the subgraph of G' induced by A_j is a tree with $s_j + r_j$ nodes (see Figure 7.2 for an illustration).

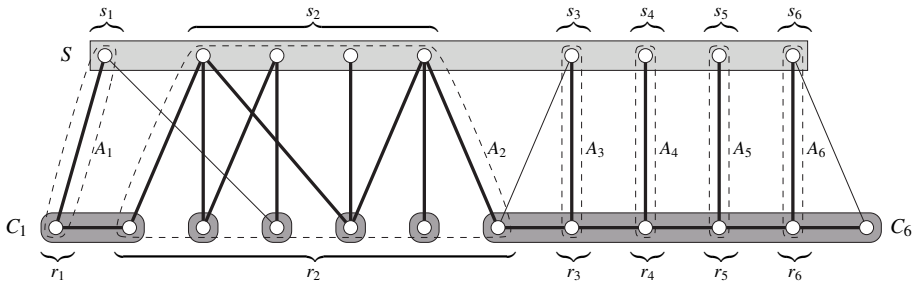


Figure 7.2: An example of graph G for the proof of Lemma 7.3. The edges in $E(T)$ are shown in bold lines.

For any fixed j, two nodes in $A_j \setminus S$ cannot belong to the same C_i (otherwise there would be a cycle formed by edges of the tree T). We derive

$$\sum_{j=1}^{\ell} (r_j - 1) = b - 1 \tag{7.9}$$

by repeatedly collapsing nodes as follows. Say A_j and $A_{j'}$ have each a node in a same C_i. Then by collapsing these two nodes we collect A_j and $A_{j'}$ together while leaving the left-hand side of Equation (7.9) unchanged because $(r_j - 1) + (r_{j'} - 1) = (r_j + r_{j'} - 1) - 1$. All such collapsings done, there remains in the left-hand side of Equation (7.9) only one term which equals $b - 1$ because exactly one node remains in each of the b sets C_i's.

We now establish a lower bound on the number p_j of edges of G connecting the r_j nodes in $A_j \setminus S$ to (whatever) nodes of S. By connectivity of A_j in the graph G', we surely have $p_j \geq s_j + r_j - 1$. Moreover $A_j \setminus S$ is stable in G (because as shown before $A_j \setminus S$ cannot have two nodes in a same C_i), so there must be at least $r_j - s_j$ nodes in the maximum stable set S which are adjacent in G to at least one node in $A_j \setminus S$. Then $p_j \geq (r_j + s_j - 1) + (r_j - s_j) = 2r_j - 1$.

Consider now any node u of G not covered by $E(T) \cap \delta(S)$, thus $u \notin S$. As $S \cup \{u\}$ cannot be stable, there is some edge connecting u to S.

Finally, any node outside S is either covered by $E(T) \cap \delta(S)$ (and then, for some j, belongs to A_j) or is not covered (and then is adjacent to some node in S). Using also $p_j \geq 2r_j - 1$ and Equation (7.9), we get

$$\begin{aligned} |\delta(S)| &\geq \left(\sum_{j=1}^{\ell}(2r_j-1)\right) + \left(n-\alpha-\sum_{j=1}^{\ell} r_j\right) \\ &= n-\alpha+\sum_{j=1}^{\ell}(r_j-1) \\ &= n-\alpha+b-1. \end{aligned}$$

□

The most important case in Theorem 7.2 comes when $c = 1$. In view of the following explicit form of the Turán number (see e.g. [28])

$$t(n,\alpha) = \left(\left\lceil \frac{n}{\alpha} \right\rceil - 1\right) \cdot \left(n - \frac{\alpha}{2}\left\lceil \frac{n}{\alpha} \right\rceil\right), \tag{7.10}$$

we thus have proved the next result.

Corollary 7.1. *The minimum number of edges for all connected graphs on n nodes with stability number* α *equals*

$$\left(\left\lceil \frac{n}{\alpha} \right\rceil - 1\right) \cdot \left(n - \frac{\alpha}{2}\left\lceil \frac{n}{\alpha} \right\rceil\right) + \alpha - 1. \tag{7.11}$$

Corollary 7.1 improves the main result in [170], which consists in a lower bound on the minimum number of edges.

We now determine the leftmost vertices and polytope edges of $P_n^{\alpha,m}$, taking advantage of the following points of $P_n^{\alpha,m}$ delivered by Corollary 7.1:

$$p_k = (k, t(n,k)+k-1), \qquad \text{for } k = 1, 2, \ldots, \left\lfloor \frac{n+1}{2} \right\rfloor. \tag{7.12}$$

A first step is to show that all of these points are boundary points. Next, we determine which of them are vertices. The polytope edges then follow at once (see Corollary 7.2).

Proposition 7.2. *The vertices of the leftmost part of the boundary of* $P_n^{\alpha,m}$ *are exactly the points provided in Equation* (7.12) *for k in* $\{2, 3, \ldots, \lfloor \frac{n-1}{2} \rfloor\}$ *satisfying*

$$\left\lceil \frac{n}{k-1} \right\rceil \neq \left\lfloor \frac{n}{k+1} \right\rfloor + 1, \tag{7.13}$$

and also for $k = 1$ *and* $k = \lfloor \frac{n+1}{2} \rfloor$.

Proof. Leaving aside the vertices $(1, \binom{n}{2})$ and $(\lfloor \frac{n+1}{2} \rfloor, n-1,)$, we assume $1 < k < \lfloor \frac{n+1}{2} \rfloor$ in the rest of the proof. To establish that the point p_k is on the boundary of $\mathcal{P}_n^{\alpha,m}$, it suffices to show that the "Turán function" (for a fixed value of n)

$$t(n,\,): \quad \left\{1,\,2,\,\ldots,\,\left\lfloor \frac{n+1}{2} \right\rfloor\right\} \to \mathbb{R}: \quad x \mapsto t(n,x) \tag{7.14}$$

is the restriction of a convex function from $\left[1,\,\lfloor \frac{n+1}{2} \rfloor\right]$ to $\mathbb{R}$ (notice that we may discard the affine term $k-1$ which appears in the second coordinate of p_k, because it does not alter convexity). In turn, we need only prove for $k = 2,\,3,\,\ldots,\,\lfloor \frac{n-1}{2} \rfloor$

$$t(n,k) \quad \leq \quad \frac{1}{2}\left(t(n,k-1)+t(n,k+1)\right). \tag{7.15}$$

Inequality (7.15) follows from the following three assertions (again for $k = 2,\,3,\,\ldots,\,\lfloor \frac{n-1}{2} \rfloor$): The line with equation

$$x_m - t(n,k) \;=\; -\binom{\left\lfloor \frac{n}{k} \right\rfloor + 1}{2}\,(x_\alpha - k) \tag{7.16}$$

1. goes through the point $p_k = (k, t(n,k))$,
2. supports the point $p_{k-1} = (k-1, t(n,k-1))$, and
3. supports the point $p_{k+1} = (k+1, t(n,k+1))$.

The first assertion is clear. Setting $s = \left\lfloor \frac{n}{k} \right\rfloor$, we see that the second one is equivalent to

$$t(n,k-1) - t(n,k) \;\geq\; \binom{s+1}{2}. \tag{7.17}$$

We derive the latter inequality by checking how many new edges are created when the Turán graph $T(n,k)$ is transformed into the similar graph $T(n,k-1)$. As explained before Theorem 7.1, the graph $T(n,k)$ is a union of cliques with size equal to s and maybe also $s+1$. To transform it into $T(n,k-1)$, we select a clique with size s and move its nodes one after the other to other cliques. At the first step, we loose $s-1$ edges in the clique to be deleted, and gain at least s edges in the augmented clique; in all, we gain at least one edge. For the second node, we have a gain of at least 2 (because we loose $s-2$ edges this time), etc. Thus the total number of edges increases by at least $1+2+\cdots+s$, that is $\binom{s+1}{2}$.

Now the third assertion about the line in Equation (7.16) translates into

$$t(n,k) - t(n,k+1) \;\leq\; \binom{s+1}{2}. \tag{7.18}$$

The graph $T(n,k+1)$ is a union of $k+1$ cliques, of size say d and (maybe also) $d+1$. To transform $T(n,k+1)$ into $T(n,k)$, we move all nodes of some clique of size d to other cliques. As we can create cliques only of size s or $s+1$ (because of the structure of $T(n,k)$), the number of edges increases by at most $s-(d-1)$ at the first move, then at most $s-(d-2)$, etc.; at the last

Table 7.1: The optimal linear inequalities in α and m for connected graphs.

$m \geq n-1$	
$k\alpha + m \leq \binom{n-k}{2} + kn$	for $k = 1, 2, \ldots, n-2$
$m - t(n,k) - (k-1) \geq \big(t(n,k) - t(n,k-1) + 1\big)\,(\alpha - k)$	for $k = 2, 3, \ldots, \lfloor \frac{n+1}{2} \rfloor$ with $\left\lceil \frac{n}{k-1} \right\rceil \neq \left\lfloor \frac{n}{k+1} \right\rfloor + 1$

move, we gain at most s edges. Hence the increase in the number of edges is bounded above by $\frac{1}{2}d(2s+1-d)$. For s fixed, this quantity is maximized for $d = s$ or $d = s+1$. We conclude that Assertion 3 is correct.

Having thus shown that all points p_k provided in (7.12) are on the boundary of $\mathcal{P}_n^{\alpha,m}$, we notice next that all leftmost vertices of $\mathcal{P}_n^{\alpha,m}$ must be among the p_k's (because two successive points p_k have their abscissas differing by 1). Also, a point p_k is not a vertex if and only if the line in Equation (7.16) goes also through the points p_{k-1} and p_{k+1}, which happens exactly if we have equality in both Inequalities (7.17) and (7.18). By inspecting the arguments which led to these inequalities, it is easily seen that both equalities occur if and only if the sizes of the maximal cliques in $T(n,k-1)$, resp. $T(n,k+1)$, are all also sizes of maximal cliques in $T(n,k)$. This is equivalent to saying that $n/(k+1)$ and $n/(k-1)$ lie in an interval with endpoints equal to two consecutive integers. Finally, p_k is not a vertex if and only if $\left\lceil \frac{n}{k-1} \right\rceil = \left\lfloor \frac{n}{k+1} \right\rfloor + 1$. □

Example 7.1. For $n = 24$, the values of k which do not provide a vertex are 7 and then 9, 10, 11. In particular, all points p_9, p_{10} and p_{11} lie on the segment joining the vertices p_8 and p_{12}.

Corollary 7.2. *The leftmost polytope edges of $\mathcal{P}_n^{\alpha,m}$ are defined by the following inequalities:*

$$x_m - t(n,k) - (k-1) \;\geq\; \big(t(n,k) - t(n,k-1) + 1\big)\,(x_\alpha - k) \tag{7.19}$$

for values $k = 2, 3, \ldots, \lfloor \frac{n+1}{2} \rfloor$ satisfying $\left\lceil \frac{n}{k-1} \right\rceil \neq \left\lfloor \frac{n}{k+1} \right\rfloor + 1$.

We have completed our analysis of the case with the two invariants α and m. The full list of optimal linear inequalities appears in Table 7.1. The polygon $\mathcal{P}_n^{\alpha,m}$ has an interesting property which is not shared by all similar polytopes constructed for other choices of invariants: Any point from $\mathcal{P}_n^{\alpha,m}$ which has integer coordinates is produced by some graph in the class considered (here, the class of connected graphs on n nodes).

7.2 Irregularity, diameter and maximum degree of connected graphs

Our methodology is now applied to the following three invariants of a connected graph G : the maximum degree $\Delta(G)$, the irregularity $A(G)$ and the diameter $D(G)$.

A tight upper bound on $A(G)$ in terms of n and m is given in Section 5 [164] when G is general and in Section 4.2 [146] when G is restricted to be a chemical tree. However, there are no similar results, to the best of our knowledge, concerning the irregularity $A(G)$ of Albertson. Consequently, it is worth studying this invariant in relation with other parameters than the number of edges. We choose the maximum degree and the diameter which express simple but different characteristics of graphs.

The polytope of graph invariants is here

$$\mathcal{P}_n^{\mathcal{C},\{\Delta,A,D\}} = \text{conv}\{(\Delta,A,D)|\exists G \in \mathcal{C}_n,$$

$$\Delta(G) = \Delta, A(G) = A, D(G) = D\}, \tag{7.20}$$

where $\mathcal{C}$ is the class of connected graphs.

Remark 7.4. In the following $\mathcal{P}_n^{\mathcal{C},\{\Delta,A,D\}}$ is denoted simply by $\mathcal{P}_n^{\Delta,A,D}$

For any fixed n, the polytope $\mathcal{P}_n^{\Delta,A,D}$ lies in the 3-dimensional space $\mathbb{R}^3$ with coordinates x_Δ, x_A and x_D. As shown in the next proposition, it is full-dimensional. Consequently, the facets of $\mathcal{P}_n^{\Delta,A,D}$ are its 2-dimensional faces.

Proposition 7.3. *Let $\mathcal{P}_n^{\Delta,A,D}$ be the polytope defined by* (7.20)*. If $n \geq 4$, then* $\dim(\mathcal{P}_n^{\Delta,A,D}) = 3$.

Proof. To exhibit four affinely independent points lying in $\mathcal{P}_n^{\Delta,A,D}$, consider the following graphs on n nodes: the complete graph K_n, the cycle C_n, the star $K_{1,n-1}$, and the path P_n. The corresponding points in $\mathbb{R}^3$ are shown below:

$$\begin{aligned} K_n: &\quad (n-1,0,1),\\ C_n: &\quad (2,0,\lfloor n/2 \rfloor),\\ K_{1,n-1}: &\quad (n-1,(n-1)(n-2),2),\\ P_n: &\quad (2,2,n-1). \end{aligned}$$

They are easily seen to be affinely independent. □

The output produced by GraPHedron for $n = 4, 5, \ldots, 10$ led us to conjecture that (under various assumptions) five families of linear inequalities are facet defining. We proceed to give the corresponding proofs.

Proposition 7.4. *If $n \geq 4$, the inequality*

$$x_\Delta \quad \leq \quad n-1 \tag{7.21}$$

is facet defining for $\mathcal{P}_n^{\Delta,A,D}$.

Proof. The validity of Inequality (7.21) for $\mathcal{P}_n^{\Delta,A,D}$ is obvious. Because the origin does not belong to the affine plane $H \equiv x_\Delta = n-1$, it is now sufficient to exhibit three linearly independent vectors of $\mathcal{P}_n^{\Delta,A,D}$ belonging to H. Let us consider the following three graphs: the complete graph K_n, the star $K_{1,n-1}$, and finally G_3, a star augmented with one edge. The corresponding vectors, namely

$$\begin{aligned} K_n: &\quad (n-1,0,1),\\ K_{1,n-1}: &\quad (n-1,(n-1)(n-2),2),\\ G_3: &\quad (n-1,n(n-3),2), \end{aligned}$$

belong to $H \cap \mathcal{P}_n^{\Delta,A,D}$ and are linearly independent. □

Proposition 7.5. *The inequality*

$$x_\Delta + x_D \quad \leq \quad n+1 \tag{7.22}$$

is facet defining for $\mathcal{P}_n^{\Delta,A,D}$ *when* $n \geq 4$.

Proof. (i) We first prove that $\Delta(G)+D(G) \leq n+1$ holds for any connected graph G on n. Take a diameter path P of G, thus P is a shortest path v_1, v_2, …, $v_{D(G)+1}$ (with length $D(G)$). Let v^* be a node having degree $\Delta(G)$ and $\mathcal{N}(v^*)$ be the neighborhood of v^*, i.e. $\mathcal{N}(v^*) = \{v \in V : \{v,v^*\} \in E\} \cup \{v^*\}$. From

$$|P| + |\mathcal{N}(v^*)| - |P \cap \mathcal{N}(v^*)| \quad \leq \quad n,$$

$|P| = D(G)+1$, and $|\mathcal{N}(v^*)| = \Delta(G)+1$, we derive

$$\Delta(G) + D(G) + 2 - |P \cap \mathcal{N}(v^*)| \quad \leq \quad n. \tag{7.23}$$

It remains to prove $|P \cap \mathcal{N}(v^*)| \leq 3$.

1. Suppose $P \cap \mathcal{N}(v^*) = \emptyset$. Then $|P \cap \mathcal{N}(v^*)| = 0 \leq 3$.
2. Suppose $P \cap \mathcal{N}(v^*) \neq \emptyset$ and $v^* \notin P$. Then $|P \cap \mathcal{N}(v^*)| \leq 3$, otherwise P could not be a shortest path.
3. Suppose $P \cap \mathcal{N}(v^*) \neq \emptyset$ and $v^* = v_k \in P$. If $1 < k < D(G)+1$, then $P \cap \mathcal{N}(v^*) = \{v_{k-1}, v^*, v_{k+1}\}$ because P is a shortest path, and so $|P \cap \mathcal{N}(v^*)| = 3$. If $v^* = v_1$ or $v^* = v_{D(G)+1}$, then $|P \cap \mathcal{N}(v^*)| = 2$.

(ii) Having proved that Inequality (7.22) is valid for $\mathcal{P}_n^{\Delta,A,D}$, we now show that it defines a facet. The path P_n, the star $K_{1,n-1}$, and the star augmented with one edge, denoted as G_3 in the proof of Proposition 7.4, produce the vectors

$$(2,2,n-1), \quad \big(n-1,(n-1)(n-2),2\big), \quad \big(n-1,n(n-3),2\big). \tag{7.24}$$

These three vectors satisfy (7.22) with equality and, when $n \geq 4$, they are linearly independent. □

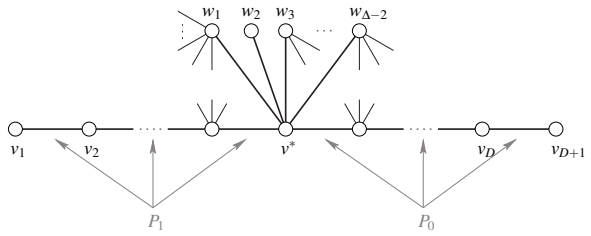
Let $H = (V(H), E(H))$ be a subgraph of $G = (V, E)$. We define the irregularity along H as $A(H) = \sum_{\{k,l\} \in E(H)} |d_k - d_l|$ where the degree d_k of node k is relative to the graph G. To establish in Proposition 7.6 the validity of our next inequality, we first improve Inequality (7.22).

Lemma 7.4. *Let $G = (V, E)$ be a connected graph such that $A(G) < \Delta(G)$. Then $\Delta(G) + D(G) \leq n$.*

Proof. We show for any connected graph G that $\Delta(G) + D(G) > n$ implies $A(G) \geq \Delta(G)$. By Inequality (7.22), we may assume $\Delta(G) + D(G) = n + 1$. From the proof of Proposition 7.5, any node of maximum degree must be on some diameter path. Let $P = v_1, v_2, \ldots, v_{D(G)+1}$ be a diameter path containing at least one node of degree $\Delta(G)$.

By construction, the diameter path P contains $D(G) + 1$ nodes. Consequently, $n - (D(G) + 1) = \Delta(G) - 2$ nodes do not belong to P. We call them exterior nodes (relatively to diameter path P) and denote them by $w_1, w_2, \ldots, w_{\Delta(G)-2}$. The extremities v_1 and $v_{D(G)+1}$ of the diameter path P are of degree at most $\Delta(G) - 1$. So, let $v^* = v_k \in P \backslash \{v_1, v_{D(G)+1}\}$ be the node of P with maximum degree such that the subpath $P_1 \subset P$ which joins v_1 to v^* does not contain any other node of maximum degree. There are $\Delta(G)$ edges incident to the node v^* and two of them belong to the diameter path P. Thus the $\Delta(G) - 2$ remaining edges are incident to the $\Delta(G) - 2$ exterior nodes. We consider two cases (see Figure 7.3).

(a)



(b)

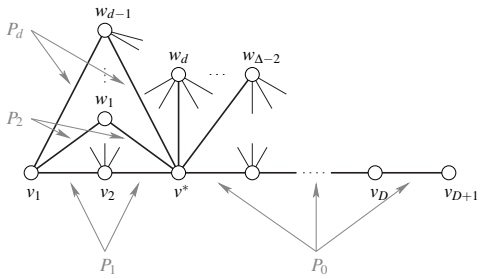


Figure 7.3: All the exterior nodes are adjacent to the node v^*. In Case (a), one has $d_{v_1} = 1$, while in Case (b), one has $1 < d_{v_1} \leq \Delta - 1$.

Case (a). Suppose $d_{v_1} = 1$. By comparing the degrees along P_1, we obtain for the irregularity along P_1

$$A(P_1) \quad \geq \quad \Delta(G) - 1.$$

Case (b). Suppose now $d_{v_1} = d > 1$. (Note that the distance between v_1 and v^* cannot be greater than two, in other terms: $v^* = v_2$ or $v^* = v_3$.) We have

$$A(P_1) \quad \geq \quad \Delta(G) - d.$$

All the nodes adjacent to v_1, but one, are exterior nodes. Let P_2, P_3, ..., P_d be the paths defined by $P_r = \{v_1, w_{r-1}, v^*\}$ (after relabeling the exterior nodes if necessary). Because the paths P_1, P_2, ..., P_d have no common edge, there follows

$$\sum_{r=1}^{d} A(P_r) \quad \geq \quad d\,(\Delta(G) - d).$$

Moreover, $2 \leq d \leq \Delta(G) - 1$, which implies

$$d\,(\Delta(G) - d) \quad \geq \quad \Delta(G) - 1.$$

This proves

$$\sum_{r=1}^{d} A(P_r) \quad \geq \quad \Delta(G) - 1.$$

Now in both Cases (a) and (b) let P_0 be the subpath of P joining v^* to $v_{D(G)+1}$. Clearly, the irregularity along P_0 is at least 1, and so

$$A(G) \quad \geq \quad \Delta(G). \qquad \square$$

Proposition 7.6. *The inequality*

$$n x_\Delta - x_A + (n-1) x_D \quad \leq \quad n^2 - 1 \tag{7.25}$$

is valid for $\mathcal{P}_n^{\Delta,A,D}$. If $n \geq 5$ and n is odd, the inequality is facet defining.

Proof. (i) We first prove the validity of Inequality (7.25). For a graph $G = (V, E)$, this inequality can be rewritten as

$$\Delta(G) + D(G) \quad \leq \quad n + 1 + \frac{A(G) - \Delta(G)}{n-1}. \tag{7.26}$$

If $\Delta(G) \leq A(G)$, Inequality (7.26) is dominated by Inequality (7.22) and is thus valid. If $\Delta(G) > A(G)$, then Lemma 7.4 gives $\Delta(G) + D(G) \leq n$, which implies (7.26) because $\Delta(G) \leq n - 1$.

(ii) We now prove that, for $n \geq 5$ and n odd, Inequality (7.25) is facet defining. The following three graphs produce vectors which are linearly independent and satisfy Inequality (7.25) with equality: K_n, P_n, and the graph on n nodes such that one of its nodes has degree $n-1$ and the $n-1$ other ones have degree $n-2$. Notice that this last graph only exists for n odd. $\square$

On the contrary, as we will see at the end of this section, Inequality (7.25) is not facet defining when n is even. Before considering two more inequalities, we establish two lemmas.

Lemma 7.5. *Let G be a connected graph with $n \geq 4$, $D(G) \geq 3$ and $\Delta(G) + D(G) = n + 1$. Then $A(G) \geq 2\Delta(G) - 2$.*

Proof. We build upon the proof of Lemma 7.4. Let $\mathcal{P} = \{v_1, v_2, \ldots, v_{D(G)+1}\}$ be a diameter path containing some node v^* of degree $\Delta(G)$. As in the proof of Lemma 7.4, we get $v^* \in \{v_2, v_3, \ldots, v_{D(G)}\}$. Let V' be the set of exterior nodes. Note that an exterior node cannot be adjacent to both v_1 and $v_{D(G)+1}$, because of our assumption $D(G) \geq 3$. Thus, V' is the disjoint union of the subset V_1 of exterior nodes adjacent to v_1, the subset V_2 of exterior nodes adjacent to $v_{D(G)+1}$, and the subset V'' of all remaining exterior nodes. Consider three cases.

(a) Suppose $V_1 \neq \emptyset \neq V_2$. (This is possible only if $D(G) = 3$ or 4 because $\Delta(G) + D(G) = n + 1$ implies that all exterior nodes are adjacent to v^*.) Consider the set $\mathbb{P}_1$ of paths v_1, t, v^* with $t \in V_1$, and the set $\mathbb{P}_2$ of paths $v_{D(G)+1}, u, v^*$ with $u \in V_2$. The same argument as in the proof of Lemma 7.4 shows that the total irregularity computed along the paths from $\mathbb{P}_1$, or from $\mathbb{P}_2$, must be at least $\Delta(G) - 1$. Because $V_1 \cap V_2 = \emptyset$, it is clear that any two paths in $\mathbb{P}_1 \cup \mathbb{P}_2$ have no common edge. Thus

$$A(G) \geq 2\Delta(G) - 2.$$

(b) Suppose $V_1 \neq \emptyset$ and $V_2 = \emptyset$ (the case where $V_1 = \emptyset$ and $V_2 \neq \emptyset$ can be treated along the same line). This implies $d_{v_{D(G)+1}} = 1$. Similarly to Case (a), the total irregularity along paths v_1, t, v^* with $t \in V_1$ is at least $\Delta(G) - 1$. Furthermore, the path $Q = v_{D(G)+1}, v_{D(G)}, \ldots, v^*$ has no common edge with those paths and it satisfies $A(Q) \geq \Delta(G) - 1$. Hence

$$A(G) \geq 2\Delta(G) - 2.$$

(c) Suppose $V_1 = \emptyset = V_2$. Then

$$A(G) \quad \geq \quad A(\{v_1, v_2, \ldots, v^*\}) + A(\{v^*, \ldots, v_{D(G)}, v_{D(G)+1}\}) \quad \geq \quad 2\Delta(G) - 2. \qquad \square$$

Lemma 7.6. *Let G be a connected graph with n even, $n \geq 4$, $D(G) = 2$, and $\Delta(G) = n - 1$. Then*

$$A(G) \quad \geq \quad 2n - 4. \tag{7.27}$$

Proof. Let $k \geq 1$ be the number of nodes having degree $n - 1$. The edges connecting the nodes of degree $n - 1$ to nodes of smaller degree lead to

$$A(G) \quad \geq \quad k(n - k). \tag{7.28}$$

If $2 \leq k \leq n - 2$, Inequality (7.27) follows from Inequality (7.28). Because the graph cannot be complete, the case $k = n - 1$ is impossible. It remains to consider the case $k = 1$.

Suppose $k = 1$, let v be the unique node of degree $n - 1$ and let ℓ denote the number of nodes of degree $n - 2$. The contribution to $A(G)$ of the edges connecting node v to the ℓ nodes of degree

$n-2$ is ℓ, and the contribution to $A(G)$ of the edges connecting node v to the $n-\ell-1$ nodes of degree lower than $n-2$ is at least $2(n-\ell-1)$. Finally, the edges connecting the ℓ nodes of degree $n-2$ to the $n-\ell-1$ nodes of degree lower than $n-2$ contribute to $A(G)$ for at least $\ell(n-\ell-2)$. Summing up, we obtain

$$A(G) \quad \geq \quad (\ell+2)(n-\ell-1).$$

If $0 \leq \ell \leq n-3$, Inequality (7.27) follows from the above inequality. Notice that $\ell = n-1$ is impossible because n is even. Thus it remains to prove the result for $k=1$ and $\ell = n-2$.

Suppose $k=1$ and $\ell = n-2$. Then, there exists a unique node of degree $d < n-2$ and

$$A(G) \quad \geq \quad (n-2)+(n-1-d)+(d-1)(n-2-d) \tag{7.29}$$

$$\quad = \quad n-2+d(n-2-d)+1. \tag{7.30}$$

Because $d(n-2-d) \geq n-3$ for $1 \leq d \leq n-3$, the expression above leads to Inequality (7.27). □

Proposition 7.7. *The inequalities*

$$2x_\Delta - x_A + 2x_D \quad \leq \quad 2n \tag{7.31}$$

and

$$2(n-1)x_\Delta - x_A + 2(n-2)x_D \quad \leq \quad 2(n^2-n-1) \tag{7.32}$$

are facet defining for $P^n_{\Delta,A,D}$ when n is even and $n \geq 4$.

Proof. (i) The validity of Inequality (7.31) can be easily proved: we show $2\Delta(G) - A + 2D \leq 2n$ for any connected graph G. If G is regular, then (7.31) reduces to $\Delta(G) + D(G) \leq n$ which is always valid by Lemma 7.4. If G is not regular, then $A(G) \geq 2$ and Inequality (7.31) is dominated by $2\Delta(G) + 2D(G) \leq 2n+2$ which is true by Proposition 7.5.

The validity of Inequality (7.32) can be shown along the same line as in the proof of Proposition 7.5. Indeed, for a connected graph G, Inequality (7.32) is equivalent to

$$\Delta(G) + D(G) \quad \leq \quad n+1+\frac{2+(A(G)-2\Delta(G))}{2(n-2)}. \tag{7.33}$$

Consider two cases.

(a) If $\Delta(G) + D(G) \leq n$, then (7.33) holds because $\Delta(G) \leq n-1$ and $A \geq 0$ imply that the right hand side is bounded from below by n.

(b) Suppose $\Delta(G) + D(G) = n+1$. Then the fact that the right hand side is bounded from below by $n+1$ follows from Lemmas 7.5 and 7.6.

(ii) We now prove that the two inequalities in the statement are facet defining. For (7.32), three linearly independent vectors giving equality are provided by the graphs K_n, P_n and the "boat" on n

nodes, i.e. the graph consisting of the path P_{n-1} with an exterior node linked to its three consecutive middle nodes (see Figure 7.4). This last graph is well defined for $n \geq 4$ and n even.

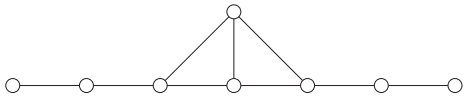


Figure 7.4: The boat on 8 nodes.

For (7.31), replace the boat with the graph sketched in Figure 7.5.

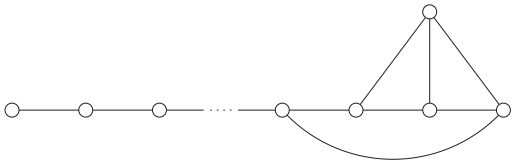


Figure 7.5: A graph used in the proof of Proposition 7.7

This graph is defined for $n \geq 6$ and it has five nodes of degree 3, one node of degree 1, and all other nodes of degree 2. In case $n = 4$, use rather the complete graph minus one edge. □

Notice that the sum of Inequalities (7.31) and (7.32) is equal to twice (7.25). Consequently, the latter inequality does not define a facet for n even.

7.3 Tight analysis of the maximal matching heuristic

As already stated in Section 6.3.2, GraPHedron can be used to compare values of heuristics and exact algorithms.

Remark 7.5. In this section, we refer to the books [74] and [132] for classical notions about algorithms and their complexity.

In this section, such a study is made to compare approximation ratios of the worst-case approximation ratio of the *maximal matching heuristic* for two classical NP-hard problems [132], under some conditions on the density of the graphs. These problems are the *minimum node cover problem* and the *minimum maximal matching problem*.

The maximal matching heuristic is an algorithm that provides a 2-approximation for the minimum node cover and minimum maximal matching problems. It is perhaps one of the simplest and best-known approximation algorithms.

We ask the question of expressing the approximation ratio in a finer way, as a function of well-chosen graph parameters. We show that density parameters are good candidates for this pur-

pose. Actually, the approximation ratio of the maximal matching heuristic is strictly less than 2 for graphs with a sufficiently high number of edges or sufficiently high minimum degree. We characterize precisely the asymptotic approximation ratio as a function of these parameters, together with tight examples. This is, to our knowledge, the tightest analysis ever done of this algorithm. This study shows that even simple heuristics might deserve nontrivial analyses. It was initiated using GraPHedron.

First, in Section 7.3.1, we recall some definitions about matchings in graphs and we recall the statement of the minimum node cover and minimum maximal matching problems. We review the best known approximation ratios of the maximal matching heuristic in Section 7.3.2 for these two problems. We introduce several lemmata in Section 7.3.3, useful for the next sections. In Section 7.3.4 we study the worst-case approximation ratio for the minimum node cover problem, considering the density of the graphs. The same kind of analysis is performed for minimum maximal matching in Section 7.3.5. Finally, we draw some remarks about our results in Section 7.3.6

7.3.1 Definitions

Matchings

Definition 7.1. *A* matching *in a graph G is a set of edges $E' \subseteq E$ that have no endpoints in common. A matching E' is* maximal *(under the set inclusion) if every edge in $E \setminus E'$ shares a common endpoint with some edge in E'.*

Definition 7.2. *Let $G = (V, E)$ be a graph with a matching $E' \subseteq E$, and a path P in G. Then P is said to be* alternating *(with respect to E') if its edges are alternately in E' and in $E \setminus E'$. An alternating path is termed* augmenting *if its end nodes are not incident with any edge of E'.*

Definition 7.3. *A* maximum matching *is a maximal matching with maximum cardinality. The cardinality of a maximum matching in G is called the* matching number *and is denoted by $\varphi(G)$.*

Definition 7.4. *A* perfect matching *is a matching which covers all nodes of the graph. That is, every node of the graph is incident to exactly one edge of the matching. A graph with an odd number of nodes is allowed one unmatched node.*

Example 7.2. Consider the graph G shown in Figure 7.6. The sets of singleton edges $\{v_1v_2\}$, $\{v_2v_3\}$, $\{v_3v_4\}$ and $\{v_1v_4\}$ are the nonempty matchings in G that are not maximal. The set $\{v_1v_3\}$ is a maximal matching in G that is not maximum, and the sets $\{v_1v_2, v_3v_4\}$, $\{v_1v_4, v_2v_3\}$ are the maximum matchings in G. These two maximum matchings are also perfect matchings. The path (v_1, v_2, v_3, v_4) is an augmenting path with respect to the matching $\{v_2v_3\}$.

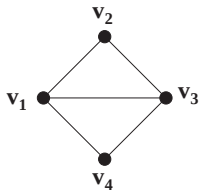


Figure 7.6: A graph G

Maximal matching heuristic

Heuristic 7.1. Let G be a graph. The *maximal matching heuristic* consists in finding a maximal matching iteratively as follows. Pick any arbitrary edge e in the graph and add it to the matching. Delete any edge sharing a node with e, and repeat until all the remaining edges are in the matching. Actually the matching itself is not needed in our case : we need only to know the cardinality of a maximal matching.

MAXIMALMATCHING(G):
Input: a graph $G(V,E)$.
Output: The cardinality of a maximal matching in G

1. $c \leftarrow 0$
2. $F \leftarrow E$
3. While $F \neq \emptyset$, do
   (a) Choose an edge $e \in F$.
   (b) $F \leftarrow F \setminus e$
   (c) $c \leftarrow c+1$
   (d) Remove all edges from F which share a common endpoint with e.
4. return c

Problems

Problem 7.1. Let G be a graph. In the *minimum node cover* problem, one is asked to find a minimum cardinality set of nodes that contains at least one endpoint of each edge of G. We denote by $\tau(G)$ the size of a minimum cardinality node cover of G.

Example 7.3. The optimal solution of the minimum node cover problem applied on the graph in Figure 7.6 is 2.

Problem 7.2. Let G be a graph. In the *minimum maximal matching* problem, one is asked to find a maximal matching of minimum cardinality in G. We denote by $\nu(G)$ the size of a minimum maximal matching of G.

Remark 7.6. The maximal matching heuristic returns an integer c such that $\nu(G) \leq c \leq \varphi(G)$.

Example 7.4. The optimal solution of the minimum maximal matching problem applied on the graph in Figure 7.6 is 1.

Tightness of the 2-approximation of the maximal matching heuristic for these problems, is witnessed by a number of examples.

Example 7.5. Suppose one uses the maximal matching heuristic on the graph G in Figure 7.6, to approximate the optimal solution of the minimum node cover problem. The idea of the approximation is that all the endpoints of the maximal matching form a node cover. The worst-case of this approximation is thus $2\varphi(G) = 4$. This solution is worse twice than the optimal solution $\tau(G) = 2$.

Example 7.6. Suppose now that one uses the maximal matching heuristic on the graph in Figure 7.6, to approximate the optimal solution of the minimum maximal matching problem. In this case, the worst-case of the maximal matching heuristic is $\varphi(G) = 2$. This solution is again worse twice than the optimal solution $\nu(G) = 1$.

7.3.2 Approximation ratios of the maximal matching heuristic

Ratios for the minimum node cover

One can easily see that the set of endpoints of a maximal matching indeed contains at least one endpoint of each edge, and that the optimal solution – since it must contain a least one node of each edge of the matching – has size not less than half the size of this set, hence the 2-approximation.

The minimum node cover problem is thus 2-approximable using the maximal matching heuristic, but no polynomial time algorithm with constant approximation ratio better than 2 is known. The problem is further known to be APX-complete [203] and not approximable within a factor of $7/6$ [173]. Monien and Speckenmeyer [199] and Bar-Yehuda and Even [22] provide algorithms that achieves a ratio of $(2-(\ln\ln n)/\ln n)$, where n is the number of nodes in the graph. Karakostas [178] later reduced the approximation ratio to $2-\Theta(1/\sqrt{\log n})$. For graphs with maximum degree Δ, Halperin [155] provides an approximation algorithm with a ratio of $2-(1-O(1))2\ln\ln\Delta/\ln\Delta$.

The problem has further been studied under the hypothesis that the input graph is *dense*.

Definition 7.5. *A graph G is* weakly ε-dense *if*

$$m(G) \geq \varepsilon \frac{n(G)^2}{2}.$$

Definition 7.6. *A graph G is* strongly ε-dense *if*

$$\delta(G) \geq \varepsilon n(G).$$

It has been shown [65] that the minimum node cover problem restricted to strongly ε-dense graphs is APX-complete. Eremeev [105] shows that it is NP-hard to approximate the minimum node cover within a ratio less than $(7+\varepsilon)/(6+2\varepsilon)$ in strongly ε-dense graphs.

Nagamochi and Ibaraki [176] provide an approximation algorithm with a ratio of $2-8m/(13n^2+8m)$, where m is the number of edges in the graph. This algorithm can also be seen as an approximation algorithm that achieves an approximation ratio that is asymptotic to $2-4\varepsilon/(13+4\varepsilon)$ for weakly ε-dense graphs.

Karpinski and Zelikovsky [180] propose an algorithm that achieves a better ratio of $2/(2-\sqrt{1-\varepsilon})$ for weakly ε-dense graphs, and a ratio of $2/(1+\varepsilon)$ for strongly ε-dense graphs.

Finally, Imamura and Iwama[177] recently proposed a randomized approximation algorithm. This algorithm yields, with high probability, an approximation factor of $2/(1+\gamma(G))$, where $\gamma(G)$ is a function of the maximum and the average degree, and runs in polynomial time if Δ, the maximum degree of the graph, is $\Omega(n\log\log n/\log n)$.

Ratios for the minimum maximal matching

It is fairly easy to see that any maximal matching has a size that is at most twice the size of the minimum maximal matching. Much less is known about the minimum maximal matching problem than about the minimum node cover problem. Chlebìk and Chlebìková [61] do nevertheless show that it is NP-hard to approximate the problem within a constant factor better than $7/6$.

7.3.3 Preliminaries

We shall make extensive use of the family of complete split graphs. These graphs arise as extremal graphs for several graph invariants (see for instance Chapter 3 and Sections 5.2 and 7.1.2).

Recall that a complete split graph $CS^{\alpha}_{n-\alpha}$ with $1 \leq \alpha \leq n-1$, is a graph that can be decomposed in an independent set of size α and a clique of size $n-\alpha$, with each node of the independent set being adjacent to each node in the clique.

Our proofs make use of the following basic results on the values of invariants of complete split graphs.

Lemma 7.7. $m(CS^{\alpha}_{n-\alpha}) = \binom{n}{2} - \binom{\alpha}{2} = (n-\alpha)(n+\alpha-1)/2$.

Proof. The proof is immediate: our complete split graphs contain all $\binom{n}{2}$ possible edges minus the $\binom{\alpha}{2}$ edges missing from the independent set. □

Lemma 7.8. $\varphi(CS^{\alpha}_{n-\alpha}) = \begin{cases} \lfloor \frac{n}{2} \rfloor & \text{if } \alpha \leq n/2, \\ n-\alpha & \text{otherwise.} \end{cases}$

Proof. We consider the two cases $\alpha \leq n/2$ and $\alpha > n/2$.

i) $\alpha \leq n/2$. The size of the independent set is lower or equal to the size of the clique. We show that any non-perfect matching can be augmented. A non-perfect matching has at least two unmatched nodes. If there are two unmatched nodes in the clique, or one in the clique and one in the independent set, the matching can trivially be augmented. Otherwise there are two unmatched nodes a and b in the independent set, and an edge xy of the clique that belongs to the matching. The matching is augmented by taking edges ax and by instead of xy.

ii) $\alpha > n/2$. The independent set is strictly greater than the clique. A matching of size $n-\alpha$ is trivially found by matching $n-\alpha$ arbitrarily chosen nodes of the independent set with the nodes of the clique. This matching is of maximum cardinality since any other matching would have at least one edge xy in the clique and would thus have less than $n-\alpha$ edges. □

Lemma 7.9. $\tau(CS^{\alpha}_{n-\alpha}) = n-\alpha$.

Proof. The nodes of the clique clearly form a node cover of size $n-\alpha$ for G. Any smaller set of nodes would contain x nodes in the independent set and y nodes in the clique, with $x+y < n-\alpha$. If $x = 0$ then at least α edges joining a node from the clique to the nodes of the independent set would not be covered. If $x > 0$, then at least some edge in the clique will not be covered since we would have $y < n-\alpha-1$. □

Lemma 7.10. $\nu(CS^{\alpha}_{n-\alpha}) = \lceil \frac{n-\alpha}{2} \rceil$.

Proof. We first show that any matching of size strictly lower than $\lceil \frac{n-\alpha}{2} \rceil$ cannot possibly be maximal: such a matching would contain at most $\lceil \frac{n-\alpha}{2} \rceil - 1$ edges taken in the clique. in which case the matching can be augmented by taking an edge joining an unmatched node of the clique and any node of the independent set. There remains to be shown that a maximal matching of size $\lceil \frac{n-\alpha}{2} \rceil$ always exists. In case $n-\alpha$ is even, taking any $\frac{n-\alpha}{2}$ disjoint edges in the clique will trivially yield a maximal matching. In case $n-\alpha$ is odd, taking $\lfloor \frac{n-\alpha}{2} \rfloor$ disjoint edges in the clique and an additional edge joining the unmatched nodes of the clique and an arbitrary node of the independent set clearly yields a maximal matching. □

The following three simple properties are also useful. Recall that the minimum degree of a graph G is denoted by $\delta(G)$.

Lemma 7.11. *For any graph G, we have* $\tau(G) \geq \delta(G)$.

Proof. By contradiction. Let us suppose we have $\tau(G) < \delta(G)$ for some graph G. Since the nodes outside of the minimum node cover form an independent set and $\tau(G) < \delta(G)$, these nodes have degree lower than $\delta(G)$, which is impossible. □

Lemma 7.12. *For any graph G, we have* $\nu(G) \geq \left\lceil \frac{\delta(G)}{2} \right\rceil$.

Proof. Since the nodes of a maximal matching form a node cover, we have, by lemma 7.11 : $2\nu(G) \geq \delta(G)$. □

Lemma 7.13. *Graphs that are isomorphic to a complete bipartite graph* $K_{x,x}$ *having at most* $x-1$ *edges removed have a perfect matching.*

Proof. Note that the graph resulting from the removal of at most $x-1$ edges from $K_{x,x}$ is still a bipartite graph. Hall's *marriage theorem* (see [97]) states that a bipartite graph with partition classes A and B has a matching of A if and only if $|N(S)| \geq |S|$ for all $S \subseteq A$.

Let A and B be the two partition classes of our graph. A matching of A is trivially a perfect matching of the graph and can easily be shown to exist using the theorem: for each $S \subseteq A$, we have $|N(S)| \geq |S|x-(x-1)$, which is trivially greater than $|S|$ when $x \geq 1$. □

7.3.4 Minimum node cover

We analyze the worst-case behavior of the maximal matching heuristic when applied to the minimum node cover problem. Recall that the idea of the approximation is that the endpoints of the maximal matching is a node cover. One has thus to count twice the matching number φ. In this case the approximation ratio is a graph invariant defined by

$$\frac{2\varphi(G)}{\tau(G)}.$$

We first consider weakly ε-dense graphs, which amounts to express the approximation ratio as a function of the number of edges. Next, we consider strongly ε-dense graphs by expressing the approximation ratio as a function of the minimum degree.

Approximation ratio vs number of edges

Remark 7.7. The following results were derived with the help of GraPHedron. In this problem, the set of invariants I is $\{\frac{2\varphi}{\tau}, m\}$ (see Section 6.1.2). The class of graphs under study are the

general graphs. In the following, we note $\mathcal{G}_{n,m}$ the class of all non-isomorphic graphs of order n and size m.

Lemma 7.14. *Let n and m be positive integers such that*

$$m(CS^{\alpha+1}_{n-\alpha-1}) < m \leq m(CS^{\alpha}_{n-\alpha})$$

for some α*. The minimum value of* $\tau(G)$ *attained by a graph G in* $\mathcal{G}_{n,m}$ *is* $\tau(CS^{\alpha}_{n-\alpha}) = n-\alpha$*.*

Proof. The proof is in two steps: we first show that a graph G in $\mathcal{G}_{n,m}$ cannot have $\tau(G) < \tau(CS^{\alpha}_{n-\alpha})$, and second that there is a graph G in $\mathcal{G}_{n,m}$ having $\tau(G) = \tau(CS^{\alpha}_{n-\alpha})$. The first step is by contradiction. We show that if $\tau(G) < \tau(CS^{\alpha}_{n-\alpha})$, then $m \leq m(CS^{\alpha+1}_{n-\alpha-1})$, contradicting our hypothesis on m:

$$m(G) \leq \sum_{i=1}^{\tau(G)} (n-i) \leq \sum_{i=1}^{n-\alpha-1} (n-i) = \binom{n}{2} - \binom{\alpha+1}{2} = m(CS^{\alpha+1}_{n-\alpha-1}).$$

The first inequality above is obtained by maximizing the number of edges covered by each node of the node cover: the first node in the cover can cover at most $n-1$ edges, the second at most $n-2$ edges, and so on. The second step is by construction. Let G be a graph formed by removing $m(CS^{\alpha}_{n-\alpha}) - m$ edges from the edges joining the clique and the independent set of a $CS^{\alpha}_{n-\alpha}$. We show that $\tau(G) = \tau(CS^{\alpha}_{n-\alpha})$. The number of edges removed is between 0 and $\alpha-1$ since $m(CS^{\alpha}_{n-\alpha}) - m(CS^{\alpha+1}_{n-\alpha-1}) = \alpha$. Thus each node in the clique has at least one neighbor in the independent set. We now show $\tau(G) = \tau(CS^{\alpha}_{n-\alpha}) = n-\alpha$. The nodes of the clique clearly form a node cover of size $n-\alpha$ for G. Any smaller set of nodes would contain x nodes in the independent set and y nodes in the clique, with $x+y < n-\alpha$. If $x = 0$ then at least one edge joining a node from the clique to the nodes of the independent set would not be covered. If $x > 0$, then at least some edge in the clique will not be covered since we would have $y < n-\alpha-1$. We thus have $\tau(G) = n-\alpha$. □

Lemma 7.15. *Let n and m be positive integers such that*

$$m(CS^{\alpha+1}_{n-\alpha-1}) < m \leq m(CS^{\alpha}_{n-\alpha})$$

for some α*. There exists a graph G in* $\mathcal{G}_{n,m}$ *such that* $\tau(G) = \tau(CS^{\alpha}_{n-\alpha})$ *and* $\varphi(G) = \varphi(CS^{\alpha}_{n-\alpha})$*.*

Proof. Let G be a graph formed by removing $m(CS^{\alpha}_{n-\alpha}) - m$ edges from the edges joining the clique and the independent set of a $CS^{\alpha}_{n-\alpha}$. The number of edges removed is between 0 and $\alpha-1$ since $m(CS^{\alpha}_{n-\alpha}) - m(CS^{\alpha+1}_{n-\alpha-1}) = \alpha$. We show that $\varphi(G) = \varphi(CS^{\alpha}_{n-\alpha})$.

When $\alpha \leq n/2$, a perfect matching for G can be obtained in the following way: any set of α nodes of the clique can be perfectly matched with the nodes of the independent set, according to lemma 7.13, and the remaining nodes of the clique can be matched inside the clique. When $\alpha > n/2$, each node of the clique can be matched with a node of the independent set, according to lemma 7.13, yielding a matching of size $n-\alpha = \varphi(CS^{\alpha}_{n-\alpha})$.

Furthermore, we have $\tau(G) = \tau(CS^{\alpha}_{n-\alpha})$ by the proof of lemma 7.14. □

The above results enable us to state Theorem 7.3:

Theorem 7.3. *Let $\beta(G)$ be the worst-case approximation ratio for graph G. Let $\beta(m,n)$ be the worst approximation ratio attained by a graph in $\mathcal{G}_{n,m}$. We have:*

$$\beta(m,n) \quad = \beta\left(CS^{\alpha^*(m,n)}_{n-\alpha^*(m,n)}\right) = \quad \begin{cases} 2 & \text{if } \alpha^*(m,n) > n/2, \\ \frac{2\lfloor \frac{n}{2} \rfloor}{n-\alpha^*(m,n)} & \text{otherwise}, \end{cases}$$

where

$$\alpha^*(m,n) = \left\lfloor 1/2 + \sqrt{n(n-1) + 1/4 - 2m} \right\rfloor$$

is the integer value α such that $m\left(CS^{\alpha+1}_{n-\alpha-1}\right) < m \leq m(CS^{\alpha}_{n-\alpha})$.

Proof. Lemma 7.15 implies the existence of a graph in $\mathcal{G}_{n,m}$ on which the ratio is given by the expression of the theorem. When $\alpha^*(m,n) > n/2$, a value of 2 trivially maximizes the ratio of our 2-approximation algorithm. When $\alpha^*(m,n) \leq n/2$, the ratio is maximized since the numerator is maximized (the matching is perfect) and the denominator is minimized (by Lemma 7.14). □

The above theorem gives a tight upper bound on the approximation ratio of the maximal matching heuristic to the minimum node cover problem in graphs of n nodes and m edges, in the form of a discrete step function of m. The function equals 2 when $\alpha > n/2$ and begins to decrease afterwards (see Figure 7.7).

Corollary 7.3. *Let $\tilde{\beta}(\varepsilon,n)$ be the worst approximation ratio attained by a graph with n nodes and an average degree at least εn. We have:*

$$\lim_{n\to\infty} \tilde{\beta}(\varepsilon,n) = \begin{cases} 2 & \text{if } \varepsilon \leq 3/4, \\ \frac{1}{1-\sqrt{1-\varepsilon}} & \text{otherwise}. \end{cases}$$

Proof. Since $\beta(m,n)$ is a monotonically decreasing function of m for each fixed n, $\tilde{\beta}(\varepsilon,n)$ can be defined as $\beta(\lceil \varepsilon n^2/2 \rceil, n)$. This yields:

$$\tilde{\beta}(\varepsilon,n) = \begin{cases} 2 & \text{if } \alpha^*(\lceil \varepsilon n^2/2 \rceil, n) > n/2, \\ \frac{2\lfloor \frac{n}{2} \rfloor}{n-\alpha^*(\lceil \varepsilon n^2/2 \rceil, n)} & \text{otherwise.} \end{cases} \tag{7.34}$$

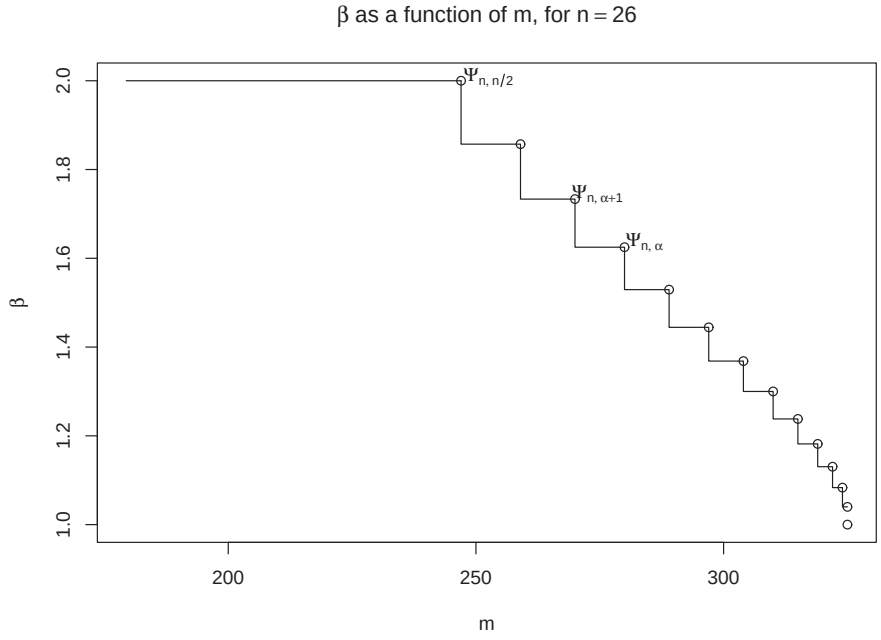


Figure 7.7: The step function

We now compute the asymptotic of the second above expression:

$$\begin{aligned}
\lim_{n\to\infty} \frac{2\lfloor n/2\rfloor}{n-\alpha^*(\lceil \varepsilon n^2/2\rceil)} &= \lim_{n\to\infty} \frac{2\lfloor n/2\rfloor}{n-\left\lfloor 1/2+\sqrt{n(n-1)+1/4-2\lceil \varepsilon n^2\rceil}\right\rfloor} \\
&= \lim_{n\to\infty} \frac{n+O(1)}{n-\sqrt{n(n-1)-\varepsilon n^2+O(1)}+O(1)} \qquad (7.35)\\
&= \lim_{n\to\infty} \frac{1+O(1/n)}{1-\sqrt{1-\varepsilon+O(1/n)}+O(1/n)} \\
&= \frac{1}{1-\sqrt{1-\varepsilon}} \qquad (7.36)
\end{aligned}$$

Let us now take a closer look at the condition $\alpha^*(\lceil \varepsilon n^2/2\rceil, n) \leq n/2$:

$$\begin{aligned}
\alpha^*(\lceil \varepsilon n^2/2\rceil, n) &\leq n/2 \\
\Leftrightarrow \lfloor 1/2+\sqrt{n(n-1)+1/4-2\lceil \varepsilon n^2/2\rceil}\rfloor &\leq n/2 \\
\Leftrightarrow \sqrt{n(n-1)-\varepsilon n^2+O(1)}+O(1) &\leq n/2 \\
\Leftrightarrow \frac{-(n/2+O(1))^2+n(n-1)+O(1)}{n^2} &\leq \varepsilon \\
\Leftrightarrow 3/4+O(1/n) &\leq \varepsilon. \qquad (7.37)
\end{aligned}$$

Results (7.34), (7.36), and (7.37) immediately imply our corollary. □

This asymptotic result is to be compared with the results of [176] and [180] quoted in Section 7.3.2 (see Figure 7.8).

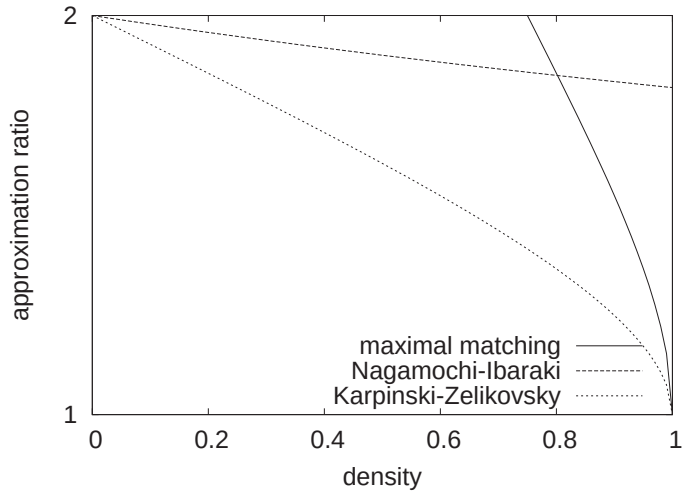


Figure 7.8: A comparison of the approximation ratios

Approximation ratio vs minimum degree

Remark 7.8. The following results were derived with the help of GraPHedron. In this problem, the set of invariants I is $\{\frac{2\varphi}{\tau}, \delta\}$ (see Section 6.1.2).

Let $A_{n,\alpha}$ be the set of all graphs of minimum degree $n-\alpha$ that can be expressed as the complete product of an independent set of order α and a graph of order $n-\alpha$. Note that $A_{n,\alpha}$ contains $CS^{\alpha}_{n-\alpha}$.

Lemma 7.16. *For all $G \in A_{n,\alpha}$ we have* $\varphi(G) = \begin{cases} n-\alpha & \text{if } \alpha \geq n/2, \\ \lfloor \frac{n}{2} \rfloor & \text{otherwise.} \end{cases}$

Proof. If $\alpha \geq n/2$, then G has an independent set of order α joined to a graph of lower or equal order $n-\alpha$. At most $n-\alpha$ nodes of the independent set can be matched, using all $n-\alpha$ nodes of the other graph.

There remains to show that if $\alpha < n/2$, then G has a perfect matching. We show that any non-perfect matching M for G can be augmented.

Let us look at a matching and fix the following notations (see Figure 7.9): let I be the independent set of size α in G, let U be the set of unmatched nodes in I, let W be the set of matched nodes in I, let X be the set of unmatched nodes in $V \backslash I$, let Y be the set of nodes in $V \backslash I$ that are matched inside $V \backslash I$, and finally let Z be the set of nodes in $V \backslash I$ that are matched with nodes of I. Our matching not being perfect, there are at least two unmatched nodes. Three cases must be distinguished.

i) *Both U and X are non-empty.* In that case the matching is augmented by adding to it an edge joining a node from U to a node from X.

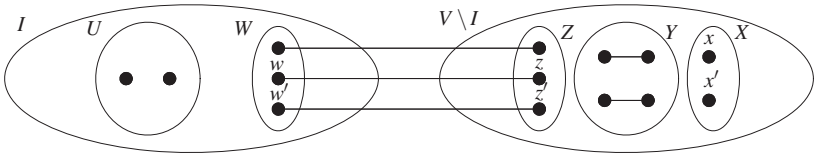


Figure 7.9: Notations for the perfect matching construction

ii) *U is empty and X is non-empty.* In that case we first prove that any node $z \in Z$ always has a neighbor in Z or in X (note that Z is non-empty in this second case). If z has no neighbors in Z nor in X, it can have at most $(n-\alpha-1)-(\alpha-1)-|X|$ neighbors in $V\backslash I$, where $n-\alpha-1$ is the number of other nodes in $V\backslash I$ and $\alpha-1$ is the number of other nodes in Z. This quantity is upper bounded by $n-2\alpha-2$ since $|X| \geq 2$. But our minimum degree hypothesis implies that z needs to have at least $n-2\alpha$ neighbors in $V\backslash I$, a contradiction. We now consider the two possible cases. If z has a neighbor z' in Z, let wz and $w'z'$ be the edges in M having z and z' as endpoints, and let $\{x,x'\} \subseteq X$. We can augment the matching by replacing edges wz and $w'z'$ in M by edges zz', xw, and $x'w'$. Otherwise, z has a neighbor x in X. Let now x' be another node in X, and wz be the edge in M having z as an endpoint. We augment the matching by replacing edge wz with edges xz and $x'w$.

iii) *X is empty and U is non-empty.* Let T be a set of $\lfloor \frac{|U|}{2} \rfloor$ edges taken in Y. Augment the matching by removing these edges and joining each one of their endpoints with a node in U (leaving one unmatched node in U if $|U|$ is odd). Note that the matching thus obtained is perfect. □

Lemma 7.17. *For all $G \in A_{n,\alpha}$ we have $\tau(G) = n-\alpha$. Furthermore, among all graphs with n nodes and minimum degree $n-\alpha$, this value of τ is minimal, and is attained only by graphs in $A_{n,\alpha}$.*

Proof. Let I be an independent set of size α in G. We know from Lemma 7.11 that a graph cannot have a node cover of size lower than its minimum degree. Since $V(G)\backslash I$ is a node cover of size $\delta(G)$, we have $\tau(G) = n-\alpha$. There remains to show that this value is attained by graphs in $A_{n,\alpha}$ only. This is done by construction. Let G' be a graph having $\tau(G') = \delta(G')$. We show that such a graph must be in $A_{n,\alpha}$ with $\alpha = n-\delta(G')$. Let X be the nodes of a minimum node cover of G'. Since X is a node cover, $V\backslash X$ is an independent set. But since $|X| = \delta(G')$, each node of the independent set must be adjacent to each node in X. This is the definition of a graph in $A_{n,\alpha}$ with $\alpha = n-\delta(G')$. □

Theorem 7.4. *Let $\gamma(\delta,n)$ be the worst approximation ratio attained by a graph with n nodes and minimum degree δ. We have:*

$$\gamma(\delta,n) = \begin{cases} 2 & \text{if } \delta \leq \frac{n}{2}, \\ \frac{2}{\delta}\lfloor \frac{n}{2} \rfloor & \text{otherwise.} \end{cases}$$

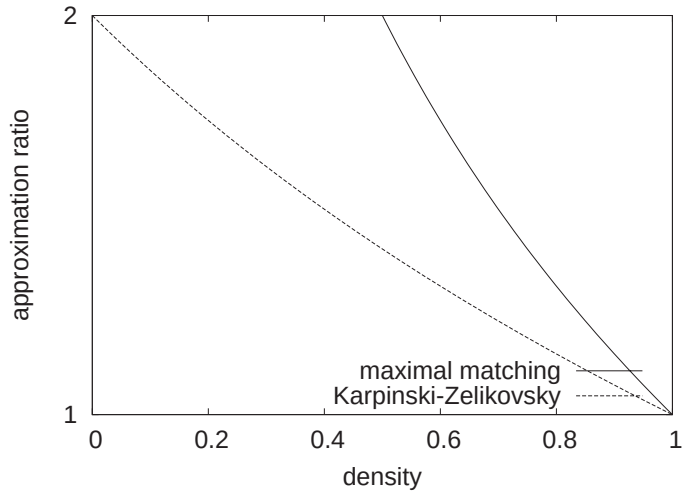


Figure 7.10: A comparison of the approximation ratios

Furthermore, the only graphs that maximize the approximation ratio among all graphs with n nodes and a minimum degree of $n-\alpha$ *when* $\delta > n/2$ *are in* $A_{n,\alpha}$.

Proof. The theorem follows directly from Lemmata 7.16 and 7.17. □

Corollary 7.4. *Let* $\tilde{\gamma}(\varepsilon,n)$ *be the worst approximation ratio attained by a strongly* ε*-dense graph. We have:*

$$\lim_{n\to\infty} \tilde{\gamma}(\varepsilon,n) = \begin{cases} 2 & \text{if } \varepsilon \leq 1/2, \\ \frac{1}{\varepsilon} & \text{otherwise .} \end{cases}$$

This asymptotic result is again to be compared with the result of [180] quoted in Section 7.3.2 (see Figure 7.10).

7.3.5 Minimum maximal matching

In this section we show that the analysis proposed for minimum node cover can be performed for minimum maximal matching as well. In this case the approximation ratio is a graph invariant defined by

$$\frac{\varphi(G)}{\nu(G)}.$$

Approximation ratio vs number of edges

Remark 7.9. The following results were again derived with the help of GraPHedron. In this problem, the set of invariants I is $\{\frac{\varphi}{\nu}, m\}$.

Lemma 7.18. *For any fixed k, the only graph G with n nodes that maximizes the number of edges among all graphs having* $\nu(G) = k$ *is* CS_{2k}^{n-2k}.

Proof. The nodes of a graph G with n nodes and $\nu(G) = k$ can be divided into two subsets: a subset A consisting of the matched nodes, and a subset $B = V \backslash A$. Note that no edge can be added inside the set B, since our matching is maximal. Clearly, adding all other possible edges to G maximizes the number of edges, without adding any edge to the minimum maximal matching. The graph we have thus constructed is clearly a CS_{2k}^{n-2k} graph, and is unique. □

Lemma 7.19. *Let n and m be positive integers such that* $m(CS_{n-\alpha-1}^{\alpha+1}) < m \leq m(CS_{n-\alpha}^{\alpha})$ *for some* α. *The minimum value for* $\nu(G)$ *attained by a graph G in* $\mathcal{G}_{n,m}$ *is* $\nu(CS_{n-\alpha}^{\alpha}) = \lceil \frac{n-\alpha}{2} \rceil$.

Proof. We first show that a graph in $\mathcal{G}_{n,m}$ cannot have a maximal matching of size less than $\nu(CS_{n-\alpha}^{\alpha})$. We then show that the value $\nu(CS_{n-\alpha}^{\alpha})$ is indeed attained by some graph in $\mathcal{G}_{n,m}$. The first step is by contradiction: we prove that if a graph $G \in \mathcal{G}_{n,m}$ has $\nu(G) < \nu(CS_{n-\alpha}^{\alpha})$ then $m(G) \leq m(CS_{n-\alpha-1}^{\alpha+1})$. We know by Lemma 7.18 that $m(G) \leq m\left(CS_{2(\nu(CS_{n-\alpha}^{\alpha})+1)}^{n-2(\nu(CS_{n-\alpha}^{\alpha})-1)}\right)$. Some basic algebra now gives us the desired result:

$$\begin{aligned} m(G) &\leq m\left(CS_{2(\nu(CS_{n-\alpha}^{\alpha})+1)}^{n-2(\nu(CS_{n-\alpha}^{\alpha})-1)}\right) = m\left(CS_{2\lceil\frac{n-\alpha}{2}\rceil-2}^{n+2-2\lceil\frac{n-\alpha}{2}\rceil}\right) \\ &\leq m\left(CS_{2(\frac{n-\alpha+1}{2})-2}^{n+2-2(\frac{n-\alpha+1}{2})}\right) = m\left(CS_{n-\alpha-1}^{\alpha+1}\right) \quad . \end{aligned}$$

There remains to present a graph in $\mathcal{G}_{n,m}$ that has a maximal matching of size $\nu(CS_{n-\alpha}^{\alpha})$. As in Lemma 7.14, let G be a graph formed by removing $m(CS_{n-\alpha}^{\alpha}) - m$ edges from the edges joining the clique and the independent set of a $CS_{n-\alpha}^{\alpha}$. The number of edges removed is between 0 and $\alpha - 1$, since $m(CS_{n-\alpha}^{\alpha}) - m(CS_{n-\alpha-1}^{\alpha+1}) = \alpha$. Thus each node in the clique has at least one neighbor in the independent set. In case $n - \alpha$ is even, taking any $\frac{n-\alpha}{2}$ disjoint edges in the clique will trivially yield a maximal matching of the desired size. In case $n - \alpha$ is odd, taking $\lfloor \frac{n-\alpha}{2} \rfloor$ disjoint edges in the clique and an additional edge joining the unmatched node of the clique and an arbitrary node of the independent set clearly yields a maximal matching of the desired size. □

Lemma 7.20. *Let n and m be positive integers such that* $m(CS_{n-\alpha-1}^{\alpha+1}) < m \leq m(CS_{n-\alpha}^{\alpha})$ *for some* α. *There exists a graph G in* $\mathcal{G}_{n,m}$ *such that* $\nu(G) = \nu(CS_{n-\alpha}^{\alpha})$ *and* $\varphi(G) = \varphi(CS_{n-\alpha}^{\alpha})$.

Proof. The graphs described in the proofs of Lemmata 7.15 and 7.19 satisfy $\nu(G) = \nu(CS_{n-\alpha}^{\alpha})$ and $\varphi(G) = \varphi(CS_{n-\alpha}^{\alpha})$. □

Using Lemmata 7.19 and 7.20 to maximize the numerator and minimize the denominator of the ratio, we obtain Theorem 7.5:

Theorem 7.5. *Let $\rho(G)$ be the worst approximation ratio for graph G. Let $\rho(m,n)$ be the worst approximation ratio attained by a graph in $\mathcal{G}_{n,m}$. For each n,m we have:*

$$\rho(m,n) \quad = \rho(CS^{\alpha^*(m,n)}_{n-\alpha^*(m,n)}) = \begin{cases} 2 & \text{if } \alpha^*(m,n) > n/2+1, \\ \frac{\lfloor \frac{n}{2} \rfloor}{\lceil \frac{n-\alpha^*(m,n)}{2} \rceil} & \text{otherwise.} \end{cases}$$

Proof. When $\alpha^*(m,n) \leq n/2+1$ or when $\alpha^*(m,n) > n/2+1$ and $n-\alpha$ is even, the proof follows the same scheme as the proof of Theorem 7.3, by using Lemmata 7.19 and 7.20. Otherwise, a value of 2 can easily be shown to be reached by a variant of the class of graphs used in the proof of Lemma 7.19. When $n \geq 4$, this variant consists in removing one edge*st* from the clique to the independent set such that $s-t-v-u$ is a path for some nodes v in the clique and u in the (formerly) independent set. □

The following corollary is easily proven using the same asymptotic arguments as in the proof of corollary 7.3:

Corollary 7.5. *Let $\tilde{\rho}(\varepsilon,n)$ be the worst approximation ratio attained by a graph with n nodes and an average degree at least εn. We have:*

$$\lim_{n\to\infty} \tilde{\rho}(\varepsilon,n) = \begin{cases} 2 & \text{if } \varepsilon \leq 3/4, \\ \frac{1}{1-\sqrt{1-\varepsilon}} & \text{otherwise.} \end{cases}$$

Approximation ratio vs minimum degree

Remark 7.10. In this problem, the set of invariants I is $\{\frac{\varphi}{v}, \delta\}$.

Let $B_{n,\delta}$ be the set of graphs of order n having minimum degree δ and a maximal matching of size $\lceil \delta/2 \rceil$. Note that $B_{n,\delta} = A_{n,n-\delta}$ when δ is even.

Lemma 7.21. *Each graph in $B_{n,\delta}$ with $\lceil \delta/2 \rceil > n/4$ has a perfect matching.*

Proof. Let us partition the nodes of G in two sets A and B such that B contains the endpoints of our maximal matching of size $\lceil \delta/2 \rceil$ (A is thus the complementary independent set, of size $n-2\lceil \delta/2 \rceil$). The formula $\lceil \delta/2 \rceil > n/4$ just expresses the condition $|A| < |B|$. In the case where δ is even, the proof is that of Lemma 7.16. If δ is odd, then $|B| = \delta+1$. We start from M, a perfect matching for B. We show that M can be augmented into a perfect matching. We show below that for any nodes a_i and a_j in A, there exists at most one $m \in M$ that does not provide an augmenting path for a_i and a_j. This implies that M can be augmented into a perfect matching for G by augmenting $\lfloor |A|/2 \rfloor$ disjoint pairs of nodes of A with different edges of M, since, having each element of a set X choose a different element of a set Y, with only one forbidden element in Y for each element of X is always possible when $|Y| > |X|$ (note here that $|M| > \lfloor |A|/2 \rfloor$).

The proof of our first assertion is by contradiction: suppose there are 2 edges $m = b_i b_j$ and $m' = b_k b_l$ that do not provide augmenting paths for a_i and a_j. This implies that in each one of the 4 following pairs of edges, at least one edge of the pair is absent from the graph:

$$\{(a_i b_i, a_j b_j), (a_i b_j, a_j b_i), (a_i b_k, a_j b_l), (a_i b_l, a_j b_k)\}$$

and therefore that: $deg(a_i) + deg(a_j) \leq 2|B| - 4 = 2\delta - 2 < 2\delta$, which contradicts the degree condition. □

Theorem 7.6 follows directly from Lemmata 7.12 and 7.21:

Theorem 7.6. *Let $\sigma(\delta, n)$ be the worst approximation ratio attained by a graph with n nodes and minimum degree δ. We have:*

$$\sigma(\delta, n) = \begin{cases} 2 & \text{if } \lceil \delta/2 \rceil \leq n/4 \\ \frac{\lfloor n/2 \rfloor}{\lceil \delta/2 \rceil} & \text{otherwise.} \end{cases}$$

Furthermore, when $\lceil \delta/2 \rceil > n/4$, $B_{n,\delta}$ is the exact set of graphs that maximize the ratio among all graphs with n nodes and minimum degree δ.

Proof. We consider the two cases $\lceil \delta/2 \rceil \leq n/4$ and $\lceil \delta/2 \rceil > n/4$.

i) $\lceil \delta/2 \rceil \leq n/4$. $CS^{n-2\lceil \delta/2 \rceil}_{2\lceil \delta/2 \rceil}$ minimizes $\nu(G)$ among all graphs G with n nodes and m edges (by noticing that $CS^{n-2\lceil \delta/2 \rceil}_{2\lceil \delta/2 \rceil} \in B_{n,\delta}$ and by applying Lemma 7.12). Further, $\varphi\left(CS^{n-2\lceil \delta/2 \rceil}_{2\lceil \delta/2 \rceil}\right) = 2\lceil \delta/2 \rceil$ (by Lemma 7.8), and thus $\sigma\left(CS^{n-2\lceil \delta/2 \rceil}_{2\lceil \delta/2 \rceil}\right) = 2\lceil \delta/2 \rceil / \lceil \delta/2 \rceil = 2$. This is the maximum value for the approximation ratio since our heuristic is a 2-approximation algorithm.

ii) $\lceil \delta/2 \rceil > n/4$. We know by Lemma 7.12 that the graphs in $B_{n,\delta}$ are the only graphs that minimize $\nu(G)$ among all graphs G with n nodes and m edges. We further know, by Lemma 7.21, that these graphs have a perfect matching. The extremal ratio is therefore given by $\lfloor n/2 \rfloor / \lceil \delta/2 \rceil$ and is attained by the graphs of $B_{n,\delta}$ only. □

Corollary 7.6. *Let $\tilde{\sigma}(\varepsilon, n)$ be the worst approximation ratio attained by a graph with n nodes and minimum degree at least εn. We have:*

$$\lim_{n \to \infty} \tilde{\sigma}(\varepsilon, n) = \begin{cases} 2 & \text{if } \varepsilon \leq 1/2, \\ \frac{1}{\varepsilon} & \text{otherwise.} \end{cases}$$

7.3.6 Remarks

We studied the worst-case approximation ratio of the maximal matching heuristic for the minimum node cover and minimum maximal matching problems in weakly and strongly ε-dense graphs.

For both problems in weakly ε-dense graphs, we characterized the exact worst-case approximation ratio as a function of ε and obtain a function that is asymptotic to 2 when $\varepsilon \leq 3/4$ and to $1/(1-\sqrt{1-\varepsilon})$ otherwise. In the case of strongly ε-dense graphs, again we characterized the exact worst-case approximation ratio as a function of ε and obtain a function that is asymptotic to 2 when $\varepsilon \leq 1/2$ and to $1/\varepsilon$ otherwise.

It is interesting to compare the approximation ratios we obtain for minimum node cover with the ones obtained by Zelikovsky and Karpinski: we note that in both the weakly and the strongly ε-dense cases the ratios differ only by one unit in the numerator and one in the denominator. The tight bounds we obtain for the minimum node cover problem are greater than those that were obtained using more sophisticated approximation algorithms [176, 180]. We nevertheless believe that a tight worst-case study of the classical heuristic is interesting as a point of comparison to the ratios obtained by other algorithms for the same problem, or by the same heuristic applied to other problems.

On the other hand, the approximation ratios obtained for minimum maximal matching are, to the best of our knowledge, the best ones known for this problem since it does not seem to have been studied under density constraints yet. Finally, the results obtained for this problem are also valid for a variant problem, namely the *minimum edge dominating set* problem since, as noted by Yannakakis and Gavril [233], both problems always admit optimal solutions of the same size, and an optimal solution to one can always be transformed into an optimal solution to the other in polynomial time.

7.4 Summary

In Section 7.1, we have done a complete analysis of the two invariants α and m, viewed as coordinates in the space. The full list of optimal linear inequalities between these two invariants was given. We have pointed out that any point from the invariants polytopes which has integer coordinates is produced by some connected graph with n nodes. This analysis gives also an answer to the following problem stated by Ore [202] in 1962 : *What is the minimum number of edges m in connected graphs with a fixed number of nodes n and a fixed stability number* α *?*.

In Section 7.2 we have established some of the optimal inequalities among the diameter, the irregularity and the maximum degree of a connected graph. Even if we are not able this time to provide a complete description of the polytope of graph invariants, we have still shown that our method leads to nontrivial and interesting results. The problem of finding all optimal linear inequalities for the diameter, the irregularity and the maximum degree is left open.

The maximal matching heuristic is an algorithm that provides a 2-approximation for the minimum node cover and minimum maximal matching problems. In Section 7.3, we have asked the

question of expressing the approximation ratio in a finer way, as a function of density parameters. We have characterized precisely the asymptotic approximation ratio as a function of these parameters, together with tight examples. This is, to our knowledge, the tightest analysis ever done of this algorithm.

All these results show that the system GraPHedron and its polyhedral approach, presented in Chapter 6, provide a fruitful strategy to derive conjectures in graph theory. Moreover, the vertex graphs produced by the system were very helpful in the writing of the proofs.

Conclusion

Several computer systems in graph theory have been very successful in helping mathematicians to formulate and explore conjectures. Sometimes such conjectures are suggested by the computer in a entirely automated way. All these systems appear to be complementary. The underlying approaches, i.e., enumeration, interactive computing, formula manipulation, generation and selection, heuristic optimization, are diverse.

In this dissertation, we used intensively the system AutoGraphiX, based on the heuristic optimization approach. It led to many new results, mainly in chemical graph theory. We also used the enumeration approach to characterize some infinite families of integral graphs. Moreover, in the context of this dissertation, we have developed the system GraPHedron. It uses a new approach, partially based on enumeration, which relies on polyhedral arguments to derive optimal conjectures, under some conditions. We also present the first new results obtained with its help. For instance, we were able to derive the full list of optimal linear inequalities, between the stability number and the number of edges of connected graphs with n nodes. We also provided a precise characterization of the asymptotic approximation ratio of the maximal matching heuristic, as a function of density parameters, for the minimum node cover and the minimum maximal matching problems.

With this work, we got closer to the initial objective : to show how computers can help researchers in graph theory. We showed by numerous examples that conjectures can often be easily derived with the help of a computer, especially in graph theory. Good examples are important both in suggesting conjectures and in proving theorems. These examples are particular graphs for a given problem. For instance, in this dissertation we intensively studied :

i) family of graphs obtained by constructive enumeration (see Section 2.1 and Chapter 3);

ii) extremal graphs provided by AutoGraphiX (see Section 2.5 and Chapters 4 and 5);

iii) vertex graphs given by GraPHedron (see Chapters 6 and 7).

The usefulness of the AutoGraphiX and GraPHedron systems was illustrated by many results.

Some of these results would probably not have been found without the help of a computer (see for instance Conjecture 2.1 and its proof in Appendix A.1). Others, even if quite intuitive, are not trivial to prove (see, e.g., Theorem 7.2, a variant of a famous Theorem of Turán). Another interesting feature of the use of computers is that the proof itself can be derived by the *way* the system gives its output, and not only using examples of graphs. It was used in Chapter 5 in a *simulated algorithm* type of proof, based on the neighborhoods used to find extremal graphs with AutoGraphiX.

One can maybe think, *a priori*, that conjectures obtained with the help of computers are trivial. Of course, if one considers connected graphs, a computer will often suggest that

$$m \geq n - 1.$$

However, the theoretical results presented in this dissertation are often nontrivial and of theoretical interest. They led to 10 publications or submitted papers. For instance, we recall that Theorem 7.2 is an answer to a problem listed as open since 1962. Another example is the interest, in the chemical community, caused by the results presented in Chapter 4 and published in [146, 163, 165].

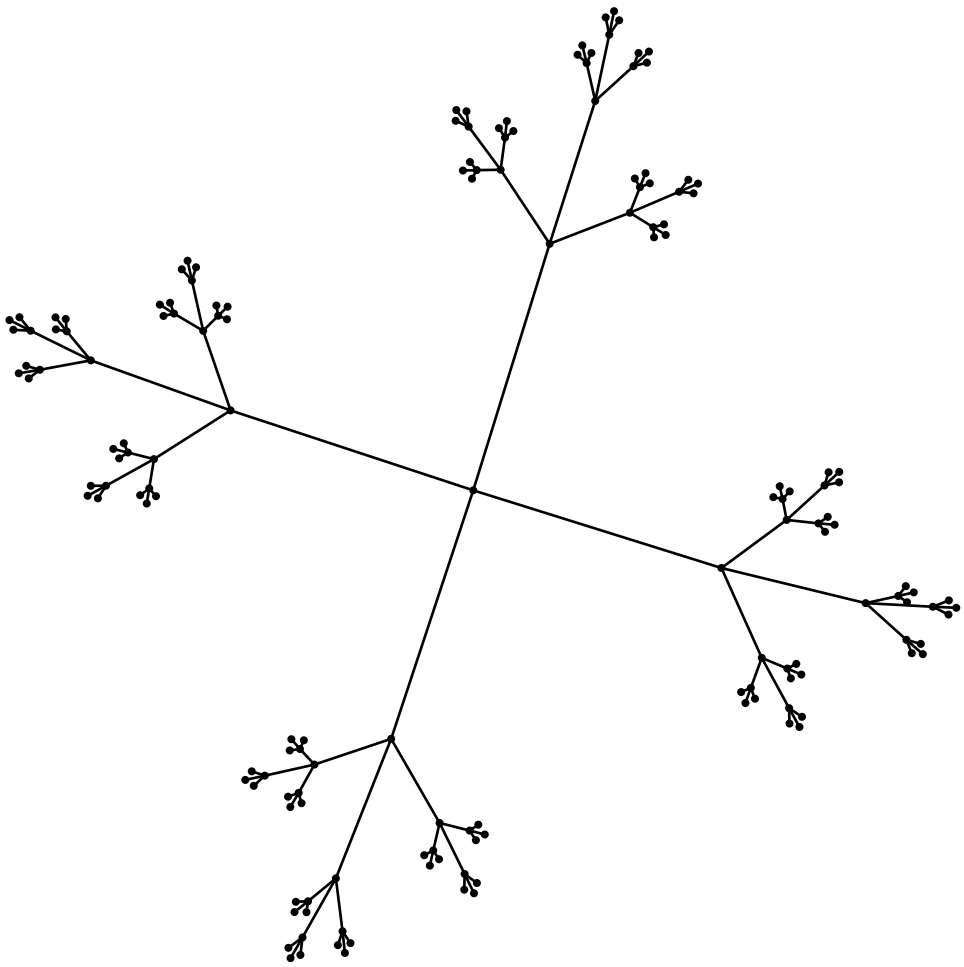
There are many future perspectives in developing computer systems in graph theory. For instance, the new version of AutoGraphiX makes a fully automatic comparison of invariants and led to many automated conjectures (see Section 2.5). The system GraPHedron can also be improved. Much can be done in the automation of conjectures, such as the recognition and generalization of families of inequalities. We also think that the current limitation on the order of the graphs used in GraPHedron, due to a complete enumeration of these graphs, can be overcome. An interesting possibility is to use a heuristic, such as the VNS used in AutoGraphiX, to construct polytopes iteratively, without enumerating all the points. Another possibility is to avoid the generation of some graphs which are not vertices of the polytopes, using a selective enumeration. Of course, it will be hard to realize : in this case, the enumeration depends on the invariants studied.

Other perspectives are theoretical. AutoGraphiX and GraPHedron have proved their efficiency. We want now to devote more time to using these systems, in order to obtain new theoretical results. We have started the study of some interesting problems. For instance, the study of approximation ratios (see Section 7.3) can be extended to other heuristics and exact algorithms. A new family of NP-hard invariants have recently been added to GraPHedron. They are related to computational geometry and are, e.g., the convex crossing number, the path width, the tree width, the queue number, ... We have begun to study them in collaboration with researchers specialized in this field.

We conclude this dissertation by two citations. GraPHedron was first presented publicly at the second *Computers and Discovery* conference (Montréal 2004). During this meeting, Stevanović said to us, after playing a little bit with GraPHedron : *"It can gives more conjectures than we have the time to prove."* The same remark can be stated for AutoGraphiX. The real challenge is

maybe to find unexplored and interesting conjectures. If time for proving (or refuting) conjectures is missing, the following statement is maybe the good answer. It was made by Hansen during the *18th mini-EURO conference on VNS* (Tenerife, 2005) to conclude an interesting talk about automated comparison of invariants made with AutoGraphiX : *"When you get many conjectures, you can let the computer prove the simplest ones, then prove yourself interesting ones, and give the most difficult ones to colleagues."*

Appendix



APPENDIX A

About AutoGraphiX

Proof of Conjecture 2.1

A.1 Proof of Conjecture 2.1

The trees with maximum degree $\Delta \leq 3$ found by AutoGraphiX with (conjectured) maximum irregularity are represented on Figure A.1.

These extremal trees are used in the following proofs, illustrating also the help provided by AutoGraphiX in getting proofs.

Theorem A.1. *For any tree T with $\Delta \leq 3$,*

$$\begin{aligned} A(T) &\leq \tfrac{4n+2}{3} && \text{if } n(mod\ 3) = 1, \\ &\leq \tfrac{4n-n(mod\ 3)}{2} && \text{otherwise.} \end{aligned}$$

Proof. Let T be a tree with maximum degree $\Delta \leq 3$ and denote by x_{ij} the number of edges of T with endnodes of degree i and j.

By definition of the irregularity,

$$A(T) = x_{12} + 2x_{13} + x_{23}. \tag{A.1}$$

We first solve the following system of five linear equations which holds for all trees with

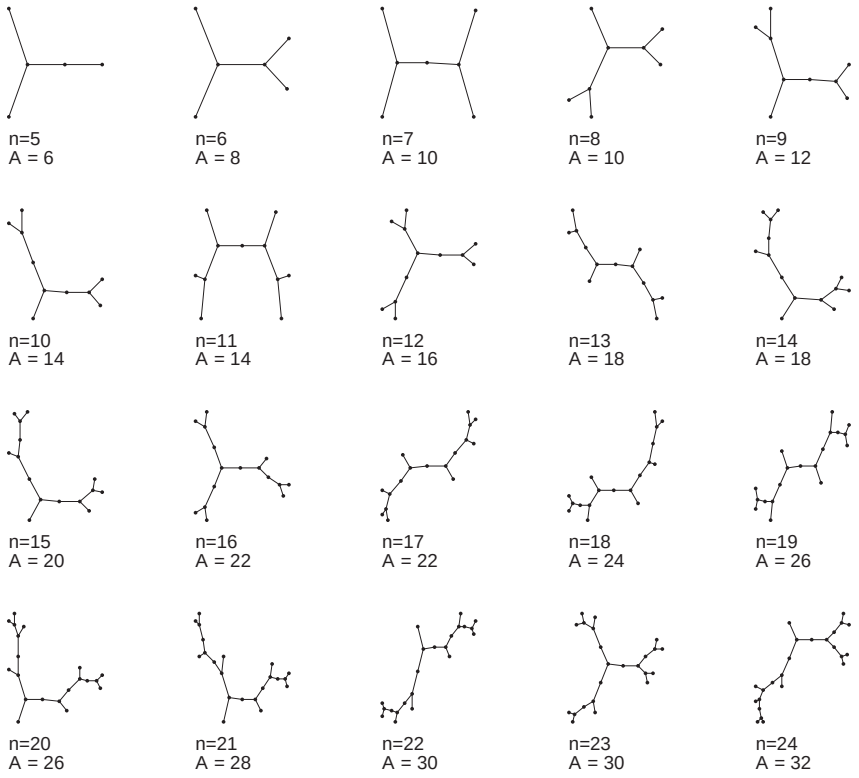


Figure A.1: Extremal trees with $\Delta \leq 3$ and maximum irregularity found by AutoGraphiX

$\Delta \leq 3$:

$$x_{12} + x_{13} = n_1 \tag{A.2}$$
$$x_{12} + 2x_{22} + x_{23} = 2n_2 \tag{A.3}$$
$$x_{13} + x_{23} + 2x_{33} = 3n_3 \tag{A.4}$$
$$n_1 + 2n_2 + 3n_3 = 2n - 2 \tag{A.5}$$
$$n_1 + n_2 + n_3 = n. \tag{A.6}$$

with unkowns x_{13}, x_{23}, n_1, n_2 and n_3. That gives :

$$x_{13} = \frac{1}{3}(n - 4x_{12} - x_{22} + x_{33} + 5) \tag{A.7}$$
$$x_{23} = \frac{1}{3}(2n + x_{12} - 2x_{22} - 4x_{33} - 8) \tag{A.8}$$
$$n_1 = \frac{1}{3}(n - x_{12} - x_{22} + x_{33} + 5) \tag{A.9}$$
$$n_2 = \frac{1}{3}(n + 2x_{12} + 2x_{22} - 2x_{33} - 4) \tag{A.10}$$
$$n_3 = \frac{1}{3}(n - x_{12} - x_{22} + x_{33} - 1). \tag{A.11}$$

Replacing x_{13} by (A.7) and x_{23} by (A.8) in (A.1) gives

$$A(G) = \frac{1}{3}(4n - 4x_{12} - 4x_{22} - 2x_{33} + 2) \tag{A.12}$$

which is maximal for a fixed number of nodes when the values x_{12} and x_{33} are equal to zero.

If $n(\text{mod } 3) = 1$, we can choose $x_{12} = 0$, $x_{22} = 0$ and $x_{33} = 0$ because the solutions given in Eqs. (A.7) – (A.11) are in integers. In this case, $x_{13} = (n+5)/3$, $x_{23} = (2n-8)/3$ and $A(T) = (4n+2)/3$.

If $n(\text{mod } 3) = 0$, x_{12}, x_{22} and x_{33} cannot be all equal to zero because the solutions are no more in integers. Looking at (A.12), the best choice is to take $x_{12} = x_{22} = 0$ and $x_{33} = 1$ which is a feasible case. In this case, $A(T) = 4n/3$.

If $n(\text{mod } 3) = 2$, there are three feasible solutions with the same irregularity value. One can choose $x_{12} = x_{22} = 0$ and $x_{33} = 2$, or $x_{12} = 1$ and $x_{22} = x_{33} = 0$, or $x_{22} = 1$ and $x_{12} = x_{33} = 0$. These solutions lead to $A(T) = (4n-2)/3$. Figure A.2 shows three different trees with maximum irregularity and $n = 8$. □

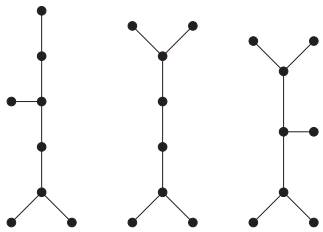


Figure A.2: Three trees with maximum irregularity, $\Delta \leq 3$ and $n = 8$

The graphs found by AutoGraphiX (see Figure A.1) are extremal for the irregularity by Theorem A.1. The proof of this theorem gives a good characterization of these graphs in terms of x_{ij}. We now prove Conjecture 2, which was obtained automatically by AutoGraphiX from these extremal trees.

Theorem A.2. *For any tree T of size m with $\Delta \leq 3$ and maximum irregularity $A(T)$, Randic index $Ra(T)$, and n_1 pending nodes,*

$$Ra(T) = -0.027421\, A(T) + 0.538005\, m - 0.110484\, n_1 + 0.614014.$$

Proof. Before proceeding to the proof itself, we find which real values AutoGraphiX has approximated. To do this, we choose 4 extremal trees given by the system (see Figure A.3), compute their values for Ra, A, m and n_1 and subsitute these values in

$$Ra = a\, A + b\, m + c\, n_1 + d \tag{A.13}$$

where a,b,c,d are the real values sought for. For instance, the tree T_1 on Figure A.3 has $Ra(T_1) = 1/\sqrt{2}+2/\sqrt{3}+1/\sqrt{6}$, $A(T_1) = 6$, $m(T_1) = 4$ and $n_1(T_1) = 3$. That gives the following system of equations with unknows a,b,c and d :

$$\begin{aligned} 6a+4b+3c+d &= \frac{1}{\sqrt{2}}+\frac{2}{\sqrt{3}}+\frac{1}{\sqrt{6}}, && \text{(A.14)}\\ 8a+5b+4c+d &= \frac{4}{\sqrt{3}}+\frac{1}{3}, && \text{(A.15)}\\ 10a+6b+4c+d &= \frac{4}{\sqrt{3}}+\frac{2}{\sqrt{6}}, && \text{(A.16)}\\ 10a+7b+5c+d &= \frac{5}{\sqrt{3}}+\frac{2}{3}. && \text{(A.17)} \end{aligned}$$

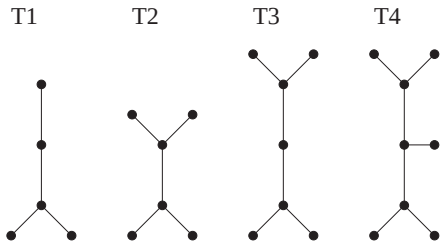


Figure A.3: Four extremal trees with $\Delta \leq 3$ and maximum irregularity found by AutoGraphiX

The unique solution of this system is

$$\begin{aligned} a &= -\frac{\sqrt{2}}{4}+\frac{\sqrt{3}}{6}+\frac{\sqrt{6}}{12}-\frac{1}{6}, && \text{(A.18)}\\ b &= \frac{\sqrt{2}}{2}-\frac{\sqrt{3}}{3}+\frac{\sqrt{6}}{6}, && \text{(A.19)}\\ c &= -\frac{\sqrt{2}}{2}+\frac{2\sqrt{3}}{3}-\frac{\sqrt{6}}{2}+\frac{2}{3}, && \text{(A.20)}\\ d &= \frac{3\sqrt{2}}{2}-\sqrt{3}+\frac{\sqrt{6}}{2}-1. && \text{(A.21)} \end{aligned}$$

A numerical approximation of these irrational values corresponds to the values given by AutoGraphiX in the conjecture.

Let T be a tree with maximum degree $\Delta \leq 3$. We have that

$$m = x_{12}+x_{13}+x_{22}+x_{23}+x_{33}, \quad \text{(A.22)}$$

and

$$n = x_{12}+x_{13}+x_{22}+x_{23}+x_{33}+1. \quad \text{(A.23)}$$

Moreover, by definition of the irregularity

$$A(T) = x_{12}+2x_{13}+x_{23}, \quad \text{(A.24)}$$

and by definition of the Randic index

$$Ra(T) = \frac{x_{12}}{\sqrt{2}} + \frac{x_{13}}{\sqrt{3}} + \frac{x_{22}}{2} + \frac{x_{23}}{\sqrt{6}} + \frac{x_{33}}{3}. \tag{A.25}$$

By Theorem A.1, if $n(\text{mod } 3) = 1$,

$$x_{13} = (n+5)/3, \tag{A.26}$$

and

$$x_{12} = x_{22} = x_{33} = 0. \tag{A.27}$$

Substituting (A.27) in (A.23) and (A.23) in (A.26) gives

$$x_{23} = 2x_{13} - 6. \tag{A.28}$$

By (A.27) and (A.28), Eqs. (A.22), (A.24) and (A.25) become

$$m = 3x_{13} - 6, \tag{A.29}$$

$$A(T) = 4x_{13} - 6, \tag{A.30}$$

and

$$Ra(T) = x_{13}\frac{\sqrt{3}+\sqrt{6}}{3} - \sqrt{6}, \tag{A.31}$$

respectively. Moreover, Eq. (A.2) gives

$$n_1 = x_{13} \tag{A.32}$$

Replace A by (A.30), m by (A.29), n_1 by (A.32) and a,b,c,d by Eqs. (A.18) – (A.21) in the right-hand-side of (A.13) and simplify. This leads to

$$x_{13}\frac{\sqrt{3}+\sqrt{6}}{3} - \sqrt{6},$$

which is equal to the Randic index of T given by (A.32).

The other cases are similar.

If $n(\text{mod } 3) = 0$, we start with $x_{13} = (n+6)/3$, $x_{33} = 1$ and $x_{12} = x_{22} = 0$ and modify the remainder of the proof in consequence.

If $n(\text{mod } 3) = 2$ we start with the three different solutions given in Theorem A.1 and apply the same ideas for each cases. □

APPENDIX B

About GraPHedron

B.1 GraPHedron's report for Problem 6.1

The following code is an input file to solve Problem 6.1 with GraPHedron.

```
/*  Which are all the best linear inequalities among D and m ? */
<ORDER> NMIN 4 NMAX 10 </ORDER>
<COORDINATES> Diam; NumEdges </COORDINATES>
<CONDITIONS> Connected </CONDITIONS>  // Optional (automatic cond. of Diam)
<REPORT>TITLE "Diameter and Number of Edges" </REPORT>
<COMPUTATION> DISTRIBUTION </COMPUTATION>
```

A copy of the report created by GraPHedron automatically is presented on the following pages.

Diameter and Number of Edges

GraPHedron 0.6.1 - Automatic Report

Thu Feb 9 13:06:13 2006

Contents

1 Problem statement

Definition 1 *A graph G is* connected *is every pair of nodes in G are joined by a path.*

Definition 2 *The* diameter *D of a graph $G = (V, E)$ is the maximum distance between two nodes of G, i.e. the "longest shortest path" in G.*

Definition 3 *The* number of edges *m of a graph $G = (V, E)$, or its* size*, is the cardinal of E.*

The class of graphs generated are all non-isomorphic *graphs* with n nodes. The dimension of the problem is 2 and the coordinates are:

$$\begin{aligned} x_1 &= D, \\ x_2 &= m. \end{aligned}$$

where D is the *Diameter* and m is the *Number of edges*.

The graphs considered respect the following condition(s):

$$C_1 : \quad \textit{Connected}.$$

where *Connected* is true if the graphs are connected. In the following, a polyhedron $\mathbb{P}_n$ is thus defined by

$$\begin{aligned} \mathbb{P}_n = \quad & \text{conv}\{(x_1, x_2) \mid \exists\, G \in \mathcal{C}_n, \\ & x_1 = D(G),\ x_2 = m(G)\}, \end{aligned}$$

where $\mathcal{C}_n$ is the set of graphs with n nodes such that C_1 is respected.

The polyhedra computed are those for which $n = 4$ to 10. Only one graph is drawn for each vertex of the polyhedron.

2 Polyhedron $n = 4$

Information about polyhedron $\mathbb{P}_4$ of graphs with $n = 4$:

# of graphs generated	11
# of gr. satisf. cond.	6
# of unique coord.	5
# of vertices of $\mathbb{P}_4$	4
# of facets of $\mathbb{P}_4$	4
Data precomputed (CPU time)	0.00 sec
Data computed now (CPU time)	0.00 sec
Total CPU time to create $\mathbb{P}_4$	0.00 sec

The polyhedron $\mathbb{P}_4$ is represented in Figure 1.

2.1 Facets

$$D + m \leq 7 \tag{1}$$
$$2D + m \leq 9 \tag{2}$$
$$m \geq 3 \tag{3}$$
$$3D + m \geq 9 \tag{4}$$

Facets	Vertices in facet
(1)	[1, 6] [2, 5]
(2)	[2, 5] [3, 3]
(3)	[2, 3] [3, 3]
(4)	[1, 6] [2, 3]

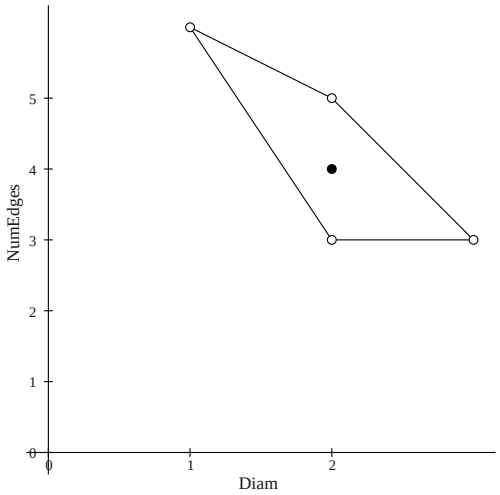


Figure 1: Representation of $\mathbb{P}_4$

2.2 Vertices

Coordinates	# of graphs	Facets
$[1, 6]$	1	(1) (4)
$[2, 3]$	1	(3) (4)
$[2, 5]$	1	(1) (2)
$[3, 3]$	1	(2) (3)

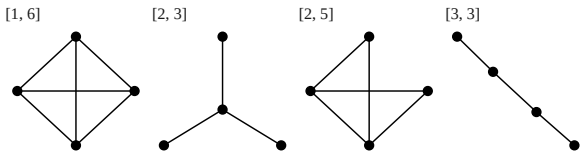


Figure 2: Graphs which are vertices of $\mathbb{P}_4$

Graphs which are polyhedral vertices of $\mathbb{P}_4$ are represented in Figure 2.

2.3 Pareto points

No pareto points found for this polytope.

3 Polyhedron $n = 5$

Information about polyhedron $\mathbb{P}_5$ of graphs with $n = 5$:

# of graphs generated	34
# of gr. satisf. cond.	21
# of unique coord.	11
# of vertices of $\mathbb{P}_5$	4
# of facets of $\mathbb{P}_5$	4
Data precomputed (CPU time)	0.00 sec
Data computed now (CPU time)	0.00 sec
Total CPU time to create $\mathbb{P}_5$	0.00 sec

The polyhedron $\mathbb{P}_5$ is represented in Figure 3.

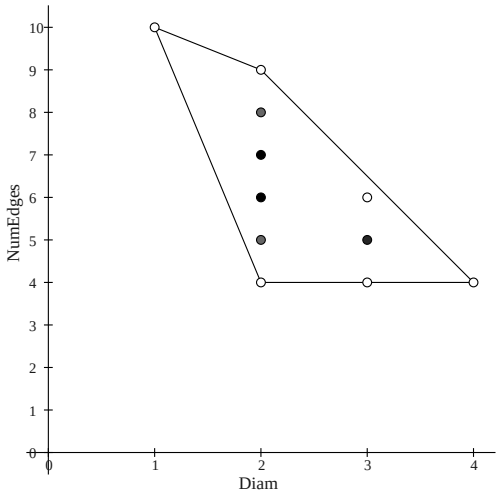


Figure 3: Representation of $\mathbb{P}_5$

3.1 Facets

$$D + m \leq 11 \quad (5)$$
$$\frac{5}{2}D + m \leq 14 \quad (6)$$
$$m \geq 4 \quad (7)$$
$$6D + m \geq 16 \quad (8)$$

Facets	Vertices in facet
(5)	$[1, 10]$ $[2, 9]$
(6)	$[2, 9]$ $[4, 4]$
(7)	$[2, 4]$ $[4, 4]$
(8)	$[1, 10]$ $[2, 4]$

3.2 Vertices

Coordinates	# of graphs	Facets
$[1, 10]$	1	(5) (8)
$[2, 4]$	1	(7) (8)
$[2, 9]$	1	(5) (6)
$[4, 4]$	1	(6) (7)

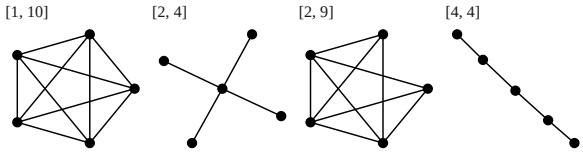


Figure 4: Graphs which are vertices of $\mathbb{P}_5$

Graphs which are polyhedral vertices of $\mathbb{P}_5$ are represented in Figure 4.

3.3 Pareto points

Pareto points related to facet $\frac{5}{2}D + m \leq 14$

Coordinates	# of graphs
$[2, 9]$	1
$[3, 6]$	1
$[4, 4]$	1

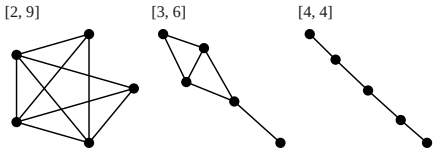


Figure 5: Pareto points of facet $\frac{5}{2}D + m \leq 14$

Pareto points of facet $\frac{5}{2}D + m \leq 14$ are represented in Figure 5.

4 Polyhedron $n = 6$

Information about polyhedron $\mathbb{P}_6$ of graphs with $n = 6$:

# of graphs generated	156
# of gr. satisf. cond.	112
# of unique coord.	21
# of vertices of $\mathbb{P}_6$	4
# of facets of $\mathbb{P}_6$	4
Data precomputed (CPU time)	0.00 sec
Data computed now (CPU time)	0.01 sec
Total CPU time to create $\mathbb{P}_6$	0.01 sec

The polyhedron $\mathbb{P}_6$ is represented in Figure 6.

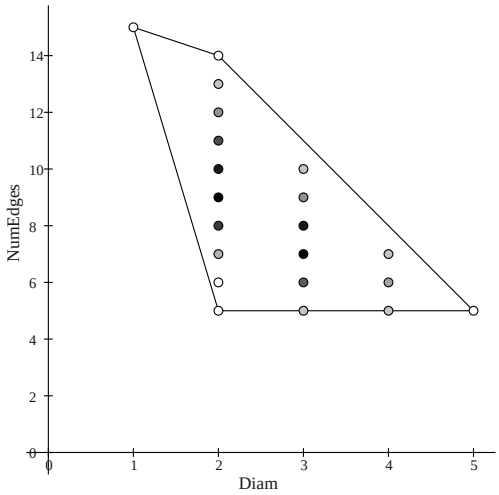


Figure 6: Representation of $\mathbb{P}_6$

4.1 Facets

$$D + m \leq 16 \quad (9)$$
$$3D + m \leq 20 \quad (10)$$
$$m \geq 5 \quad (11)$$
$$10D + m \geq 25 \quad (12)$$

Facets	Vertices in facet
(9)	[1, 15] [2, 14]
(10)	[2, 14] [5, 5]
(11)	[2, 5] [5, 5]
(12)	[1, 15] [2, 5]

4.2 Vertices

Coordinates	# of graphs	Facets
[1, 15]	1	(9) (12)
[2, 5]	1	(11) (12)
[2, 14]	1	(9) (10)
[5, 5]	1	(10) (11)

Graphs which are polyhedral vertices of $\mathbb{P}_6$ are represented in Figure 7.

4.3 Pareto points

Pareto points related to facet $3D + m \leq 20$

Coordinates	# of graphs
[2, 14]	1
[3, 10]	2
[4, 7]	2
[5, 5]	1

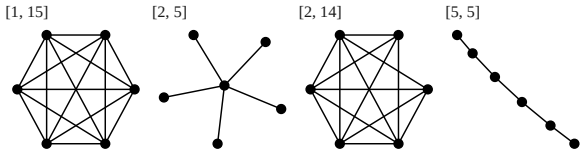


Figure 7: Graphs which are vertices of $\mathbb{P}_6$

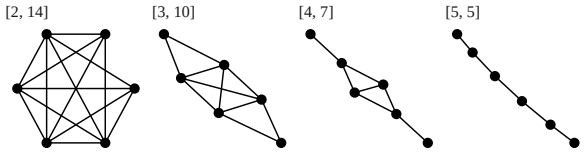


Figure 8: Pareto points of facet $3D + m \leq 20$

Pareto points of facet $3D+m \leq 20$ are represented in Figure 8. All these Pareto points respect the following equality :

$$m = \frac{1}{2}D^2 - \frac{13}{2}D + 25 \tag{13}$$

5 Polyhedron $n = 7$

Information about polyhedron $\mathbb{P}_7$ of graphs with $n = 7$:

# of graphs generated	1044
# of gr. satisf. cond.	853
# of unique coord.	36
# of vertices of $\mathbb{P}_7$	4
# of facets of $\mathbb{P}_7$	4
Data precomputed (CPU time)	0.01 sec
Data computed now (CPU time)	0.03 sec
Total CPU time to create $\mathbb{P}_7$	0.04 sec

The polyhedron $\mathbb{P}_7$ is represented in Figure 9.

5.1 Facets

$$D + m \leq 22 \tag{14}$$
$$\frac{7}{2}D + m \leq 27 \tag{15}$$
$$m \geq 6 \tag{16}$$
$$15D + m \geq 36 \tag{17}$$

Facets	Vertices in facet
(14)	$[1, 21]$ $[2, 20]$
(15)	$[2, 20]$ $[6, 6]$
(16)	$[2, 6]$ $[6, 6]$
(17)	$[1, 21]$ $[2, 6]$

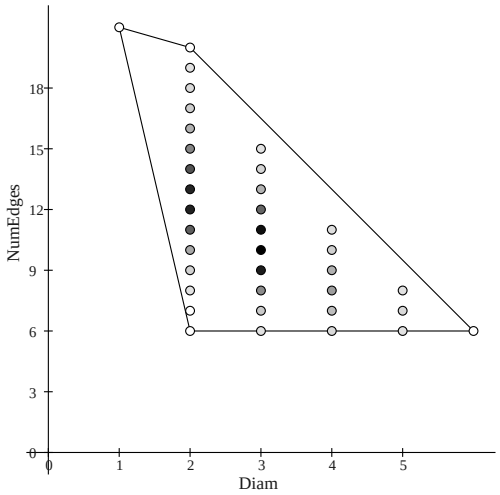


Figure 9: Representation of $\mathbb{P}_7$

5.2 Vertices

Coordinates	# of graphs	Facets
$[1, 21]$	1	(14) (17)
$[2, 6]$	1	(16) (17)
$[2, 20]$	1	(14) (15)
$[6, 6]$	1	(15) (16)

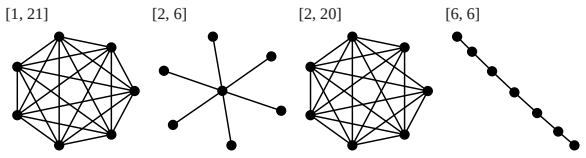


Figure 10: Graphs which are vertices of $\mathbb{P}_7$

Graphs which are polyhedral vertices of $\mathbb{P}_7$ are represented in Figure 10.

5.3 Pareto points

Pareto points related to facet $\frac{7}{2}D + m \leq 27$

Coordinates	# of graphs
$[2, 20]$	1
$[3, 15]$	2
$[4, 11]$	3
$[5, 8]$	2
$[6, 6]$	1

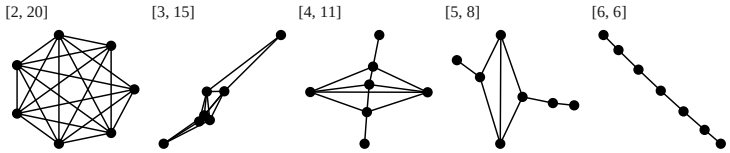


Figure 11: Pareto points of facet $\frac{7}{2}D + m \leq 27$

Pareto points of facet $\frac{7}{2}D + m \leq 27$ are represented in Figure 11. All these Pareto points respect the following equality :

$$m = \frac{1}{2}D^2 - \frac{15}{2}D + 33 \tag{18}$$

6 Polyhedron $n = 8$

Information about polyhedron $\mathbb{P}_8$ of graphs with $n = 8$:

# of graphs generated	12346
# of gr. satisf. cond.	11117
# of unique coord.	57
# of vertices of $\mathbb{P}_8$	4
# of facets of $\mathbb{P}_8$	4
Data precomputed (CPU time)	0.15 sec
Data computed now (CPU time)	0.22 sec
Total CPU time to create $\mathbb{P}_8$	0.37 sec

The polyhedron $\mathbb{P}_8$ is represented in Figure 12.

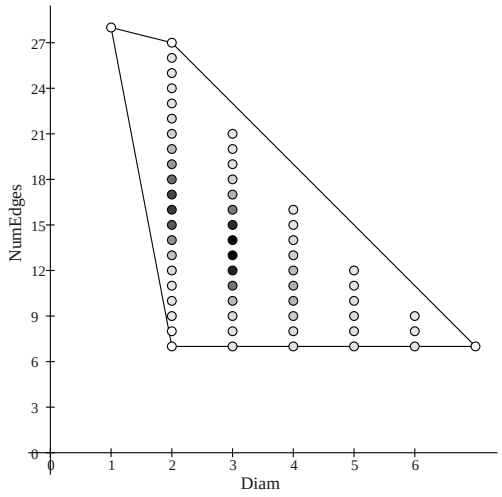


Figure 12: Representation of $\mathbb{P}_8$

6.1 Facets

$$D + m \leq 29 \quad (19)$$
$$4D + m \leq 35 \quad (20)$$
$$m \geq 7 \quad (21)$$
$$21D + m \geq 49 \quad (22)$$

Facets	Vertices in facet
(19)	[1, 28] [2, 27]
(20)	[2, 27] [7, 7]
(21)	[2, 7] [7, 7]
(22)	[1, 28] [2, 7]

6.2 Vertices

Coordinates	# of graphs	Facets
[1, 28]	1	(19) (22)
[2, 7]	1	(21) (22)
[2, 27]	1	(19) (20)
[7, 7]	1	(20) (21)

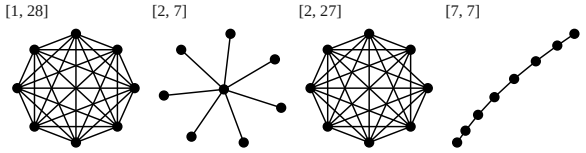


Figure 13: Graphs which are vertices of $\mathbb{P}_8$

Graphs which are polyhedral vertices of $\mathbb{P}_8$ are represented in Figure 13.

6.3 Pareto points

Pareto points related to facet $4D + m \leq 35$

Coordinates	# of graphs
[2, 27]	1
[3, 21]	3
[4, 16]	4
[5, 12]	4
[6, 9]	3
[7, 7]	1

Pareto points of facet $4D + m \leq 35$ are represented in Figure 14. All these Pareto points respect the following equality :

$$m = \frac{1}{2}D^2 - \frac{17}{2}D + 42 \quad (23)$$

7 Polyhedron $n = 9$

Information about polyhedron $\mathbb{P}_9$ of graphs with $n = 9$:

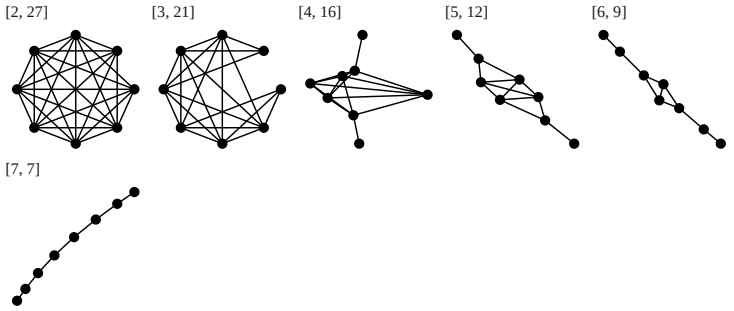


Figure 14: Pareto points of facet $4D + m \leq 35$

# of graphs generated	274668
# of gr. satisf. cond.	261080
# of unique coord.	85
# of vertices of $\mathbb{P}_9$	4
# of facets of $\mathbb{P}_9$	4
Data precomputed (CPU time)	4.17 sec
Data computed now (CPU time)	4.69 sec
Total CPU time to create $\mathbb{P}_9$	8.86 sec

The polyhedron $\mathbb{P}_9$ is represented in Figure 15.

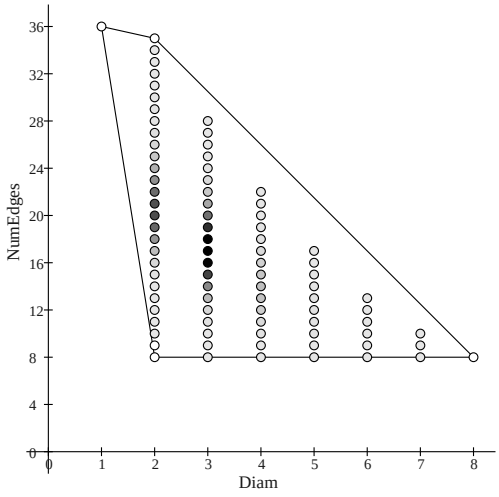


Figure 15: Representation of $\mathbb{P}_9$

7.1 Facets

$$D + m \leq 37 \quad (24)$$
$$\frac{9}{2}D + m \leq 44 \quad (25)$$
$$m \geq 8 \quad (26)$$
$$28D + m \geq 64 \quad (27)$$

Facets	Vertices in facet
(24)	$[1, 36]$ $[2, 35]$
(25)	$[2, 35]$ $[8, 8]$
(26)	$[2, 8]$ $[8, 8]$
(27)	$[1, 36]$ $[2, 8]$

7.2 Vertices

Coordinates	# of graphs	Facets
$[1, 36]$	1	(24) (27)
$[2, 8]$	1	(26) (27)
$[2, 35]$	1	(24) (25)
$[8, 8]$	1	(25) (26)

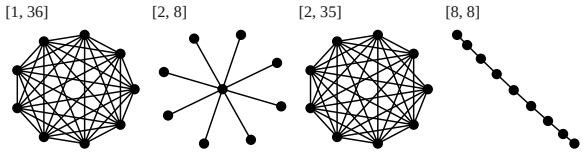


Figure 16: Graphs which are vertices of $\mathbb{P}_9$

Graphs which are polyhedral vertices of $\mathbb{P}_9$ are represented in Figure 16.

7.3 Pareto points

Pareto points related to facet $\frac{9}{2}D + m \leq 44$

Coordinates	# of graphs
$[2, 35]$	1
$[3, 28]$	3
$[4, 22]$	5
$[5, 17]$	5
$[6, 13]$	5
$[7, 10]$	3
$[8, 8]$	1

Pareto points of facet $\frac{9}{2}D + m \leq 44$ are represented in Figure 17. All these Pareto points respect the following equality :

$$m = \frac{1}{2}D^2 - \frac{19}{2}D + 52 \quad (28)$$

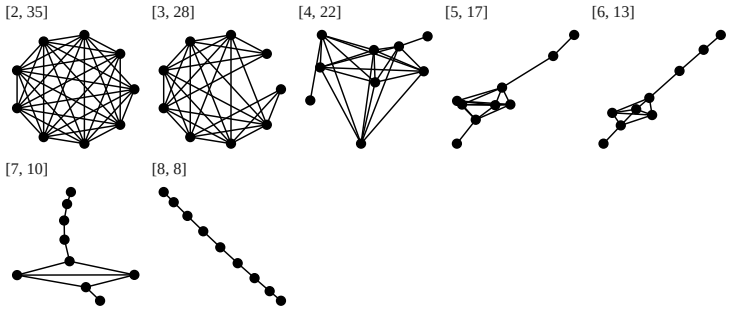


Figure 17: Pareto points of facet $\frac{9}{2}D + m \leq 44$

8 Polyhedron $n = 10$

Information about polyhedron $\mathbb{P}_{10}$ of graphs with $n = 10$:

# of graphs generated	12005168
# of gr. satisf. cond.	11716571
# of unique coord.	121
# of vertices of $\mathbb{P}_{10}$	4
# of facets of $\mathbb{P}_{10}$	4
Data precomputed (CPU time)	3 min 49.37 sec
Data computed now (CPU time)	3 min 25.95 sec
Total CPU time to create $\mathbb{P}_{10}$	7 min 15.32 sec

The polyhedron $\mathbb{P}_{10}$ is represented in Figure 18.

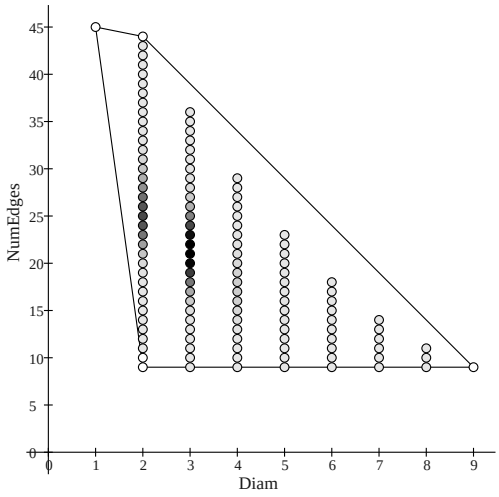


Figure 18: Representation of $\mathbb{P}_{10}$

8.1 Facets

$$D + m \leq 46 \quad (29)$$
$$5D + m \leq 54 \quad (30)$$
$$m \geq 9 \quad (31)$$
$$36D + m \geq 81 \quad (32)$$

Facets	Vertices in facet
(29)	$[1, 45]$ $[2, 44]$
(30)	$[2, 44]$ $[9, 9]$
(31)	$[2, 9]$ $[9, 9]$
(32)	$[1, 45]$ $[2, 9]$

8.2 Vertices

Coordinates	# of graphs	Facets
$[1, 45]$	1	(29) (32)
$[2, 9]$	1	(31) (32)
$[2, 44]$	1	(29) (30)
$[9, 9]$	1	(30) (31)

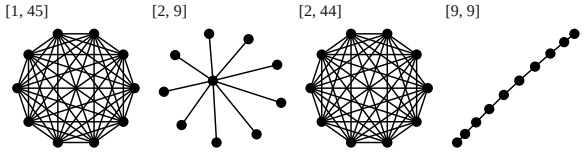


Figure 19: Graphs which are vertices of $\mathbb{P}_{10}$

Graphs which are polyhedral vertices of $\mathbb{P}_{10}$ are represented in Figure 19.

8.3 Pareto points

Pareto points related to facet $5D + m \leq 54$

Coordinates	# of graphs
$[2, 44]$	1
$[3, 36]$	4
$[4, 29]$	6
$[5, 23]$	7
$[6, 18]$	7
$[7, 14]$	6
$[8, 11]$	4
$[9, 9]$	1

Pareto points of facet $5D + m \leq 54$ are represented in Figure 20. All these Pareto points respect the following equality :

$$m = \frac{1}{2}D^2 - \frac{21}{2}D + 63 \quad (33)$$

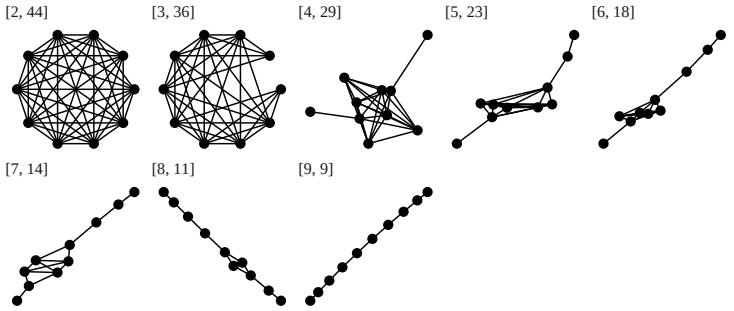


Figure 20: Pareto points of facet $5D + m \leq 54$

9 Automated conjectures

GPH Conjecture 1 *For all graphs with $n \geq 4$ nodes such that C_1 is respected, the number of vertices of $\mathbb{P}_n$ is*

4.

GPH Conjecture 2 *For all graphs with $n \geq 4$ nodes such that C_1 is respected, the number of facets of $\mathbb{P}_n$ is*

4.

GPH Conjecture 3 *For all graphs with $n \geq 4$ nodes such that C_1 is respected, the following inequality is a facet of $\mathbb{P}_n$*

$$D + m \leq \frac{1}{2}n^2 - \frac{1}{2}n + 1 \tag{34}$$

GPH Conjecture 4 *For all graphs with $n \geq 4$ nodes such that C_1 is respected, the following inequality is a facet of $\mathbb{P}_n$*

$$\frac{1}{2}nD + m \leq \frac{1}{2}n^2 + \frac{1}{2}n - 1 \tag{35}$$

GPH Conjecture 5 *For all graphs with $n \geq 4$ nodes such that C_1 is respected, the following inequality is a facet of $\mathbb{P}_n$*

$$m \geq n - 1 \tag{36}$$

GPH Conjecture 6 *For all graphs with $n \geq 4$ nodes such that C_1 is respected, the following inequality is a facet of $\mathbb{P}_n$*

$$\left(\frac{1}{2}n^2 - \frac{3}{2}n + 1\right) D + m \geq n^2 - 2n + 1 \tag{37}$$

The following conjecture was derived from the generalization of the pareto points.

GPH Conjecture 7 *For all graphs with $n \geq 6$ nodes such that C_1 is respected,*

$$m \leq \frac{1}{2}D^2 + \left(-n - \frac{1}{2}\right) D + \left(\frac{1}{2}n^2 + \frac{3}{2}n - 2\right) \tag{38}$$

10 CPU times

Here are the CPU times used to solve this problem. Note that, if some data have been computed on different computers, the following times are inconsistents. To get consistent CPU times for a given problem, all computation should be made on the same computer.

There are two types of CPU times in the following tables, depending on *when* the data were computed :

1. *Precomp.* (precomputed data). CPU times to compute data which were already computed to solve previously a similar problem (with, for instance, the same class of graphs or the same invariant). The CPU times are thus given for information but were not spended to solve this specific problem.
2. *Comp. now* (data computed now). CPU times to compute data which are not (yet) precomputed.

10.1 Generate graphs

Here are the times used to generate (with *geng*) all non-isomorphic graphs with n nodes.

$n =$	Precomp.	Comp. now
4	0.00 sec	0.00 sec
5	0.00 sec	0.00 sec
6	0.00 sec	0.00 sec
7	0.00 sec	0.00 sec
8	0.02 sec	0.00 sec
9	0.45 sec	0.00 sec
10	21.38 sec	0.00 sec

10.2 Compute invariants

Here are the times to compute the invariants.

$n =$	*Connected*	*D*	*m*	Total Precomp.	Total Comp. now
4	0.00 sec	0.00 sec	0.00 sec	0.00 sec	0.00 sec
5	0.00 sec	0.00 sec	0.00 sec	0.00 sec	0.00 sec
6	0.00 sec	0.00 sec	0.00 sec	0.00 sec	0.00 sec
7	0.00 sec	0.01 sec	0.00 sec	0.01 sec	0.00 sec
8	0.04 sec	0.07 sec	0.02 sec	0.13 sec	0.00 sec
9	1.07 sec	2.10 sec	0.55 sec	3.72 sec	0.00 sec
10	56.99 sec	1 min 58.98 sec	32.02 sec	3 min 27.99 sec	0.00 sec

10.3 Compute polyhedra

The computation of a polyhedron requires three steps :

1. *Filter points*. Get a set S of unique points. In this problem, a *point* represent one or more graphs with n nodes such that C_1 is respected. This step is important (and often slow) as the number of resulting points in S is often much lesser than the number of graphs with n nodes. Some optional information (e.g. statistics about coordinates) are also collected during the filtering.
2. *Compute facets*. Compute the facets of the polytope defined by $conv(S)$.
3. *Compute vertices*. Compute the vertices of the polytope defined by $conv(S)$.

$n =$	Filter	Cp. Facets	Cp. Vert.	Total Precp.	Tot. Cp. now
4	0.00 sec	0.00 sec	0.00 sec	0.00 sec	0.00 sec
5	0.00 sec	0.00 sec	0.00 sec	0.00 sec	0.00 sec
6	0.00 sec	0.00 sec	0.01 sec	0.00 sec	0.01 sec
7	0.02 sec	0.00 sec	0.01 sec	0.00 sec	0.03 sec
8	0.20 sec	0.00 sec	0.02 sec	0.00 sec	0.22 sec
9	4.63 sec	0.00 sec	0.06 sec	0.00 sec	4.69 sec
10	3 min 25.85 sec	0.00 sec	0.10 sec	0.00 sec	3 min 25.95 sec

Diameter and Number of Edges 18

10.4 Global CPU times

Compute polyhedra. The total times to compute completely the polytopes are mentionned here.

$n =$	Precomp.	Comp. now	Total
4	0.00 sec	0.00 sec	0.00 sec
5	0.00 sec	0.00 sec	0.00 sec
6	0.00 sec	0.01 sec	0.01 sec
7	0.01 sec	0.03 sec	0.04 sec
8	0.15 sec	0.22 sec	0.37 sec
9	4.17 sec	4.69 sec	8.86 sec
10	3 min 49.37 sec	3 min 25.95 sec	7 min 15.32 sec

Search conjectures. The CPU time of the automated conjecture's search is 0.01 sec.
Total. The total CPU time used to compute all data used in this report is 7 min 24.61 sec (52.56 % was precomputed, so the real CPU time for this run was 3 min 30.91 sec).

B.2 GraPHedron's report for Problem 6.2

The following code is an input file to solve Problem 6.2 with GraPHedron.

```
<COORDINATES>
 StableMax; NumEdges
</COORDINATES>

<CONDITIONS> Connected </CONDITIONS>

<REPORT>
 TITLE "Stability number and Number of edges of Connected graphs"
 COLOR Degree
</REPORT>

<ORDER>
  NMIN 4
  NMAX 10
</ORDER>

<COMPUTATION>
  STATISTICS
  DISTRIBUTION
  REFS 5
</COMPUTATION>
```

A copy of the report created by GraPHedron automatically is presented on the following pages.

Stability number and Number of edges of Connected graphs

GraPHedron 0.6.1 - Automatic Report

Thu Feb 9 13:24:56 2006

Contents

1 Problem statement

Definition 1 *A graph G is* connected *is every pair of nodes in G are joined by a path.*

Definition 2 *The* number of edges *m of a graph $G = (V, E)$, or its* size*, is the cardinal of E.*

Definition 3 *The* stability number *of a graph G is the maximum cardinality of an independent set in G*

The class of graphs generated are all non-isomorphic *graphs* with n nodes. The dimension of the problem is 2 and the coordinates are:

$$\begin{aligned} x_1 &= \alpha, \\ x_2 &= m. \end{aligned}$$

where m is the *Number of edges* and α is the *Stability Number*.

The graphs considered respect the following condition(s):

$$C_1 : \quad \textit{Connected}.$$

where *Connected* is true if the graphs are connected. In the following, a polyhedron $\mathbb{P}_n$ is thus defined by

$$\begin{aligned} \mathbb{P}_n = \quad & \text{conv}\{(x_1, x_2) \mid \exists\, G \in \mathcal{C}_n, \\ & x_1 = \alpha(G),\ x_2 = m(G)\}, \end{aligned}$$

where $\mathcal{C}_n$ is the set of graphs with n nodes such that C_1 is respected.

The polyhedra computed are those for which $n = 4$ to 10. A maximum of 5 graphs are drawn for one vertex of the polyhedron.

2 Statistics for coordinates

2.1 Coordinate x_1

The following statistics about $x_1 = \alpha$ are represented in Figure 1. Figure 2 shows the distributions of x_1 (for each computed n).

n	Min	Max	Mean	Std. Dev.
4	1	3	2	0.57735
5	1	4	2.42857	0.659829
6	1	5	2.8125	0.675215
7	1	6	3.1653	0.661886
8	1	7	3.45057	0.654493
9	1	8	3.69208	0.649242
10	1	9	3.90161	0.636624

GPH Conjecture 1 *For all graphs with $n \geq 4$ nodes such that C_1 is respected,*

$$1 \leq \alpha.$$

GPH Conjecture 2 *For all graphs with $n \geq 4$ nodes such that C_1 is respected,*

$$\alpha \leq n - 1.$$

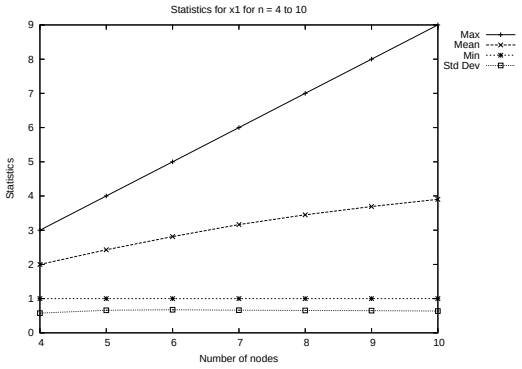


Figure 1: Statistics for x_1

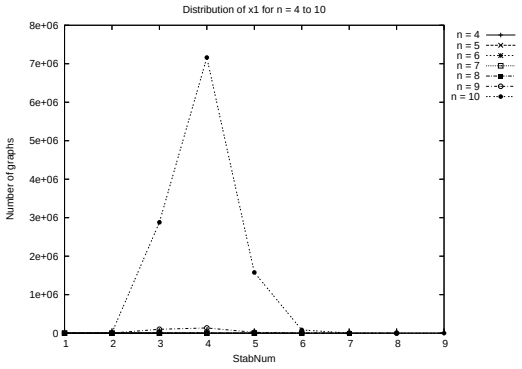


Figure 2: Distribution of values for x_1

2.2 Coordinate x_2

The following statistics about $x_2 = m$ are represented in Figure 3. Figure 4 shows the distributions of x_2 (for each computed n).

n	Min	Max	Mean	Std. Dev.
4	3	6	4.16667	1.06719
5	4	10	6.19048	1.59221
6	5	15	8.49107	2.0485
7	6	21	11.1981	2.44362
8	7	28	14.4122	2.79586
9	8	36	18.2193	3.11345
10	9	45	22.6145	3.4266

GPH Conjecture 3 *For all graphs with $n \geq 4$ nodes such that C_1 is respected,*

$$n - 1 \leq m.$$

GPH Conjecture 4 *For all graphs with $n \geq 4$ nodes such that C_1 is respected,*

$$m \leq \frac{1}{2}n^2 - \frac{1}{2}n.$$

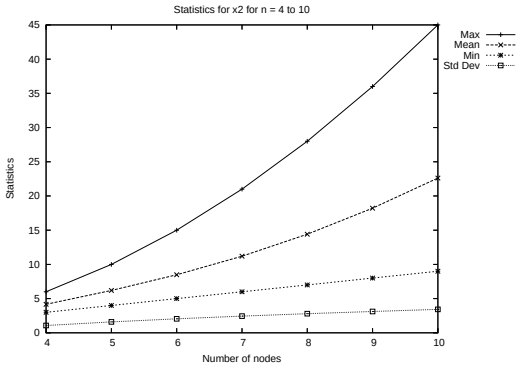


Figure 3: Statistics for x_2

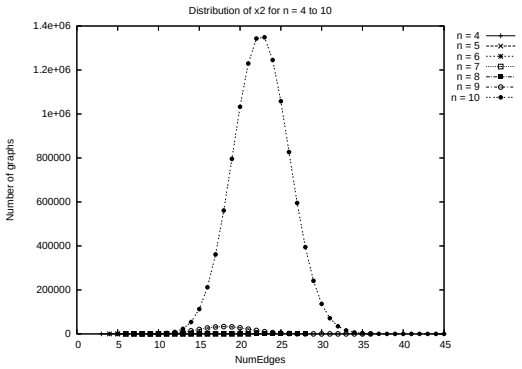


Figure 4: Distribution of values for x_2

3 Polyhedron $n = 4$

Information about polyhedron $\mathbb{P}_4$ of graphs with $n = 4$:

# of graphs generated	11
# of gr. satisf. cond.	6
# of unique coord.	5
# of vertices of $\mathbb{P}_4$	4
# of facets of $\mathbb{P}_4$	4
Data precomputed (CPU time)	0.00 sec
Data computed now (CPU time)	0.00 sec
Total CPU time to create $\mathbb{P}_4$	0.00 sec

The polyhedron $\mathbb{P}_4$ is represented in Figure 5.

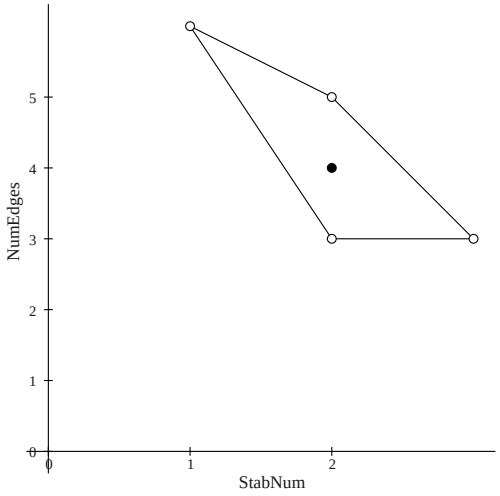


Figure 5: Representation of $\mathbb{P}_4$

3.1 Facets

$$\alpha + m \leq 7 \quad (1)$$
$$2\alpha + m \leq 9 \quad (2)$$
$$m \geq 3 \quad (3)$$
$$3\alpha + m \geq 9 \quad (4)$$

Facets	Vertices in facet
(1)	[1, 6] [2, 5]
(2)	[2, 5] [3, 3]
(3)	[2, 3] [3, 3]
(4)	[1, 6] [2, 3]

3.2 Vertices

Coordinates	# of graphs	Facets
$[1, 6]$	1	(1) (4)
$[2, 3]$	1	(3) (4)
$[2, 5]$	1	(1) (2)
$[3, 3]$	1	(2) (3)

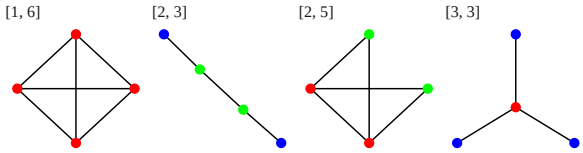


Figure 6: Graphs which are vertices of $\mathbb{P}_4$

Graphs which are polyhedral vertices of $\mathbb{P}_4$ are represented in Figure 6.

3.3 Pareto points

No pareto points found for this polytope.

4 Polyhedron $n = 5$

Information about polyhedron $\mathbb{P}_5$ of graphs with $n = 5$:

# of graphs generated	34
# of gr. satisf. cond.	21
# of unique coord.	11
# of vertices of $\mathbb{P}_5$	6
# of facets of $\mathbb{P}_5$	6
Data precomputed (CPU time)	0.00 sec
Data computed now (CPU time)	0.01 sec
Total CPU time to create $\mathbb{P}_5$	0.01 sec

The polyhedron $\mathbb{P}_5$ is represented in Figure 7.

4.1 Facets

$$\alpha + m \leq 11 \quad (5)$$
$$2\alpha + m \leq 13 \quad (6)$$
$$3\alpha + m \leq 16 \quad (7)$$
$$m \geq 4 \quad (8)$$
$$\alpha + m \geq 7 \quad (9)$$
$$5\alpha + m \geq 15 \quad (10)$$

Facets	Vertices in facet
(5)	$[1, 10]$ $[2, 9]$
(6)	$[2, 9]$ $[3, 7]$
(7)	$[3, 7]$ $[4, 4]$
(8)	$[3, 4]$ $[4, 4]$
(9)	$[2, 5]$ $[3, 4]$
(10)	$[1, 10]$ $[2, 5]$

Stability number and Number of edges of Connected graphs 8

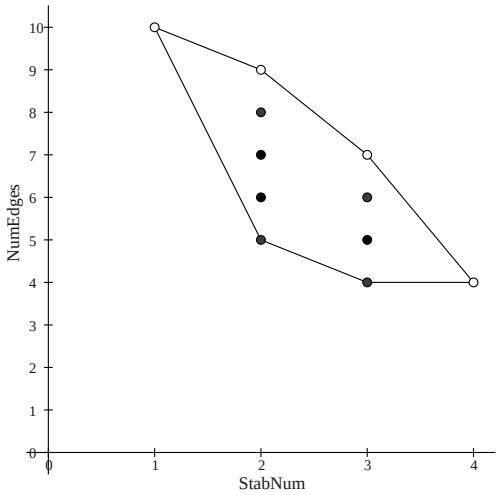


Figure 7: Representation of $\mathbb{P}_5$

4.2 Vertices

Coordinates	# of graphs	Facets
$[1, 10]$	1	(5) (10)
$[2, 5]$	2	(9) (10)
$[2, 9]$	1	(5) (6)
$[3, 4]$	2	(8) (9)
$[3, 7]$	1	(6) (7)
$[4, 4]$	1	(7) (8)

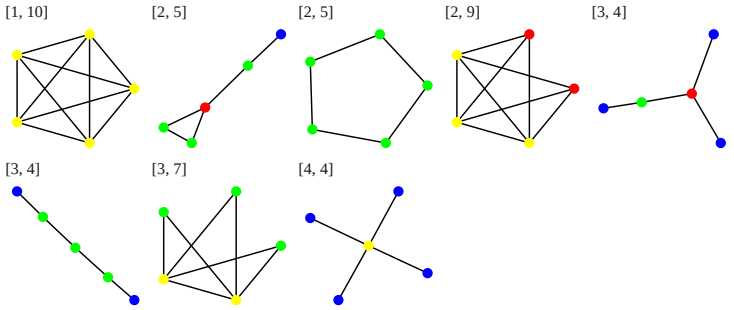


Figure 8: Graphs which are vertices of $\mathbb{P}_5$

Graphs which are polyhedral vertices of $\mathbb{P}_5$ are represented in Figure 8.

4.3 Pareto points

No pareto points found for this polytope.

5 Polyhedron $n = 6$

Information about polyhedron $\mathbb{P}_6$ of graphs with $n = 6$:

# of graphs generated	156
# of gr. satisf. cond.	112
# of unique coord.	23
# of vertices of $\mathbb{P}_6$	7
# of facets of $\mathbb{P}_6$	7
Data precomputed (CPU time)	0.00 sec
Data computed now (CPU time)	0.01 sec
Total CPU time to create $\mathbb{P}_6$	0.01 sec

The polyhedron $\mathbb{P}_6$ is represented in Figure 9.

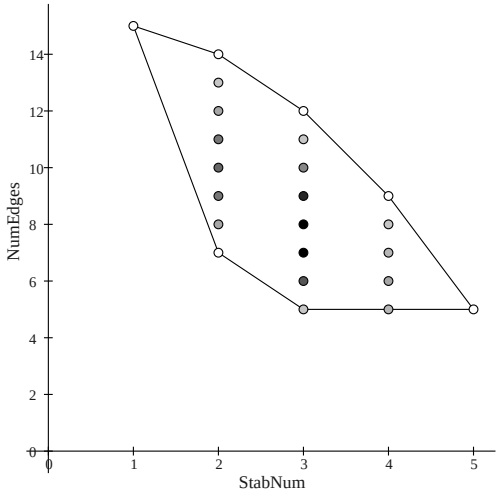


Figure 9: Representation of $\mathbb{P}_6$

5.1 Facets

$$\alpha + m \leq 16 \quad (11)$$
$$2\alpha + m \leq 18 \quad (12)$$
$$3\alpha + m \leq 21 \quad (13)$$
$$4\alpha + m \leq 25 \quad (14)$$
$$m \geq 5 \quad (15)$$
$$2\alpha + m \geq 11 \quad (16)$$
$$8\alpha + m \geq 23 \quad (17)$$

Stability number and Number of edges of Connected graphs 10

Facets	Vertices in facet
(11)	[1, 15] [2, 14]
(12)	[2, 14] [3, 12]
(13)	[3, 12] [4, 9]
(14)	[4, 9] [5, 5]
(15)	[3, 5] [5, 5]
(16)	[2, 7] [3, 5]
(17)	[1, 15] [2, 7]

5.2 Vertices

Coordinates	# of graphs	Facets
[1, 15]	1	(11) (17)
[2, 7]	1	(16) (17)
[2, 14]	1	(11) (12)
[3, 5]	2	(15) (16)
[3, 12]	1	(12) (13)
[4, 9]	1	(13) (14)
[5, 5]	1	(14) (15)

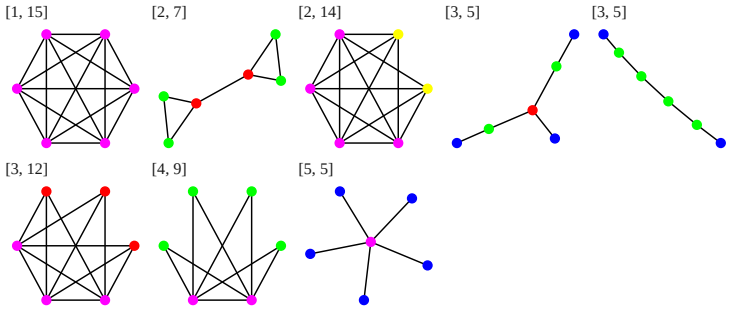


Figure 10: Graphs which are vertices of $\mathbb{P}_6$

Graphs which are polyhedral vertices of $\mathbb{P}_6$ are represented in Figure 10.

5.3 Pareto points

No pareto points found for this polytope.

6 Polyhedron $n = 7$

Information about polyhedron $\mathbb{P}_7$ of graphs with $n = 7$:

# of graphs generated	1044
# of gr. satisf. cond.	853
# of unique coord.	41
# of vertices of $\mathbb{P}_7$	9
# of facets of $\mathbb{P}_7$	9
Data precomputed (CPU time)	0.00 sec
Data computed now (CPU time)	0.04 sec
Total CPU time to create $\mathbb{P}_7$	0.04 sec

The polyhedron $\mathbb{P}_7$ is represented in Figure 11.

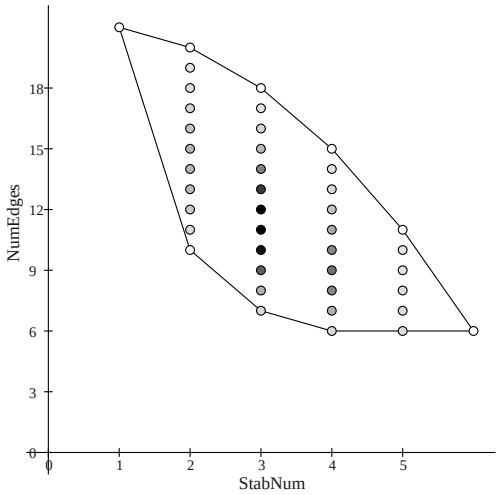


Figure 11: Representation of $\mathbb{P}_7$

6.1 Facets

$$\alpha + m \leq 22 \quad (18)$$
$$2\alpha + m \leq 24 \quad (19)$$
$$3\alpha + m \leq 27 \quad (20)$$
$$4\alpha + m \leq 31 \quad (21)$$
$$5\alpha + m \leq 36 \quad (22)$$
$$m \geq 6 \quad (23)$$
$$\alpha + m \geq 10 \quad (24)$$
$$3\alpha + m \geq 16 \quad (25)$$
$$11\alpha + m \geq 32 \quad (26)$$

Facets	Vertices in facet
(18)	$[1, 21]$ $[2, 20]$
(19)	$[2, 20]$ $[3, 18]$
(20)	$[3, 18]$ $[4, 15]$
(21)	$[4, 15]$ $[5, 11]$
(22)	$[5, 11]$ $[6, 6]$
(23)	$[4, 6]$ $[6, 6]$
(24)	$[3, 7]$ $[4, 6]$
(25)	$[2, 10]$ $[3, 7]$
(26)	$[1, 21]$ $[2, 10]$

Stability number and Number of edges of Connected graphs 12

6.2 Vertices

Coordinates	# of graphs	Facets
[1, 21]	1	(18) (26)
[2, 10]	1	(25) (26)
[2, 20]	1	(18) (19)
[3, 7]	6	(24) (25)
[3, 18]	1	(19) (20)
[4, 6]	6	(23) (24)
[4, 15]	1	(20) (21)
[5, 11]	1	(21) (22)
[6, 6]	1	(22) (23)

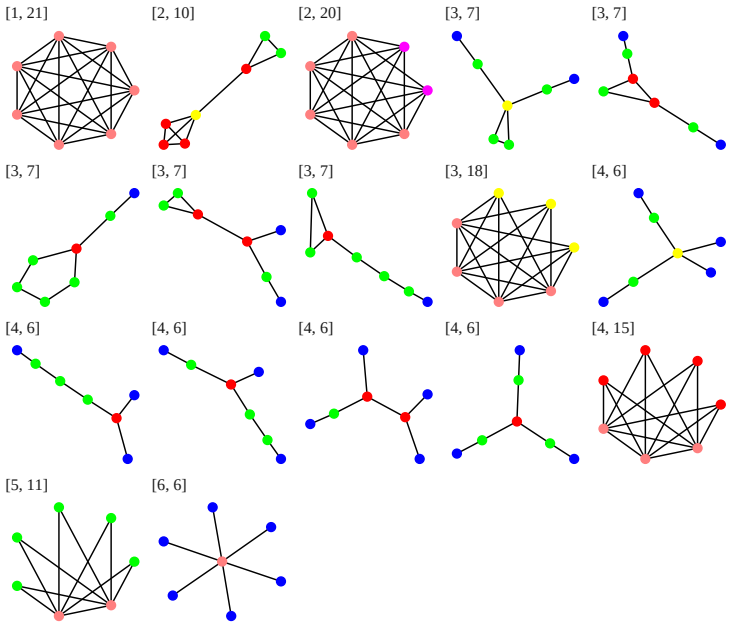


Figure 12: Graphs which are vertices of $\mathbb{P}_7$

Graphs which are polyhedral vertices of $\mathbb{P}_7$ are represented in Figure 12.

6.3 Pareto points

No pareto points found for this polytope.

7 Polyhedron $n = 8$

Information about polyhedron $\mathbb{P}_8$ of graphs with $n = 8$:

# of graphs generated	12346
# of gr. satisf. cond.	11117
# of unique coord.	69
# of vertices of $\mathbb{P}_8$	10
# of facets of $\mathbb{P}_8$	10
Data precomputed (CPU time)	0.15 sec
Data computed now (CPU time)	0.24 sec
Total CPU time to create $\mathbb{P}_8$	0.39 sec

The polyhedron $\mathbb{P}_8$ is represented in Figure 13.

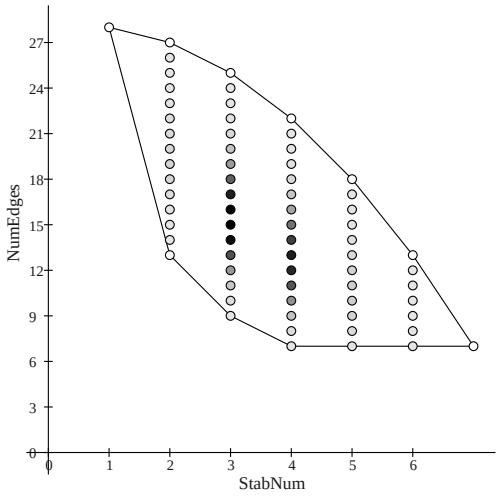


Figure 13: Representation of $\mathbb{P}_8$

7.1 Facets

$$\alpha + m \leq 29 \quad (27)$$
$$2\alpha + m \leq 31 \quad (28)$$
$$3\alpha + m \leq 34 \quad (29)$$
$$4\alpha + m \leq 38 \quad (30)$$
$$5\alpha + m \leq 43 \quad (31)$$
$$6\alpha + m \leq 49 \quad (32)$$
$$m \geq 7 \quad (33)$$
$$2\alpha + m \geq 15 \quad (34)$$
$$4\alpha + m \geq 21 \quad (35)$$
$$15\alpha + m \geq 43 \quad (36)$$

Stability number and Number of edges of Connected graphs 14

Facets	Vertices in facet
(27)	[1, 28] [2, 27]
(28)	[2, 27] [3, 25]
(29)	[3, 25] [4, 22]
(30)	[4, 22] [5, 18]
(31)	[5, 18] [6, 13]
(32)	[6, 13] [7, 7]
(33)	[4, 7] [7, 7]
(34)	[3, 9] [4, 7]
(35)	[2, 13] [3, 9]
(36)	[1, 28] [2, 13]

7.2 Vertices

Coordinates	# of graphs	Facets
[1, 28]	1	(27) (36)
[2, 13]	1	(35) (36)
[2, 27]	1	(27) (28)
[3, 9]	5	(34) (35)
[3, 25]	1	(28) (29)
[4, 7]	5	(33) (34)
[4, 22]	1	(29) (30)
[5, 18]	1	(30) (31)
[6, 13]	1	(31) (32)
[7, 7]	1	(32) (33)

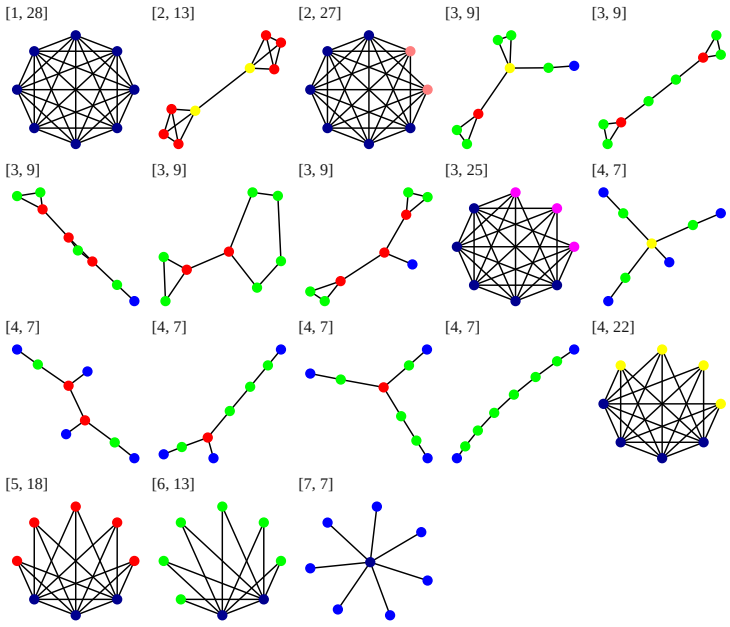


Figure 14: Graphs which are vertices of $\mathbb{P}_8$

Graphs which are polyhedral vertices of $\mathbb{P}_8$ are represented in Figure 14.

7.3 Pareto points

No pareto points found for this polytope.

8 Polyhedron $n = 9$

Information about polyhedron $\mathbb{P}_9$ of graphs with $n = 9$:

# of graphs generated	274668
# of gr. satisf. cond.	261080
# of unique coord.	107
# of vertices of $\mathbb{P}_9$	12
# of facets of $\mathbb{P}_9$	12
Data precomputed (CPU time)	3.88 sec
Data computed now (CPU time)	4.55 sec
Total CPU time to create $\mathbb{P}_9$	8.43 sec

The polyhedron $\mathbb{P}_9$ is represented in Figure 15.

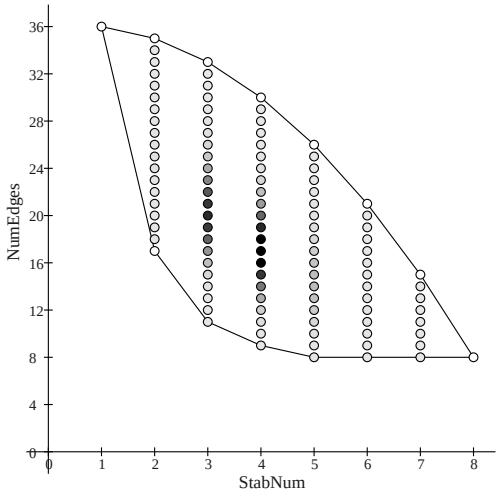


Figure 15: Representation of $\mathbb{P}_9$

Stability number and Number of edges of Connected graphs 16

8.1 Facets

$$\begin{aligned}
\alpha + m &\leq 37 && (37)\\
2\alpha + m &\leq 39 && (38)\\
3\alpha + m &\leq 42 && (39)\\
4\alpha + m &\leq 46 && (40)\\
5\alpha + m &\leq 51 && (41)\\
6\alpha + m &\leq 57 && (42)\\
7\alpha + m &\leq 64 && (43)\\
m &\geq 8 && (44)\\
\alpha + m &\geq 13 && (45)\\
2\alpha + m &\geq 17 && (46)\\
6\alpha + m &\geq 29 && (47)\\
19\alpha + m &\geq 55 && (48)
\end{aligned}$$

Facets	Vertices in facet
(37)	[1, 36] [2, 35]
(38)	[2, 35] [3, 33]
(39)	[3, 33] [4, 30]
(40)	[4, 30] [5, 26]
(41)	[5, 26] [6, 21]
(42)	[6, 21] [7, 15]
(43)	[7, 15] [8, 8]
(44)	[5, 8] [8, 8]
(45)	[4, 9] [5, 8]
(46)	[3, 11] [4, 9]
(47)	[2, 17] [3, 11]
(48)	[1, 36] [2, 17]

8.2 Vertices

Coordinates	# of graphs	Facets
[1, 36]	1	(37) (48)
[2, 17]	1	(47) (48)
[2, 35]	1	(37) (38)
[3, 11]	2	(46) (47)
[3, 33]	1	(38) (39)
[4, 9]	21	(45) (46)
[4, 30]	1	(39) (40)
[5, 8]	20	(44) (45)
[5, 26]	1	(40) (41)
[6, 21]	1	(41) (42)
[7, 15]	1	(42) (43)
[8, 8]	1	(43) (44)

Graphs which are polyhedral vertices of $\mathbb{P}_9$ are represented in Figure 16.

8.3 Pareto points

No pareto points found for this polytope.

9 Polyhedron $n = 10$

Information about polyhedron $\mathbb{P}_{10}$ of graphs with $n = 10$:

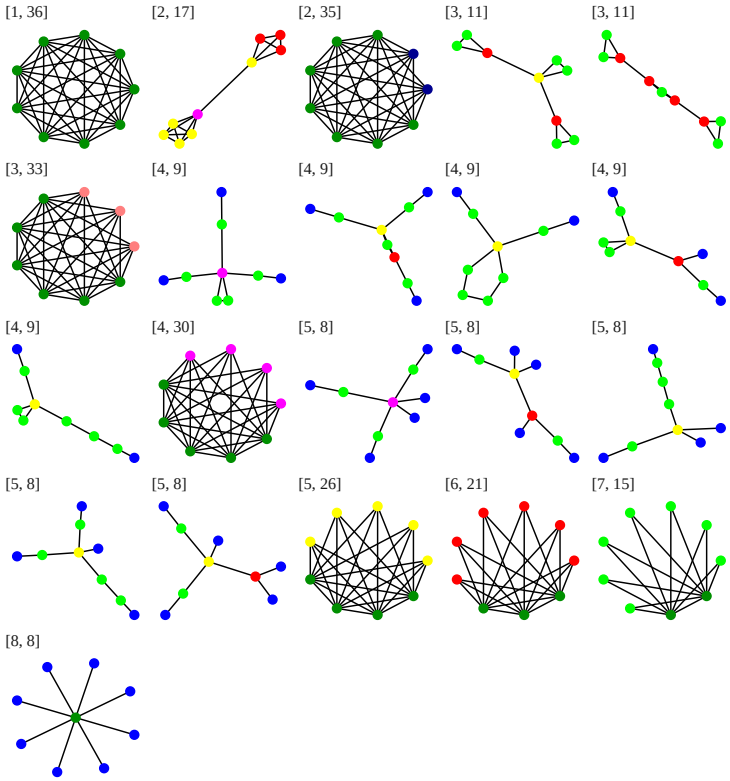


Figure 16: Graphs which are vertices of $\mathbb{P}_9$

# of graphs generated	12005168
# of gr. satisf. cond.	11716571
# of unique coord.	158
# of vertices of $\mathbb{P}_{10}$	13
# of facets of $\mathbb{P}_{10}$	13
Data precomputed (CPU time)	3 min 40.14 sec
Data computed now (CPU time)	3 min 22.91 sec
Total CPU time to create $\mathbb{P}_{10}$	7 min 3.05 sec

The polyhedron $\mathbb{P}_{10}$ is represented in Figure 17.

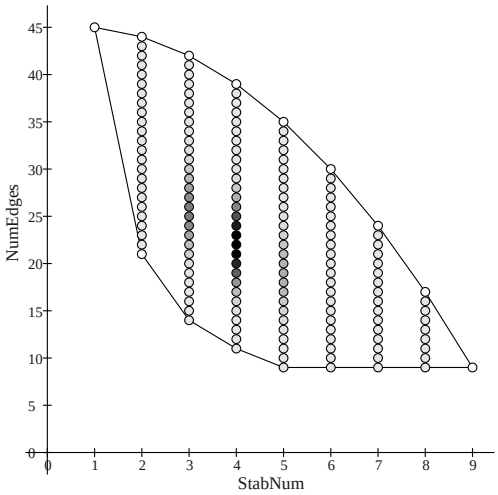


Figure 17: Representation of $\mathbb{P}_{10}$

9.1 Facets

$$\alpha + m \leq 46 \quad (49)$$
$$2\alpha + m \leq 48 \quad (50)$$
$$3\alpha + m \leq 51 \quad (51)$$
$$4\alpha + m \leq 55 \quad (52)$$
$$5\alpha + m \leq 60 \quad (53)$$
$$6\alpha + m \leq 66 \quad (54)$$
$$7\alpha + m \leq 73 \quad (55)$$
$$8\alpha + m \leq 81 \quad (56)$$
$$m \geq 9 \quad (57)$$
$$2\alpha + m \geq 19 \quad (58)$$
$$3\alpha + m \geq 23 \quad (59)$$
$$7\alpha + m \geq 35 \quad (60)$$
$$24\alpha + m \geq 69 \quad (61)$$

Facets	Vertices in facet
(49)	[1, 45] [2, 44]
(50)	[2, 44] [3, 42]
(51)	[3, 42] [4, 39]
(52)	[4, 39] [5, 35]
(53)	[5, 35] [6, 30]
(54)	[6, 30] [7, 24]
(55)	[7, 24] [8, 17]
(56)	[8, 17] [9, 9]
(57)	[5, 9] [9, 9]
(58)	[4, 11] [5, 9]
(59)	[3, 14] [4, 11]
(60)	[2, 21] [3, 14]
(61)	[1, 45] [2, 21]

9.2 Vertices

Coordinates	# of graphs	Facets
[1, 45]	1	(49) (61)
[2, 21]	1	(60) (61)
[2, 44]	1	(49) (50)
[3, 14]	4	(59) (60)
[3, 42]	1	(50) (51)
[4, 11]	30	(58) (59)
[4, 39]	1	(51) (52)
[5, 9]	15	(57) (58)
[5, 35]	1	(52) (53)
[6, 30]	1	(53) (54)
[7, 24]	1	(54) (55)
[8, 17]	1	(55) (56)
[9, 9]	1	(56) (57)

Graphs which are polyhedral vertices of $\mathbb{P}_{10}$ are represented in Figure 18.

9.3 Pareto points

No pareto points found for this polytope.

10 Automated conjectures

GPH Conjecture 5 *For all graphs with $n \geq 4$ nodes such that C_1 is respected, the number of vertices of $\mathbb{P}_n$ is*

$$\frac{3}{2}n - 2 \text{ if } n(mod\ 2) = 0,$$
$$\frac{3}{2}n - \frac{3}{2} \text{ if } n(mod\ 2) = 1.$$

GPH Conjecture 6 *For all graphs with $n \geq 4$ nodes such that C_1 is respected, the number of facets of $\mathbb{P}_n$ is*

$$\frac{3}{2}n - 2 \text{ if } n(mod\ 2) = 0,$$
$$\frac{3}{2}n - \frac{3}{2} \text{ if } n(mod\ 2) = 1.$$

Stability number and Number of edges of Connected graphs 20

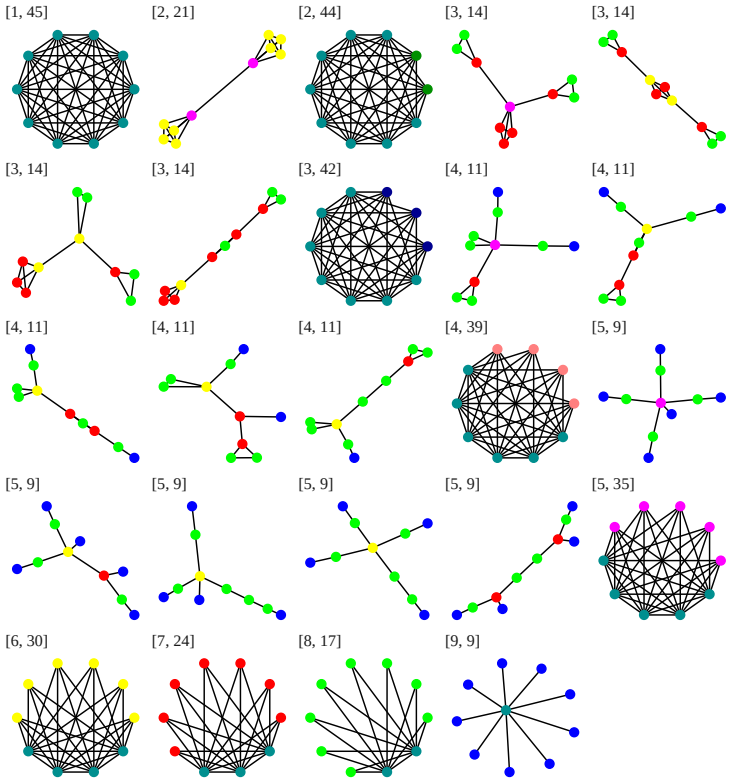


Figure 18: Graphs which are vertices of $\mathbb{P}_{10}$

11 CPU times

Here are the CPU times used to solve this problem. Note that, if some data have been computed on different computers, the following times are inconsistents. To get consistent CPU times for a given problem, all computation should be made on the same computer.

There are two types of CPU times in the following tables, depending on *when* the data were computed :

1. *Precomp.* (precomputed data). CPU times to compute data which were already computed to solve previously a similar problem (with, for instance, the same class of graphs or the same invariant). The CPU times are thus given for information but were not spended to solve this specific problem.
2. *Comp. now* (data computed now). CPU times to compute data which are not (yet) precomputed.

11.1 Generate graphs

Here are the times used to generate (with *geng*) all non-isomorphic graphs with n nodes.

$n =$	Precomp.	Comp. now
4	0.00 sec	0.00 sec
5	0.00 sec	0.00 sec
6	0.00 sec	0.00 sec
7	0.00 sec	0.00 sec
8	0.02 sec	0.00 sec
9	0.45 sec	0.00 sec
10	21.38 sec	0.00 sec

11.2 Compute invariants

Here are the times to compute the invariants.

$n =$	*Connected*	m	α	Total Precomp.	Total Comp. now
4	0.00 sec	0.00 sec	0.00 sec	0.00 sec	0.00 sec
5	0.00 sec	0.00 sec	0.00 sec	0.00 sec	0.00 sec
6	0.00 sec	0.00 sec	0.00 sec	0.00 sec	0.00 sec
7	0.00 sec	0.00 sec	0.00 sec	0.00 sec	0.00 sec
8	0.04 sec	0.02 sec	0.07 sec	0.13 sec	0.00 sec
9	1.07 sec	0.55 sec	1.81 sec	3.43 sec	0.00 sec
10	56.99 sec	32.02 sec	1 min 49.75 sec	3 min 18.76 sec	0.00 sec

11.3 Compute polyhedra

The computation of a polyhedron requires three steps :

1. *Filter points*. Get a set S of unique points. In this problem, a *point* represent one or more graphs with n nodes such that C_1 is respected. This step is important (and often slow) as the number of resulting points in S is often much lesser than the number of graphs with n nodes. Some optional information (e.g. statistics about coordinates) are also collected during the filtering.
2. *Compute facets*. Compute the facets of the polytope defined by $conv(S)$.
3. *Compute vertices*. Compute the vertices of the polytope defined by $conv(S)$.

$n =$	Filter	Cp. Facets	Cp. Vert.	Total Precp.	Tot. Cp. now
4	0.00 sec	0.00 sec	0.00 sec	0.00 sec	0.00 sec
5	0.00 sec	0.00 sec	0.01 sec	0.00 sec	0.01 sec
6	0.00 sec	0.00 sec	0.01 sec	0.00 sec	0.01 sec
7	0.01 sec	0.00 sec	0.03 sec	0.00 sec	0.04 sec
8	0.19 sec	0.00 sec	0.05 sec	0.00 sec	0.24 sec
9	4.45 sec	0.00 sec	0.10 sec	0.00 sec	4.55 sec
10	3 min 22.72 sec	0.00 sec	0.19 sec	0.00 sec	3 min 22.91 sec

Stability number and Number of edges of Connected graphs 22

11.4 Global CPU times

Compute polyhedra. The total times to compute completely the polytopes are mentionned here.

$n =$	Precomp.	Comp. now	Total
4	0.00 sec	0.00 sec	0.00 sec
5	0.00 sec	0.01 sec	0.01 sec
6	0.00 sec	0.01 sec	0.01 sec
7	0.00 sec	0.04 sec	0.04 sec
8	0.15 sec	0.24 sec	0.39 sec
9	3.88 sec	4.55 sec	8.43 sec
10	3 min 40.14 sec	3 min 22.91 sec	7 min 3.05 sec

Search conjectures. The CPU time of the automated conjecture's search is 0.00 sec.
Total. The total CPU time used to compute all data used in this report is 7 min 11.93 sec (51.90 % was precomputed, so the real CPU time for this run was 3 min 27.76 sec).

B.3 Operators available in GraPHedron

Tables B.1 – B.2 list the operators currently available in GraPHedron where x and y are arithmetic expressions, p an integer and a and b are boolean expressions. Operator precedence determines the

order in which the terms of an expression will be evaluated. Operators with the highest precedence will be evaluated before operators with lower precedence. One can use parentheses to break out precedence. There are three types of operators : arithmetic operators can then be used as arithmetic expressions and relational and logical operators can be used as boolean expressions.

Table B.1: Arithmetic operators

Description	**Accepted syntax**	**Arity**	**Prec.**
power	`x^p`	binary	11
unary minus	`-x`	unary	10
multiplication	`x*y`	binary	9
division	`x/y`	binary	9
modulo	`x%p` or `x mod y`	binary	9
plus	`x+y`	binary	8
minus	`x-y`	binary	8
maximum	`MAX(x,y,...)`	*k*-ary	7
minimum	`MIN(x,y,...)`	*k*-ary	7
mean	`MEAN(x,y,...)`	*k*-ary	7
variance	`VAR(x,y,...)`	*k*-ary	7
floor	`FLOOR(x)`	unary	7
ceil	`CEIL(x)`	unary	7
round	`ROUND(x)`	unary	7

Table B.2: Relational operators

Description	**Accepted syntax**	**Arity**	**Prec.**
less	`x < y`	binary	6
greater	`x > y`	binary	6
less or equal	`x <= y`	binary	6
greater or equal	`x >= y`	binary	6
equal	`a = b` or `a == b`	binary	5
not equal	`a != b` or `a <> b`	binary	5

Table B.3: Logical operators

Description	**Accepted syntax**	**Arity**	**Prec.**
not	`NOT a` or `!a`	unary	4
and	`a AND b` or `a && b`	binary	3
exclusive or	`a XOR b`	binary	2
or	`a OR b` or `a || b`	binary	1

B.4 Invariants available in GraPHedron

When designing GraPHedron, we wanted to keep the code and the documentation about invariants in the same place. This is the detailed information about invariants implemented in GraPHedron as produced by the program itself.

Remark B.1. The notations used in this list can differ from the notation used in the dissertation. It comes from the fact that the following notations are those used by GraPHedron to produce the reports, and is not related to the text of the dissertation.

List of invariants available in GraPHedron
Date of creation : Mon May 29 23:12:43 2006
Version of GraPHedron : 0.6.1
Number of invariants: 59

Here is the list of invariants currently available in GraPHedron.

Note about the complexities : The following worst case complexities are given when a graph G is already constructed, and following the GraPHedron's representation of a graph (which keeps n, m and the degrees as instance fields). For instance, computing the average degree $\bar{d}$ of a graph G is made in constant time as

$$\bar{d} = \sum_{v_i \in V(G)} d_{v_i}/n = \frac{2m}{n}$$

where d_{v_i} is the degree of node v_i, n is the number of nodes of G, and m its number of edges.

Note also that graphs are stored in an array representing the upper triangle of the adjacency matrix. Checking the presence of an edge is thus in constant time. However, a Depth First Search has a complexity $O(n^2)$ (with adjacency lists, it will be in $O(n+m)$).

Note about the conditions : Some invariants require conditions to be properly computed. For instance, a graph G has to be connected to have a finite diameter. If such a condition is not present in a problem definition, GraPHedron will automatically add it.

1. **Keyword:** `Acyclic`
   Name: *Acyclic*
   Type: Boolean
   Definition: A graph is said to be *acyclic* if it contains no cycle.
   Notation: *Acyclic*
   Complexity: $O(n^2)$

Note: We use a Breadth First Search.

2. **Keyword:** `AlgConnectivity`
   Name: *Algebraic connectivity*
   Type: Real
   Definition: The *algebraic connectivity* λ_{n-1}^{L} of a graph $G=(V,E)$ is the second smallest eigenvalue of its Laplacian matrix L, defined by

$$L = D - A$$

   where D is a diagonal matrix where $D_{i,i} = d_{v_i}$ (degree of node v_i) and A is the adjacency matrix of G.
   Notation: λ_{n-1}^{L}
   Complexity: $O(50n^2)$

3. **Keyword:** `AveDeg`
   Name: *Average degree*
   Type: Rational
   Definition: The *average degree* $\bar{d}(G)$ of a graph $G=(V,E)$ with n nodes is the (arithmetic) mean of the degrees of its nodes :

$$\bar{d}(G) = \frac{\sum_{v_i \in V}(d_{v_i})}{n}$$

   where d_{v_i} is the degree of node v_i.
   Notation: $\bar{d}$
   Complexity: $O(1)$

4. **Keyword:** `AveDist`
   Name: *Average distance*
   Type: Rational
   Definition: The *average distance* of a graph G is the mean of the distances between any pair of distinct nodes of G.
   Notation: $\bar{l}$
   Condition: *Connected*
   Complexity: $O(n^3)$
   Note: We use the Floyd-Marshall algorithm to compute all pairs shortest paths.

5. **Keyword:** `BandWidth`
   Name: *Bandwidth*
   Type: Integer

Definition: Let ν be a *numbering* on the nodes of a graph $G = (V,E)$, i.e. a function that assigns a distinct number from 1 to n to each of the nodes of G. The *bandwidth* of G is defined by

$$\max\{|\nu(v) - \nu(w)| : \{v,w \in E\}\},$$

between all possible numberings ν.
Notation: *bw*
Complexity: NP hard
Note: Implemented by S. Langerman.

6. **Keyword:** `Bipartite`
   Name: *Bipartite*
   Type: Boolean
   Definition: A graph $G = (V,E)$ is *bipartite* if it is possible to partition V into two subsets V_1 and V_2 such that every edge of E joins a vertex of V_1 to a vertex of V_2.
   Notation: *Bipartite*
   Complexity: $O(n^2)$
   Note: We use a Breadth First Search.

7. **Keyword:** `Chromatic`
   Name: *Chromatic Number*
   Type: Integer
   Definition: The (node) *chromatic number* of a graph G is the minimum number of different colors required to color the nodes of G such that two adjacent nodes have different colors.
   Notation: χ
   Complexity: NP-hard

8. **Keyword:** `CliqueMax`
   Name: *Clique Number*
   Type: Integer
   Definition: The *clique number* of a graph G is the maximum cardinality of a clique in G
   Notation: ω
   Complexity: NP hard

9. **Keyword:** `CompleteSplit`
   Name: *Complete Split graph*
   Type: Boolean
   Definition: A graph G is a *complete split graph* if it constructed from a clique on $n - \alpha$ nodes, an independent set on α nodes, and all possible edges between these two sets.

Notation: *CSG*
Complexity: $O(n)$

10. **Keyword:** `Connected`
Name: *Connected*
Type: Boolean
Definition: A graph G is *connected* is every pair of nodes in G are joined by a path.
Notation: *Connected*
Complexity: $O(n^2)$
Note: We use a Depth First Search.

11. **Keyword:** `ConvexCN`
Name: *Convex Crossing Number*
Type: Integer
Definition: A convex drawing of a graph G with n nodes is a drawing in which the nodes are placed on the corners of a convex n-gon in the plane and each edge is drawn using one straight line segment. The *convex crossing number* is the minimum number of crossings in any convex drawings of G.
Notation: *C*
Note: Implemented by M. Myldraz and S. Langerman.

12. **Keyword:** `CutWidth`
Name: *Cut width*
Type: Integer
Definition: Let ν be a *numbering* on the nodes of a graph $G = (V,E)$, i.e. a function that assigns a distinct number from 1 to n to each of the nodes of G. The *cutwidth* of G is defined by

$$\max_l \{\{v,w \in E\} : |\nu(v) \leq l < \nu(w)| :\},$$

between all possible numbering ν.
Notation: *ct*
Complexity: NP hard
Note: Implemented by S. Langerman.

13. **Keyword:** `CycleRank`
Name: *Cycle rank*
Type: Integer
Definition: The *cycle rank* of a graph G is the dimension of its cycle space. It is computed from

$$m - n + p,$$

where m is the number of edges, n the number of nodes and p the number of connected components of G.
Notation: $\mathcal{C}$
Complexity: $O(n^2)$
Note: We use a Depth First Search to compute p.

14. **Keyword:** `DNC`
Name: *Karpinski's dense node cover algorithm*
Type: Integer
Definition: The *Karpinski* value of a graph G is the worst case result of an approximation algorithm, from Karpinski and Zelikovsky for minimum vertex cover, applied on G
Notation: DVC
Note: Implemented by J. Cardinal. Use a brute-force algorithm.

15. **Keyword:** `DegreeGreedyMaxMatching`
Name: *Worst-case of a maximal matching computed by greedy method*
Type: Integer
Notation: $DGMM$
Note: Implemented by E. Levy. Use a brute-force algorithm.

16. **Keyword:** `Density`
Name: *Density of edges*
Type: Rational
Definition: The *density* of a graph G is

$$\frac{m}{m^*}$$

where m is the number of edges of G and $m^* = \frac{n(n-1)}{2}$ (number of edges of K_n).
Notation: d
Complexity: $O(1)$

17. **Keyword:** `Diam`
Name: *Diameter*
Type: Integer
Definition: The *diameter* D of a graph $G = (V,E)$ is the maximum distance between two nodes of G, i.e. the "longest shortest path" in G.
Notation: D
Condition: $Connected$
Complexity: $O(n^3)$

Note: We use the Floyd-Marshall algorithm to compute all pairs shortest paths.

18. **Keyword:** `EdgeConn`
Name: *Edge Connectivity*
Type: Integer
Definition: The *edge connectivity* of a graph G is the smallest number of edges of G whose deletion will cause G to not be connected.
Notation: λ
Complexity: Polynomial
Note: Implemented by K. Kontos.

19. **Keyword:** `Energy`
Name: *Energy*
Type: Real
Definition: The *energy* E of a graph $G = (V,E)$ is defined by

$$\sum_{i=1}^{n} |\lambda_i|$$

where λ_i is the i^{th} eigenvalue of the adjacency matrix of G.
Notation: E
Complexity: $O(50n^2)$

20. **Keyword:** `Eulerian`
Name: *Eulerian*
Type: Boolean
Definition: A graph is *Eulerian* if it contains a cycle passing by each edges of G exactly once.
Notation: $Eulerian$
Complexity: $O(n^2)$
Note: We check if G is connected and if all of its degrees are even.

21. **Keyword:** `Girth`
Name: *Girth*
Type: Integer
Definition: The *girth* of G is the length of the shortest cycle (if any) in G.
Notation: g
Condition: $!(Acyclic)$

22. **Keyword:** `Harmonic`
    Name: *Harmonic Mean of degrees*
    Type: Rational
    Definition: The *harmonic mean of degrees* of a graph G is the number h defined by
    $$\frac{1}{h} : \frac{1}{n}\sum_{i=1}^{n}\frac{1}{d_{v_i}}$$
    where d_{v_i} is the degree of node v_i.
    Notation: h
    Condition: $\delta \geq 1$
    Complexity: $O(n)$

23. **Keyword:** `Index`
    Name: *Index*
    Type: Real
    Definition: The *index* or *spectral radius* λ_1 of a graph $G = (V,E)$ is the greatest eigenvalue of its adjacency matrix.
    Notation: λ_1
    Complexity: $O(50n^2)$

24. **Keyword:** `Irr`
    Name: *Irregularity*
    Type: Integer
    Definition: The *irregularity Irr* of a graph $G = (V,E)$ is
    $$\sum_{(v_i,v_j)\in E} |d_{v_i} - d_{v_j}|$$
    where d_{v_i} is the degree of node v_i.
    Notation: *Irr*
    Complexity: $O(n^2)$
    Note: Introduced by Albertson (1997).

25. **Keyword:** `LapLargest`
    Name: *Laplacian largest eigenvalue*
    Type: Real
    Definition: The *Laplacian largest eigenvalue* λ_1^L of a graph $G = (V,E)$ is the largest eigenvalue of its Laplacian matrix L, defined by
    $$L = D - A$$
    where D is a diagonal matrix where $D_{i,i} = d_{v_i}$ (degree of node v_i) and A is the adjacency matrix of G.

Notation: λ_1^L
Complexity: $O(50n^2)$

26. **Keyword:** `MaxDeg`
Name: *Maximum Degree*
Type: Integer
Definition: The *maximum degree* Δ of a graph $G = (V,E)$ is

$$\max_{v_i \in V} d_{v_i}$$

where d_{v_i} is the degree of node v_i.
Notation: Δ
Complexity: $O(n)$

27. **Keyword:** `MaxMatching`
Name: *Maximum Matching*
Type: Integer
Definition: The *maximum matching* or *matching numbber* of a graph G is the maximum cardinality of the maximal matchings of G.
Notation: M
Note: Implemented by E. Levy. It represents the worst-case of the maximal matching heuristic (found by an exhaustive search).

28. **Keyword:** `MinDeg`
Name: *Minimum Degree*
Type: Integer
Definition: The *minimum degree* δ of a graph $G = (V,E)$ is

$$\min_{v_i \in V} d_{v_i}$$

where d_{v_i} is the degree of node v_i.
Notation: δ
Complexity: $O(n)$

29. **Keyword:** `MinDisjointCliques`
Name: *Minimum Number of Disjoint Cliques*
Type: Integer
Definition: The *minimum number of disjoint cliques* is the minimum cardinality of a clique partition of G. A *clique partition* of $G = (V,E)$ is a partition of V into disjoint subsets $V_1, V_2, \ldots, V_k$ such that, for $1 \leq i \leq k$, the subgraph induced by V_i is a complete graph.

Notation: κ
Complexity: NP hard

30. **Keyword:** `MinMaxMatching`
Name: *Minimum Maximal Matching*
Type: Integer
Definition: The *minimum maximal matching* of a graph G is the minimum cardinality of the maximal matchings of G.
Notation: *MMM*
Complexity: NP hard
Note: Implemented by J. Cardinal.

31. **Keyword:** `MinNodeCover`
Name: *Minimum Node Cover*
Type: Integer
Definition: The *minimum node cover* of a graph G is the minimum cardinality of a set of nodes that contains at least one endpoint of each edge of G.
Notation: C
Complexity: NP hard
Note: Implemented by J. Cardinal.

32. **Keyword:** `Nagamochi`
Name: *Worst case of Nagamochi's node cover algorithm*
Type: Integer
Definition: The *Nagamochi* value of a graph G is the worst case result of an approximation algorithm, from Nagamochi and Ibaraki for minimum vertex cover, applied on G.
Notation: N
Note: Implemented by J. Cardinal. Use a brute-force algorithm.

33. **Keyword:** `NodesEachMinNC`
Name: *Number of nodes belonging to all the node covers*
Type: Integer
Definition: The all_{nc} value of a graph G is the number of nodes belonging to all of its node covers.
Notation: all_{nc}
Note: Implemented by E. Levy.

34. **Keyword:** NodesNoneMinNC
Name: *Number of nodes belonging to none node cover*
Type: Integer
Definition: The $none_{nc}$ value of a graph G is the number of nodes belonging to none of its node covers.
Notation: $none_{nc}$
Note: Implemented by E. Levy. Use a brute-force algorithm.

35. **Keyword:** NodesSomeMinNC
Name: *Number of nodes belonging to some node covers*
Type: Integer
Definition: The $some_{nc}$ value of a graph G is the number of nodes belonging to some (but not all) of its node covers.
Notation: $some_{nc}$
Note: Implemented by E. Levy. Use a brute-force algorithm.

36. **Keyword:** NumConnComp
Name: *Number of Connected Components*
Type: Integer
Definition: The *number of connected components* of a graph G is the number of maximal connected subgraphs of G.
Notation: nc
Complexity: $O(n^2)$
Note: We use a Depth First Search.

37. **Keyword:** NumDeg1
Name: *Number of nodes with degree 1*
Type: Integer
Definition: The number n_1 of a graph G is the number of nodes with a degree equals to 1.
Notation: n_1
Complexity: $O(n)$

38. **Keyword:** NumDeg2
Name: *Number of nodes with degree 2*
Type: Integer
Definition: The number n_2 of a graph G is the number of nodes with a degree equals to 2.
Notation: n_2
Complexity: $O(n)$

39. **Keyword:** `NumDegNm1`
Name: *Number of nodes with degree n-1*
Type: Integer
Definition: The number n_{n-1} of a graph G, with n nodes, is the number of nodes with a degree equals to $n-1$.
Notation: n_{n-1}
Complexity: $O(n)$

40. **Keyword:** `NumEdges`
Name: *Number of edges*
Type: Integer
Definition: The *number of edges* m of a graph $G = (V,E)$, or its *size*, is the cardinal of E.
Notation: m
Complexity: $O(1)$

41. **Keyword:** `NumNodes`
Name: *Number of nodes*
Type: Integer
Definition: The *number of nodes* n of a graph $G = (V,E)$, or its *order*, is the cardinal of V.
Notation: n
Complexity: $O(1)$

42. **Keyword:** `NumOddCycles`
Name: *Number of Odd Cycles*
Type: Integer
Definition: The *number of odd cycles* of G is the number of cycles in G which have an odd length.
Notation: *OddC*
Note: Implemented by J. Cardinal.

43. **Keyword:** `PathWidth`
Name: *Path width*
Type: Integer
Definition: A *path decomposition* of a graph $G = (V,E)$ is a pair $(\{X_i : i \in I\}, T)$ where $T = (I,F)$ is a path and $\{X_i\}$ is a collection of subsets of V, such that

- $\bigcup_{i \in I} X_i = V$,
- for any $\{v,w\} \in E$, there exists an $i \in I$ with $u,v \in X_i$,
- for any $v \in V$, the set $\{i \in I : v \in X_i\}$ forms a connected subpath of T.

The *path width* of G is $\max_{i \in I} |X_i| - 1$.
Notation: *pw*
Complexity: NP hard
Note: Implemented by S. Langerman.

44. **Keyword:** `Proximity`
Name: *Proximity*
Type: Rational
Definition: The *proximity* of a graph $G = (V,E)$ is its minimum transmission, i.e.,

$$\min_{v \in V} \frac{1}{n-1} \sum_{u \in V} d(u,v)$$

where $d(u,v)$ is the distance between u and v.
Notation: T_{min}
Condition: *Connected*
Complexity: $O(n^3)$
Note: We use the Floyd-Marshall algorithm to compute all pairs shortest paths.

45. **Keyword:** `QueueNumber`
Name: *Queue number*
Type: Integer
Definition: A *queue layout* of a graph G consists of a linear order of the nodes and a partition of the edges into non-nested queues. Edge {x, y} is *nested* inside edge {v,w} if $v < x < y < w$ in the linear order. The *queue number* of a graph G is the minimum number of queues in a queue layout of G.
Notation: *qn*
Complexity: NP hard
Note: Implemented by S. Langerman.

46. **Keyword:** `Radius`
Name: *Radius*
Type: Integer
Definition: Let $G = (V,E)$ be a graph. The *eccentricity* of a node $v \in V$ is the maximum distance between v and any other node $u \in V$. The *radius* of G is the minimum eccentricity of its nodes.
Notation: r
Condition: *Connected*
Complexity: $O(n^3)$
Note: We use the Floyd-Marshall algorithm to compute all pairs shortest paths.

47. **Keyword:** `RandMaxMatching`
Name: *Expected randomized maximal matching*
Type: Real
Definition: Average result of the maximal matching heuristic. The heuristic iteratively removes an edge chosen uniformly at random among the remaining edges.
Notation: *RMM*
Note: Implemented by E. Levy. Use a brute-force algorithm.

48. **Keyword:** `Randic`
Name: *Randic index*
Type: Real
Definition: The *Randić index* $Ra(G)$ or *connectivity index* of a graph $G = (V,E)$ is defined by
$$Ra(G) = \sum_{(v_i,v_j)\in E} \frac{1}{\sqrt{d_{v_i}\, d_{v_j}}}$$
where d_{v_i} is the degree of node v_i.
Notation: *Ra*
Condition: $\delta \geq 1$
Complexity: $O(n^2)$

49. **Keyword:** `Regular`
Name: *Regular*
Type: Boolean
Definition: A graph is said to be *regular* if all of its nodes's degrees are equals.
Notation: *Regular*
Complexity: $O(n)$

50. **Keyword:** `Remoteness`
Name: *Remoteness*
Type: Rational
Definition: The *remoteness* of a graph $G = (V,E)$ is its maximum transmission, i.e.,
$$\max_{v\in V} \frac{1}{n-1} \sum_{u\in V} d(u,v)$$
where $d(u,v)$ is the distance between u and v.
Notation: T_{max}
Condition: *Connected*
Complexity: $O(n^3)$
Note: We use the Floyd-Marshall algorithm to compute all pairs shortest paths.

51. **Keyword:** `SecondEig`
Name: *Second largest eigenvalue*
Type: Real
Definition: The *second largest eigenvalue* λ_2 of a graph $G = (V,E)$ is the second greatest eigenvalue of its adjacency matrix.
Notation: λ_2
Complexity: $O(50n^2)$

52. **Keyword:** `SmallestEig`
Name: *Smallest eigenvalue*
Type: Real
Definition: The *smallest eigenvalue* λ_n of a graph $G = (V,E)$ is the smallest eigenvalue of its adjacency matrix.
Notation: λ_n
Complexity: $O(50n^2)$

53. **Keyword:** `StabNum`
Name: *Stability Number*
Type: Integer
Definition: The *stability number* of a graph G is the maximum cardinality of an independent set in G
Notation: α
Complexity: NP hard

54. **Keyword:** `SumDeg`
Name: *Sum of degrees*
Type: Integer
Definition: The *sum of degrees* of a graph G is the sum of the node's degrees of G and it is well known that
$$\sum_{i=1}^{n} d_{v_i} = 2m.$$
Notation: $\sum d_{v_i}$
Complexity: $O(1)$

55. **Keyword:** `Tree`
Name: *Tree*
Type: Boolean
Definition: A graph G is a *tree* iff it is connected and acyclic.

Notation: *Tree*
Complexity: $O(n^2)$
Note: We check if G is connected and if $m = n - 1$.

56. **Keyword:** `TreeWidth`
Name: *Tree width*
Type: Integer
Definition: A *tree decomposition* of a graph $G = (V,E)$ is a pair $(\{X_i : i \in I\}, T)$ where $T = (I,F)$ is a tree and $\{X_i\}$ is a collection of subsets of V, such that

- $\bigcup_{i \in I} X_i = V$,
- for any $\{v,w\} \in E$, there exists an $i \in I$ with $u,v \in X_i$,
- for any $v \in V$, the set $\{i \in I : v \in X_i\}$ forms a connected subtree of T.

The *tree width* of G is $\max_{i \in I} |X_i| - 1$.
Notation: *tw*
Complexity: NP hard
Note: Implemented by S. Langerman.

57. **Keyword:** `Unicyclic`
Name: *Unicyclic*
Type: Boolean
Definition: A graph is said to be *unicyclic* is it contains one (and only one) cycle.
Notation: *Unicyclic*
Complexity: $O(n^2)$
Note: We use a Depth First Search.

58. **Keyword:** `VarDeg`
Name: *Variance of degrees*
Type: Rational
Definition: The *variance of degrees* of a graph G is the variance of the degrees of its nodes.
Notation: σ^2
Complexity: $O(n)$

59. **Keyword:** `VarDist`
Name: *Variance of distances*
Type: Rational
Definition: The *variance of distances* of a graph G is the variance of the distances between any pair of distinct nodes of G.

Notation: $\sigma^2(l)$

Condition: *Connected*

Complexity: $O(n^3)$

Note: We use the Floyd-Marshall algorithm to compute all pairs shortest paths.

Bibliography

[1] ALBERTSON, M.O. The Irregularity of a Graph. *Ars Combinatoria 46* (1997), 219 – 225.

[2] AOUCHICHE, M., BELL, F.K., CVETKOVIĆ, D., HANSEN, P., ROWLINSON, P., SIMIĆ, S., AND STEVANOVIĆ, D. Variable Neighborhood Search for Extremal Graphs 16. Some Conjectures Related to the Largest Eigenvalue of a Graph. Submitted, 2005.

[3] AOUCHICHE, M., BONNEFOY, J.M., FIDAHOUSSEN, A., CAPOROSSI, G., HANSEN, P., HIESSE, L., LACHERÉ, J., AND MONHAIT, A. Variable Neighborhood Search for Extremal Graphs 14. The AutoGraphiX 2. In *Global optimization: from theory to implementation*, L. Liberti and N. Maculan, Eds. Springer, 2005. to appear.

[4] AOUCHICHE, M., CAPOROSSI, G., AND HANSEN, P. Variable Neighborhood Search for Extremal Graphs 8. Variations on Graffiti 105. *Congr. Numer. 148* (2001), 129 – 144.

[5] AOUCHICHE, M., CAPOROSSI, G., HANSEN, P., AND LAFFAY, M. AutoGraphiX: a survey. *Electronic Notes in Discrete Mathematics 22* (2005), 515 – 520.

[6] AOUCHICHE, M., AND HANSEN, P. Recherche à voisinage variable de graphes extrémaux 13. A propos de la maille. Submitted, 2005.

[7] AOUCHICHE, M., HANSEN, P., AND STEVANOVIĆ, D. Variable Neighborhood Search for Extremal Graphs 17. Further Conjectures and Results about the Index. Submitted, 2005.

[8] APPEL, K., AND HAKEN, W. Every Planar Map is Four Colorable. Part I. Discharging. *Illinois J. Math. 21* (1977), 429 – 490.

[9] APPEL, K., AND HAKEN, W. Every Planar Map is Four Colorable. Part II. Reducibility. *Illinois J. Math. 21* (1977), 491 – 567.

[10] APPEL, K., AND HAKEN, W. Every Planar Map is Four Colorable. *Contemp. Math. 98* (1989), 1 – 741.

[11] ARAUJO, O., AND DE LA PEÑA, J. A. Some Bounds for the Connectivity Index of a Chemical Graph. *J. Chem. Inf. Comp. Sci. 38* (1998), 827–831.

[12] ARCHIMEDES. The method of mechanical theorem proving, cited by J. Gray in "Sale of the Century ?". *The Mathematical Intelligencer 21* (1999), 12 – 15.

[13] AVIS, D., BREMNER, D., AND SEIDEL, R. How good are convex hull algorithms. *Computational Geometry 7* (1997), 265 – 301.

[14] BALABAN, A.T., Ed. *Chemical Applications of Graph Theory*. Academic Press, New York, 1976.

[15] BALABAN, A.T. Topological Indices and Their Uses: A New Approach for Coding of Alkanes. *J. Mol. Struct. (Theochem) 165* (1988), 243 – 253.

[16] BALIŃSKA, K.T., CVETKOVIĆ, D., LEPOVIĆ, M., AND SIMIĆ, S.K. There Are Exactly 150 Connected Integral Graphs up to 10 Vertices. *Univ. Beograd Publ. Elektrotehn. Fak. Ser. Mat. 10* (1999), 95 – 105.

[17] BALIŃSKA, K.T.., CVETKOVIĆ, D., RADOSAVLJEVIĆ, Z., SIMIĆ, S., AND STEVANOVIĆ, D. A survey on integral graphs. *Univ. Beograd Publ. Elektrotehn. Fak. Ser. Mat. 13* (2002), 42 – 65.

[18] BALIŃSKA, K.T., KUPCZYK, M., SIMIĆ, S.K., AND ZWIERZYŃSKI, K.T. On Generating All Integral Graphs on 11 Vertices. Tech. rep., Computer Science Center, Techn. Univ. of Poznań, 1999/2000.

[19] BALIŃSKA, K.T., KUPCZYK, M., AND ZWIERZYŃSKI, K.T. Methods of Generating Integral Graphs. Tech. rep., Computer Science Center, Techn. Univ. of Poznań, 1997.

[20] BALIŃSKA, K.T., LEPOVIĆ, M., AND SIMIĆ, S.K. The Nonregular, Bipartite, Integral Graphs with Maximum Degree Four. Part II: Computer Search. In preparation.

[21] BALIŃSKA, K.T., AND SIMIĆ, S.K. The Nonregular, Bipartite, Integral Graphs with Maximum Degree Four. Part I: Basic Properties. *Discrete Math. 236* (2001), 13 – 24.

[22] BAR-YEHUDA, R., AND S.EVEN. A local-ratio theorem for approximating the weighted vertex cover problem. *Ann. Discrete Math. 25* (1985), 27–45.

[23] BAREFOOT, C.A., ENTRINGER, R.C., AND MULLHAUPT, A.P. Computer Based Proofs by Induction in Graph Theory – A House of Cards ? *Math. Comput. Modelling 17*, 11 (1993), 17 – 23.

[24] BARGHOUTI, N.S., MOCENIGO, J.M., AND WENKE, L. Grappa : A GRAPh PAckage in Java. *Lecture Notes in Computer Science 1353* (1997), 336 – 343.

[25] BEHZAD, M., CHARTRAND, G., AND LESNIAK-FOSTER, L. *Graphs and Digraphs*. Prindle, Weber & Schmidt, Boston, 1979.

[26] BELHAIZA, S., DE ABREU, N.M.M., HANSEN, P., AND OLIVEIRA, C.S. Variable Neighborhood Search for Extremal Graphs 11. Bounds on algebraic connectivity. In *Graph Theory and Combinatorial Optimization*, D. Avis, A. Hertz, and O. Marcotte, Eds. Springer, 2005, pp. 1 – 15. to appear.

[27] BELL, F.K. A Note on the Irregularity of Graphs. *Linear Algebra Appl. 161* (1992), 45 – 54.

[28] BERGE, C. *Graphes et Hypergraphes*. Gauthier-Villars, Paris, 1983.

[29] BERQUE, D., CECCHINI, R., GOLDBERG, M., AND RIVENBURGH, R. The SetPlayer System for Symbolic Computation on Power Sets. *J. Symb. Comp. 14* (1992), 645 – 662.

[30] BERRY, J.W. Improving Discrete Mathematics and Algorithms Curricula with LINK. *SIGCSE Bulletin (ACM Special Interest Group on Computer Science Education) 29* (1997), 14 – 20.

[31] BERRY, J.W. LINK: A Software System for Experimentation with Graphs and Hypergraphs. *SIAM Discrete Math. Newsletter 2*, 7 (1997).

[32] BERRY, J.W., DEAN, N., GOLDBERG, M.K.., SHANNON, G.E., AND SKIENA, S. Graph Drawing and Manipulation with LINK. *Lecture Notes in Computer Science 1353* (1997), 425 – 437.

[33] BERRY, J.W., DEAN, N., GOLDBERG, M.K.., SHANNON, G.E., AND SKIENA, S. LINK: A System for Graph Computation. *Software - Practice and Experience* (1999).

[34] BIGGS, N.L., LLOYD, E.K., AND WILSON, R.J. *Graph Theory 1736 – 1936*. Clarendon Press, Oxford, 1977.

[35] BOLLOBAS, B. *Graph Theory*. Springer, New York, 1979.

[36] BOLLOBÁS, B., AND ERDÖS, P. Graphs of Extremal Weights. *Ars Combinatoria 50* (1998), 225–233.

[37] BONDY, J.A.., AND MURTY, U.S.R.. *Graph Theory with Applications*. McMillan, London, 1972.

[38] BOUVIER, A., AND GEORGE, M. *Dictionnaire des Mathématiques (in French)*. Presses Universitaires de France, Paris, 1979.

[39] BREWSTER, T.L., DINNEEN, M.J., AND FABER, V. A Computational Attack on the Conjectures of Graffiti: New Counterexamples and Proofs. *Discrete Math. 147* (1994), 35 – 55.

[40] BRIGHAM, R.C., AND DUTTON, R.D. Ingrid: A Software Tool for Extremal Graph Theory Research. *Congressum Numerantium 39* (1983), 337 – 352.

[41] BRIGHAM, R.C., AND DUTTON, R.D. A Compilation of Relations Between Graph Invariants. *Networks 15* (1985), 73 – 107.

[42] BRIGHAM, R.C., AND DUTTON, R.D. A Compilation of Relations Between Graph Invariants. Supplement 1. *Networks 21* (1991), 421 – 455.

[43] BRIGHAM, R.C., DUTTON, R.D., AND GOMEZ, F. INGRID: A Graph Invariant Manipulator. *J. Symb. Comp. 7* (1989), 163 – 177.

[44] BRINKMANN, G. Fast Generation of Cubic Graphs. *J. Graph Theory 23* (1996), 139 – 149.

[45] BRINKMANN, G. Isomorphism Rejection in Structure Generation Programs. In *Discrete Mathematical Chemistry*, P. Hansen, P. Fowler, and M. Zheng, Eds., vol. 51 of *DIMACS Series on Discrete Mathematics and Theoretical Computer Science*. Amer. Math. Soc., Providence RI, 2000, pp. 25 – 38.

[46] BRINKMANN, G., AND DRESS, A.W.M. A Constructive Enumeration of Fullerenes. *J. Algorithms 23* (1997), 345 – 358.

[47] CAGE. The Chemical and Abstract Graph Environment. Homepage: `http://www.mathematik.uni-bielefeld.de/~CaGe/`.

[48] CAPOBIANCO, M., AND MOLLUZZO, J.C. *Examples and Counterexamples in Graph Theory*. North Holland, New York, 1978.

[49] CAPOROSSI, G. *Découverte par ordinateur en théorie des graphes*. PhD thesis, Université de Montréal, 2000.

[50] CAPOROSSI, G., CVETKOVIĆ, D., GUTMAN, I., AND HANSEN, P. Variable Neighborhood Search for Extremal Graphs 2. Finding Graphs with Extremal Energy. *J. Chem. Inf. Comput. Sci. 39* (1999), 984 – 996.

[51] CAPOROSSI, G., DOBRYNIN, A.A., HANSEN, P., AND GUTMAN, I. Trees with Palindromic Hosoya Polynomials. *Graph Theory Notes N. Y. 37* (1999), 10 – 16.

[52] CAPOROSSI, G., GUTMAN, I., AND HANSEN, P. Variable Neighborhood Search for Extremal Graphs 4. Chemical Trees with Extremal Connectivity Index. *Computers and Chemistry 23* (1999), 469 – 477.

[53] CAPOROSSI, G., GUTMAN, I., HANSEN, P., AND PAVLOVIĆ, L. Graphs with maximum connectivity index. *Computers and Chemistry 27* (2003), 85 – 90.

[54] CAPOROSSI, G., AND HANSEN, P. Finding Relations in Polynomial Time. In *XVIth International Joint Conference on Artificial Intelligence (IJCAI)* (Stockholm, 1999), vol. 2, pp. 780 – 785.

[55] CAPOROSSI, G., AND HANSEN, P. Variable Neighborhood Search for Extremal Graphs 1. The AutoGraphiX System. *Discrete Math. 212* (2000), 29 – 44.

[56] CAPOROSSI, G., AND HANSEN, P. Variable Neighborhood Search for Extremal Graphs 5. Three Ways to Automate Finding Conjectures. *Discrete Math. 276* (2004), 81 – 94.

[57] CARBONNEAUX, Y., LABORDE, J.-M., AND MADANI, M. Cabri-Graphs: a Tool for Research and Teaching in Graph Theory. In *Lecture Notes in Computer Science*, vol. 1027. Berlin: Springer, 1995, pp. 123 – 127.

[58] CARDINAL, J., LABBÉ, M., LANGERMAN, S., LEVY, E., AND MÉLOT, H. A tight analysis of the maximal matching heuristic. In *Computing and Combinatorics: 11th Annual International Conference, COCOON 2005* (Kunming, China, 2005), vol. 3595 of *Lecture Notes in Computer Science*, Springer–Verlag, pp. 701 – 709.

[59] CHARTRAND, G., KRONK, H.V., AND WALL, C.E.. The point arboricity of a graph. *Isr. J. Math. 6* (1968), 169 – 175.

[60] CHINN, P.Z. Disovery-method teaching in graph theory. *Annals Discrete Math. 55* (1993), 375 – 384.

[61] CHLEBÌK, M., AND CHLEBÌKOVÁ, J. Approximation hardness of minimum edge dominating set and minimum maximal matching. In *Proceedings of 14th International Symposium on Algorithms and Computation (ISAAC)* (2003), Lecture Notes in Computer Science, Springer–Verlag.

[62] CHRISTOPHE, J., DEWEZ, S., DOIGNON, J.-P., ELLOUMI, S., FASBENDER, G., GRÉGOIRE, P., HUYGENS, D., LABBÉ, M., MÉLOT, H., AND YAMAN, H. Linear inequalities among graph invariants: using *GraPHedron* to uncover optimal relationships. *Submitted, 2005*.

[63] CHUNG, F.R.K.. The average distance and the independence number. *J. Graph Theory 12* (1988), 229 – 235.

[64] CHVÁTAL, V. The Smallest Triangle-free 4-chromatic 4-regular Graph. *J. Combinatorial Theory 9* (1970), 93 – 94.

[65] CLEMENTI, A., AND TREVISAN, L. Improved non-approximability results for vertex cover problems with density constraints. *Theoretical Computer Science 225*, 1–2 (1999), 113–128.

[66] CODDING, P.W., Ed. *Structure-Based Drug Design: Experimental and Computational Approaches*. Kluwer, Dordrecht, 1998.

[67] COLLATZ, L., AND SINOGOWITZ, U. Spektren endlicher Grafen. *Abh. Math. Sem. Univ. Hamburg 21* (1957), 63 – 77.

[68] CONDER, M. Semi - Automated Theorem Proving. The Impact of Computers on Research in Pure Mathematics. *First Asian Technology Conference in Mathematics* (Singapore, December 1995).

[69] CONDER, M. An Infinite Family of 5-arc-transitive Cubic Graphs. *Ars Combinatoria 25A* (1988), 95 – 108.

[70] CONDER, M. A Surprising Isomorphism. *J. of Algebra 129* (1990), 494 – 501.

[71] CONDER, M. Hexagon Free Subgraphs of Hypercubes. *J. Graph Theory 17* (1993), 477 – 479.

[72] CONDER, M., AND LORIMER, P. Automorphism Groups of Symmetric Graphs of Valency 3. *J. Combinatorial Theory Ser. B 47* (1989), 60 – 72.

[73] COOK, R.J. The Theory and Applications of Graphs. In *Proceedings of the Fourth International Conference on the Theory and Applications of Graphs, 6 – 9 May 1980* (Western Michigan Univ., Kalamazoo, Michigan, 1981), G. Chartrand, Y. Alavi, D.L. Goldsmith, L. Lesniak-Foster, and D.R. Lick, Eds., New York: John Wiley, p. 610.

[74] CORMEN, T.H., LEISERSON, C.E., AND RIVEST, R. *Introduction to algorithms*. The MIT Press, Mc Graw-Hill, 1991.

[75] CVETKOVIĆ, D. Chromatic Number and the Spectrum of a Graph. *Publ. Inst. Math. (Beograd) 14 (28)* (1972), 25 – 38.

[76] CVETKOVIĆ, D. Discussing Graph Theory with a Computer, II: Theorems Suggested by the Computer. *Publ. Inst. Math. (Beograd) 33 (47)* (1983), 29 – 33.

[77] CVETKOVIĆ, D. Discussing Graph Theory with a Computer, IV: Knowledge Organisation and Examples of Theorem Proving. In *Proc. Fourth Yugoslav Seminar on Graph Theory* (Novi Sad, 1983), pp. 43 – 68.

[78] CVETKOVIĆ, D. Discussing Graph Theory with a Computer, VI: Theorems Proved with the Aid of the Computer. *Cl. Sci. Math. Natur., Sci. Math. T. XCVII* (1988), No. 16, 51 – 70.

[79] CVETKOVIĆ, D., DOOB, M., GUTMAN, I., AND TORGASEV, A. Recent Results in the Theory of Graph Spectra. *Annals of Discrete Mathematics 36* (1988), 1 – 306.

[80] CVETKOVIĆ, D., DOOB, M., AND SACHS, H. *Spectra of Graphs*. Johann Ambrosius Barth, Heidelberg, 1995.

[81] CVETKOVIĆ, D., AND KRAUS, L. "Graph" an Expert System for the Classification and Extension of the Knowledge in the Field of Graph Theory, User's Manual. Elektrotehn. Fak., Beograd, 1983.

[82] CVETKOVIĆ, D., KRAUS, L., AND SIMIĆ, S. Discussing Graph Theory with a Computer, I: Implementation of Graph Theoretic Algorithms. *Univ. Beograd Publ. Elektrotehn. Fak, Ser. Mat. Fiz. No. 716 – No. 734* (1981), 100 – 104.

[83] CVETKOVIĆ, D., AND PEVAC, I. Discussing Graph Theory with a Computer, III: Man-Machine Theorem Proving. *Publ. Inst. Math. (Beograd) 34 (48)* (1983), 37 – 47.

[84] CVETKOVIĆ, D., AND PEVAC, I. Man-Machine Theorem Proving in Graph Theory. *Artificial Intell. 35* (1988), 1 – 23.

[85] CVETKOVIĆ, D., AND ROWLINSON, P. On Connected Graphs with Maximal Index. *Publ. Inst. Math. (Beograd) 44* (1988), 29 – 34.

[86] CVETKOVIĆ, D., AND SIMIĆ, S. Graph Theoretical Results Obtained by the Support of the Expert System "Graph". *Bull. Acad. Serbe Sci. Arts Cl. Sci. Math. Natur. 19* (1994), 19 – 41.

[87] CVETKOVIĆ, D., AND SIMIĆ, S. Graph Theoretical Results Obtained by the Support of the Expert System "Graph" – An Extended Survey. In *Graphs and Discovery*, Fajtlowicz, S. *et al.*, Ed., vol. 69 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, Providence, 2005, pp. 39 – 70.

[88] CVETKOVIĆ, D., SIMIĆ, S., CAPOROSSI, G., AND HANSEN, P. Variable Neighborhood Search for Extremal Graphs 3. On the Largest Eigenvalue of Color-Constrained Trees. *Lin. and Multilin. Algebra 2* (2001), 143 – 160.

[89] DEAN, N., AND LIU, Y. Discrete Mathematics Software for the K-12 Education. In *Discrete Mathematics in the Schools: Making an Impact*, D.S. Franzblau, F.S. Roberts, and J.G. Rosenstein, Eds., vol. 36 of *DIMACS Series*. AMS, 1997.

[90] DE LA VINA, E. Bibliography on Conjectures of Graffitti. Available at `http://cms.dt.uh.edu/faculty/delavinae/research/wowref.htm`, 2000.

[91] DELAVINA, E. Some history of the development of Graffiti. In *Graphs and Discovery*, Fajtlowicz, S. *et al.*, Ed., vol. 69 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, Providence, 2005, pp. 81 – 118.

[92] DEVILLERS, J. *Neural Networks in QSAR and Drug Design*. Academic Press, London, 1996.

[93] DEVILLERS, J., AND BALABAN, A.T., Eds. *Topological Indices and Related Descriptors in QSAR and QSPR*. Gordon and Breach, The Nederlands, 1999.

[94] DEVILLERS, J., AND BALADAN, A., Eds. *Topological Indices and Related Descriptors in QSAR and QSPAR*. Gordon and Breach, Newark, 2000.

[95] DI BATTISTA, G., EADES, P., TAMASSIA, R., AND TOLLIS, I.G. Algorithms for Drawing Graphs: an Annotated Bibliography. *Computational Geometry: Theory and Applications 4*, 5 (1994), 235 – 282.

[96] DI BATTISTA, G., EADES, P., TAMASSIA, R., AND TOLLIS, I.G. *Graph Drawing : Algorithms for the Visualization of Graphs*. Prentice Hall, 1999.

[97] DIESTEL, R. *Graph Theory*, second edition ed. Springer-Verlag, 2000.

[98] DOBRYNIN, A.A., ENTRINGER, R., AND GUTMAN, I. Wiener Index of Trees: Theory and Applications. *Acta Appl. Math. 66* (2001), 211 – 249.

[99] DOUCET, J.-P.., AND WEBER, J. *Computer-Aided Molecular Design*. Academic Press, London, 1996.

[100] EADES, P., AND SUGIYAMA, K. How to Draw a Directed Graph. *J. Information Processing 13*, 4 (1990), 424 – 434.

[101] EDWARDS, C.S., AND ELPHICK, C.H. Lower bounds for the clique and the chromatic numbers of a graph. *Discrete Appl. Math. 5* (1983), 51 – 64.

[102] EPSTEIN, S.L. On the Discovery of Mathematical Theorems. In *Proceedings of the Tenth International Joint Conference on Articial Intelligence* (Milan, Italy, 1987), pp. 194 – 197.

[103] EPSTEIN, S.L. Learning and Discovery : One System's Search for Mathematical Knowledge. *Comput. Intell. 4* (1988), 42 – 53.

[104] EPSTEIN, S.L., AND SRIDHARAN, N.S. Knowledge Representation for Mathematical Discovery : Three Experiments in Graph Theory. *J. Applied Intelligence 1* (1991), 7 – 33.

[105] EREMEEV, A. On some approximation algorithms for dense vertex cover problem. In *Proceedings of SOR* (1999), Springer–Verlag.

[106] EULER, L. Solutio problematis ad geometriam situs pertinentis. *Comment. Academiae Sci. I. Petropolitanae 8* (1736), 128 – 140.

[107] FAJTLOWICZ, S. Written on the Wall. A regulary updated file accessible from `http://www.math.uh.edu/~clarson/`.

[108] FAJTLOWICZ, S. On Conjectures of Graffiti – II. *Congr. Numer. 60* (1987), 187 – 197.

[109] FAJTLOWICZ, S. On Conjectures of Graffiti. *Discrete Math. 72* (1988), 113 – 118.

[110] FAJTLOWICZ, S. On Conjectures of Graffiti – III. *Congr. Numer. 66* (1988), 23 – 32.

[111] FAJTLOWICZ, S. On Conjectures of Graffiti – IV. *Congr. Numer. 70* (1990), 231 – 240.

[112] FAJTLOWICZ, S. On Conjectures of Graffiti – V. In *Seventh International Quadrennial Conference on Graph Theory* (1995), vol. 1, pp. 367 – 376.

[113] FARADZEV, I.A. Constructive Enumeration of Combinatorial Object. In *Problèmes combinatoires et Théorie des Graphes, Colloque Internat. CNRS 260* (1978), pp. 131 – 135.

[114] FARADZEV, I.A. *Generation of nonisomorphic graphs with given degree sequence (Russian).* Nauka, Moscow, 1978, in Algorithmic Studies in Combinatorics, pp. 11 – 19.

[115] FARADZEV, I.A., AND KLIN, M.H. Computer Package for computations with Coherent Configurations. In *Proc. ISSAC-91* (Bonn, 1991), ACM Press, pp. 219 – 223.

[116] FARADZEV, I.A., KLIN, M.H., AND MUZYCHUCK, M.E. Cellular Rings and Groups of Automorphism of Graphs. In *Investigations in algebraic theory of combinatorial objects* (Dordrecht, 1994), vol. 84 of *Mathematics and its Applications*, Kluwer Academic Publishers, pp. 1 – 152.

[117] FAVARON, O., MAHÉO, M., AND SACLÉ, J.-F. Some Results on Conjectures of Graffiti. I. *Ars Combin. 29* (1990), 90 – 106.

[118] FAVARON, O., MAHÉO, M., AND SACLÉ, J.-F. On the Residue of a Graph. *J. Graph Theory 15* (1991), 39 – 64.

[119] FAVARON, O., MAHÉO, M., AND SACLÉ, J.-F. Some Eigenvalue Properties in Graphs (Conjectures of Graffiti. II). *Discrete Math. 111* (1993), 197 – 220.

[120] FINCK, H.-J.., AND GROHMANN, G. Vollständiges Produkt, chromatische Zahl und charakteristisches Polynom regulärer Graphen, I. *Wiss. Z. TH Ilmenau 11* (1965), 1 – 3.

[121] FIORINI, S., AND WILSON, R.J. On the Chromatic Index of a Graph II. In *Combinatorics (Proc. British Combinatorial Conf.)* (Univ. Coll. Wales, Aberystwyth, 1973, 1974), vol. 13 of *London Math. Soc. Lecture Note Ser.*, Cambridge Univ. Press, London, pp. 37 – 51.

[122] FISCHERMANN, M., GUTMAN, I., HOFFMANN, A., RAUTENBACH, D., VIDOVIĆ, D., AND VOLKMANN, L. Extremal Chemical Trees. *Z. Naturforsch. 57a* (2002), 49 – 52.

[123] FISCHERMANN, M., HOFFMANN, A., RAUTENBACH, D., SZÉKELY, L., AND VOLKMANN, L. Wiener Index Versus Maximum Degree in Trees. *Discrete Appl. Math. 122* (2002), 127 – 137.

[124] FISCHERMANN, M., HOFFMANN, A., RAUTENBACH, D., AND VOLKMANN, L. A Linear-Programming Approach to the Generalized Randic Index. *Discrete Appl. Math. 128* (2003), 375 – 385.

[125] FOULDAS, C.A., AND PARDALOS, P.M., Eds. *Optimization in Computational Chemistry and Molecular Biology: Local and Global Approaches.* Kluwer, Dordrecht, 2000.

[126] FOULDS, L.R. *Graph Theory Applications.* Springer-Verlag, New York, 1992.

[127] FOWLER, P.W., HANSEN, P., CAPOROSSI, G., AND SONCINI, A. Polyenes with Maximum HOMO-LUMO Gap. (Variable Neighborhood Search for Extremal Graphs 7). *Chemical Physics Letters 342* (2001), 105 – 112.

[128] FOWLER, P.W., AND MANOPOULOS, D.E. *An Atlas of Fullerenes*. Oxford Univ. Press, Oxford, 1995.

[129] FRÖLICH, M., AND WERNER, M. Demonstration of the Interactive Graph-Visualization System da Vinci. In *Lectures Notes in Computer Science*, vol. 894. Springer - Verlag, 1994, pp. 266 – 269.

[130] FUKUDA, K. cdd/cdd+ reference manual. cdd ver. 0.61, cdd+ ver. 0.76 (1999), available at `http://www.cs.mcgill.ca/~fukuda/soft/cddman/cddman.html`.

[131] GAP. Groups, Algorithms and Programming. Homepage : `http://www-groups.dcs.st-andrews.ac.uk/~gap/`.

[132] GAREY, M.R.., AND JOHNSON, D.S.. *Computers and intractability. A guide to the theory of NP-completeness*. Freeman and Company, 1979.

[133] GAWRILOW, E., AND JOSWIG, M. polymake: a Framework for Analyzing Convex Polytopes. In *Polytopes — Combinatorics and Computation*, G. Kalai and G.M.. Ziegler, Eds. Birkhäuser, 2000, pp. 43–74.

[134] GAWRILOW, E., AND JOSWIG, M. polymake: an Approach to Modular Software Design in Computational Geometry. In *Proceedings of the 17th Annual Symposium on Computational Geometry* (2001), ACM, pp. 222–231. June 3-5, 2001, Medford, MA.

[135] GMP. GNU Multiple Precision Arithmetic Library. Homepage : http://www.swox.com/gmp/.

[136] GRAPHEDRON. Web portal at `http://www.graphedron.net`.

[137] GREGORY, C., AND LATKA, B. Minimal Antichains in Well-Founded Quasi-Orders with an Application to Tournaments. *J. Combinatorial Theory Series B 80* (2000), 258 – 276.

[138] GROSS, J.L., AND YELLEN, J. *Graph Theory and its Applications*. CRC Press, Inc., Boca Raton, FL, USA, 1999.

[139] GRUND, R., KERBER, A., AND LAUE, R. MOLGEN : Ein Computeralgebra-System für die Konstruktion molekularer Graphen. *Communications in Math. Chemistry (MATCH) 27* (1992), 87 – 131.

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

[141] GUTMAN, I. Total π-Electron Energy of Benzenoid Hydrocarbon. *Topics in Current Chemistry 162* (1992), 29 – 63.

[142] GUTMAN, I., ARAUJO, O., AND MORALES, D.A. Bounds for the Randić Connectivity Index. *J. Chem. Inf. Comp. Sci.*, 40 (2000), 572–598.

[143] GUTMAN, I., ARAUJO, O., AND RADA, J. . *Acta Chimica Hungarica Models in Chemistry*, 137 (2000), 653.

[144] GUTMAN, I., AND CYVIN, S. *Introduction to the Theory of Benzenoid Hydrocarbons*. Springer - Verlag, 1989.

[145] GUTMAN, I., AND DAS, K.C.. The First Zagreb Index 30 Years After. *MATCH Commun. Math. Comput. Chem. 50* (2004), 83 – 92.

[146] GUTMAN, I., HANSEN, P., AND MÉLOT, H. Variable Neighborhood Search for Extremal Graphs 10. Comparison of Irregularity Indices for Chemical Trees. *J. Chem. Inf. Mod. 45* (2005), 222 – 230.

[147] GUTMAN, I., AND HOU, Y. Bipartite Unicyclic Graphs with Greatest Energy. *Comm. Math. Comp. Chem. 43* (2001), 17 – 28.

[148] GUTMAN, I., AND MILJKOVIĆ. Molecules with Smallest Connectivity Indices. *Comm. in Math. Chem. (MATCH)*, 41 (2000), 57–70.

[149] GUTMAN, I., MILJKOVIĆ, O., CAPOROSSI, G., AND HANSEN, P. Alkanes with Small and Large Randić Connectivity Index. *Chemical Physics Letters*, 306 (1999), 366–372.

[150] GUTMAN, I., MILJKOVIĆ, O., AND PAVLOVIĆ, L. On Graphs with Extremal Connectivity Indices. *Bull. Acad. Serbe Sci. Arts 121* (2000), 1 – 14.

[151] GUTMAN, I., AND POLANSKY, O.E. *Mathematical Concepts in Organic Chemistry*. Springer-Verlag, Berlin, 1988.

[152] GUTMAN, I., RUŠČIĆ, B., TRINAJSTIĆ, N., AND WILCOX JR, C.F. Graph Theory and Molecular Orbitals. XII. Acyclic Polyenes. *J. Chem. Phys. 62* (1975), 3399 – 3405.

[153] GUTMAN, I., AND TRINAJSTIĆ, N. Graph Theory and Molecular Orbitals. Total π-Electron Energy of Alternant Hydrocarbons. *Chem. Phys. Lett. 17* (1972), 535 – 538.

[154] GUTMAN, I., AND VIDOVIĆ, D. The Largest Eigenvalues of Adjacancy and Laplacian Matrices, and Ionization Potentials of Alkanes. *Indian J. Chem. 41A* (2002), 893 – 896.

[155] HALPERIN, E. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. *Siam Journal on Computing 31* (2002), 1608–1623.

[156] HANSCH, C., LEO, A., AND HOEKMAN, D. *Exploring QSAR, Vols. I and II*. ACS Professional Reference Book, Oxford University Press, New York, 1995.

[157] HANSEN, P. Automated Comparison of Graph Invariants. Communication presented at *Computers and Discovery III*, Ghent, Belgium, 6 – 9 February 2006, Joint work with M. AOUCHICHE and G. CAPOROSSI.

[158] HANSEN, P. Bornes et algorithmes pour les stables d'un graphe. In *A look at graph theory (Proc. Colloq., Cerisy, 1980) (French)*. Presses Polytech. Romandes, Lausanne, 1980, pp. 39–53.

[159] HANSEN, P. How Far Should, Is And Could Be Conjecture-Making Automated in Graph Theory ? In *Graphs and Discovery*, Fajtlowicz, S. *et al.*, Ed., vol. 69 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, Providence, 2005, pp. 189 – 230.

[160] HANSEN, P., AOUCHICHE, M., CAPOROSSI, G., MÉLOT, H., AND STEVANOVIĆ, D. What Forms Do Interesting Conjectures Have in Graph Theory? In *Graphs and Discovery*, Fajtlowicz, S. *et al.*, Ed., vol. 69 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, Providence, 2005, pp. 231 – 252.

[161] HANSEN, P., LABBÉ, M., AND MÉLOT, H. Automatization of conjecture-making in GraPHedron. In preparation, 2006.

[162] HANSEN, P., AND MÉLOT, H. Computers and Discovery in Algebraic Graph Theory. *Linear Algebra and its Applications 356* (2002), 211 – 230.

[163] HANSEN, P., AND MÉLOT, H. Variable Neighborhood Search for Extremal Graphs 6. Analysing Bounds for the Connectivity Index. *J. Chem. Inf. Comput. Sci. 43* (2003), 1 – 14.

[164] HANSEN, P., AND MÉLOT, H. Variable Neighborhood Search for Extremal Graphs 9. Bounding the Irregularity of a Graph. In *Graphs and Discovery*, Fajtlowicz, S. *et al.*, Ed., vol. 69 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, Providence, 2005, pp. 253 – 264.

[165] HANSEN, P., MÉLOT, H., AND GUTMAN, I. Variable Neighborhood Search for Extremal Graphs 12. A Note on the Variance of Bounded Degrees in Graphs. *MATCH Comm. Math. Comp. Chem. 54* (2005), 221 – 232.

[166] HANSEN, P., MÉLOT, H., AND STEVANOVIĆ, D. Integral Complete Split Graphs. *Univ. Beograd Publ. Elektrotehn. Fak. Ser. Mat. 13* (2002), 89 – 95.

[167] HANSEN, P., AND MLADENOVIĆ, N. Variable Neighborhood Search : Principles and Applications. *European J. of Oper. Res. 130* (2001), 449 – 467.

[168] HANSEN, P., AND STEVANOVIĆ, D. Variable Neighborhood Search for Extremal Graphs 15. On bags and bugs. Submitted, 2005.

[169] HANSEN, P., AND ZHENG, M.L. Sharp bounds on the order, size, and stability number of graphs. *Networks 23* (1993), 99 – 102.

[170] HARANT, J., AND SCHIERMEYER, I. On the independence number of a graph in terms of order and size. *Discrete Math. 232*, 1-3 (2001), 131–138.

[171] HARARY, F., AND PALMER, E.M. *Graphical Enumeration*. Academic Press, New York, 1973.

[172] HARARY, F., AND SCHWENK, A.J.. Which graphs have intergral spectra ? In *Graphs and Combinatorics*, R. Bari and F. Harary, Eds. Springer-Verlag, Berlin, 1974, pp. 45 – 51.

[173] HASTAD, J. Some optimal inapproximability results. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing* (1997), pp. 1–10.

[174] HIMSOLT, M. GraphEd: A Graphical Platform for the Implementation of Graph Algorithms. In *Lectures Notes in Computer Science*, vol. 894. Springer - Verlag, 1994, pp. 182 – 193.

[175] HIMSOLT, M. The graphlet sytem. In *Graph Drawing*, S. North, Ed., vol. 1190 of *Lectures Notes in Computer Science*. Springer - Verlag, 1996, pp. 233 – 240.

[176] IBARAKI, T., AND NAGAMOCHI, H. An approximation of the minimum vertex cover in a graph. *Japan J. Indust. Appl. Math. 16* (1999), 369–375.

[177] IMAMURA, T., AND IWAMA, K. Approximating vertex cover on dense graphs. In *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms (SODA)* (2005).

[178] KARAKOSTAS, G. A better approximation ratio for the vertex cover problem. *ECCC Report TR04-084* (2004).

[179] KARELSON, M. *Molecular Descriptors in GSAR/QSPR*. Wiley, New York, 2000.

[180] KARPINSKI, M., AND ZELIKOVSKY, A. Approximating dense cases of covering problems. In *Proceedings of the DIMACS Workshop on Network Design: Connectivity and Facilites Location* (1997), P. Pardalos and D. Du, Eds., vol. 40 of *DIMACS series in Disc. Math. and Theor. Comp. Sci.*, pp. 169–178.

[181] KIER, L.B., AND HALL, L.H. *Molecular Connectivity in Chemistry and Drug Research*. Academic Press, New York, 1976.

[182] KIER, L.B., AND HALL, L.H. *Molecular Connectivity in Structure-Activity Analysis*. Research Studies Press, John Wiley and Sons, Letchworth, England, 1986.

[183] KIER, L.B., AND HALL, L.H. *Molecular Structure Description*. Academic Press, London, 1999.

[184] KNUTH, D. *The Stanford Graphbase : a Platform for Combinatorial Computing*. Addison - Wesley, Reading, Massachusetts, 1993.

[185] LATKA, B. Finitely Constrained Classes of Homogeneous Directed Graphs. *J. Symb. Logic 59 (1)* (1994), 124 – 139.

[186] LEE, L.-Q., SIEK, J.G., AND LUMSDAINE, A. The Generic Graph Component Library. *ACM SIGPLAN Notices 34* (1999), 399 – 414.

[187] LICK, D.R. A Class of Point Partition Numbers. In *Recent Trends in Graph Theory (Proc. Conf. New York 1970)* (1971), vol. 186 of *Lectures Notes in Mathematics*, Springer, Berlin, pp. 185 – 190.

[188] LOVÁSZ, L. *Combinatorial Problems and Exercices*. North-Holland, New York, 1979.

[189] LOVÁSZ, L., AND PELIKAN, J. On the eigenvalues of trees. *Periodica Math. Hungar. 3* (1973), 175 – 182.

[190] LUTZ, M., AND KENADIN, T. *Quantitative Molecular Pharmacology and Informatics in Drug Discovery*. Wiley, New York, 1999.

[191] MATHWORKS, INC. Matlab: The language of Technical Computing. The MathWorks, Inc.

[192] MCKAY, B.D. Nauty User's Guide (version 1.5). Tech. rep., Department of Computer Science, Australian National University, 1990.

[193] MCKAY, B.D. Isomorph-free Exhaustive Generation. *J. Algorithms 26* (1998), 306 – 324.

[194] MEHLHORN, K., AND NÄHGER, S. Leda: A Platform for Combinatorial and Geometric Computing. *Communications of the ACM 38 (1)* (1995), 96 – 102.

[195] MÉLOT, H. The Evolution of GraPHedron. Communication presented at *Computers and Discovery III*, Ghent, Belgium, 6 – 9 February 2006.

[196] MÉLOT, H. Facet defining inequalities among graph invariants: the system GraPHedron. *Submitted*, 2005.

[197] MEVENKAMP, M., DEAN, N., AND MONMA, C. NETPAD User's Guide and Reference Guide. Bellcore, 1990.

[198] MLADENOVIĆ, N., AND HANSEN, P. Variable Neighbourhood Search. *Computers and Operations Research 24* (1997), 1097 – 1100.

[199] MONIEN, B., AND SPECKENMEYER, E. Ramsey numbers and an approximation algorithm for the vertex cover problem. *Acta Inf. 22* (1985), 115–123.

[200] NEWBERRY, P.F. The Design of an Extendible Graph Editor. In *Lectures Notes in Computer Science*, vol. 704. Springer - Verlag, 1993.

[201] NIKOLIĆ, S., KOVACEVIĆ, G., MILICEVIĆ, A., AND TRINAJSTIĆ, N. The Zagreb Indices 30 Years After. *Croat. Chem. Acta 76* (2003), 113 – 124.

[202] ORE, O. *Theory of graphs*. American Mathematical Society Colloquium Publications, Vol. XXXVIII. American Mathematical Society, Providence, R.I., 1962.

[203] PAPADIMITRIOU, C. H., AND YANNAKAKIS, M. Optimization, approximation, and complexity classes. *J. Comput. System. Sci. 43*, 3 (1991), 425–440.

[204] POLYA, G., AND READ, R.C. *Combinatorial Enumeration of Groups*. Springer, 1987.

[205] PRAEGER, C.E., AND SOICHER, L.H. Low rank representations and graphs for sporadic groups. In *Australian Math. Soc. Lecture Series*, vol. 8 of *Australian Math. Soc. Lecture Series*. Cambridge University Press, 1997.

[206] PUJOL, F. Etude de conjectures en théorie des graphes. Internship Report under direction of P. Hansen (GERAD), 1998.

[207] RADOSAVLJEVIĆ, Z., AND SIMIĆ, S. There Are Just Thirteen Connected Nonregular Nonbipartite Integral Graphs Having Maximum Vertex Degree Four. In *Graph Theory, Proc. VI Yugoslav Seminar on Graph Theory, Dubrovnik, April 18 – 19, 1985* (University of Novi Sad, Institute of Mathematics, Novi Sad, 1987), R. Tosić, D. Acketa, and V. Petrović, Eds., pp. 183 – 187.

[208] RADOSAVLJEVIĆ, Z., AND SIMIĆ, S. The Nonregular, Nonbipartite Integral Graphs with the Maximum Degree Four. *J. Comb. Infor. Syst. Sci. 20* (1995), 9 – 26.

[209] RADZISZOWSKI, S.P. Small Ramsey Numbers. Dynamic Survey 1. *Electr. J. of Combin.* (1994). Updated 1998.

[210] RANDIĆ, M. On characterization of molecular branching. *J. Am. Chem. Soc. 97* (1975), 6609 – 6615.

[211] READ, R.C. Every One a Winner. *Annals of Discrete Math. 2* (1978), 107 – 120.

[212] READ, R.C. Some Applications of Computers in Graph Theory. In *Selected Topics in Graph Theory*, L.W. Beineke and R.J. Wilson, Eds. London - New York: Academic Press, 1978, pp. 417 – 444.

[213] ROBERTSON, N., SANDERS, D., SEYMOUR, P., AND THOMAS, R. The Four-Color Theorem. *J. Combinatorial Theory, Ser. B 70* (1997), 2 – 44.

[214] ROWLINSON, P. On the Index of Certain Outerplanar Graphs. *Ars Combinatoria 29C* (1990), 271 – 275.

[215] SANDER, G. Graph Layout through the VCG Tool. In *Lectures Notes in Computer Science*, vol. 894. Springer - Verlag, 1994, pp. 194 – 205.

[216] SHANNON, G., MEEDEN, L., AND FRIEDMAN, D. SchemeGraph : an Object-Oriented Environment for Manipulating Graphs. Software and documentation, 1990.

[217] SHAO, J.Y., AND HONG, Y. Bounds on the smallest positive eigenvalue of a tree with perfect matchings. *Chinese Sci. Bull.*, 37(9) (1992), 713 – 717.

[218] SIMIĆ, S. On the Largest Eigenvalue of Unicyclic Graphs. *Publ. Inst. Math. (Beograd) 42 (56)* (1987), 13 – 19.

[219] SIMIĆ, S. Some Results on the Largest Eigenvalue of a Graph. *Ars Combinatoria 24A* (1987), 211 – 219.

[220] SKIENA, S. *Implementing Discrete Mathematics : Combinatorics and Graph Theory with Mathematica*. Addison-Wesley, 1990.

[221] SOICHER, L.H. GRAPE: a System for Computing with Graphs and Groups. In *Groups and Computation*, L. Finkelstein and W.M. Kantor, Eds., vol. 11 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. AMS, 1993, pp. 287 – 291.

[222] SOICHER, L.H. Three new distance-regular graphs. *Europ. J. Combinatorics 14* (1993), 501 – 505.

[223] STANLEY, R.P. A Bound on the Spectral Radius of Graphs with *e* Edges. *Linear Algebra Appl. 87* (1987), 267 – 269.

[224] STEVANOVIĆ, D. *Some Compositions of Graphs and Integral Graphs (in Serbian)*. PhD Thesis, University of Niš, 2000, 2000.

[225] SYLVESTER, J.J. Chemistry and Algebra. *Nature 17* (1878), 284.

[226] TODESCHINI, R., AND CONSONNI, V. *Handbook of Molecular Descriptors*, vol. 11 of *Methods and Principles in Medicinal Chemistry*. Wiley-VCH, 2000.

[227] TODESCHINI, R., AND CONSONNI, V. *Handbook of Molecular Descriptors*. Wiley, New York, 2001.

[228] TURÁN, P. Eine Extremalaufgabe aus der Graphentheorie. *Mat. Fiz. Lapok 48* (1941), 436–452.

[229] VAN DE WATERBEEMD, H., Ed. *Structure-Property Correlations in Drug Research*. Academic Press, London, 1996.

[230] VAN DE WATERBEEMD, H., TESTA, B., AND FOLKERS, G., Eds. *Computer-Assisted Lead Finding and Optimization*. Proceedings of the 11th European Symposium on QSAR, VCH, Weinheim, 1996.

[231] VEGA. A System for Manipulating Discrete Mathematical Structures. Project Homepage : `http://vega.ijp.si/`.

[232] WOLFRAM, RESEARCH INC. Mathematica Language and Software. Wolfram Research, Inc.

[233] YANNAKAKIS, M., AND GAVRIL, F. Edge dominating sets in graphs. *SIAM J. Appl. Math. 38*, 3 (1980), 364–372.

[234] YU, P. (in chinese). *J. Math. Studies*, 31 (1998), 225.

[235] ZHOU, B., AND GUTMAN, I. Relations Between Wiener, Hyper-Wiener and Zagreb Indices. *Chem. Phys. Lett. 394* (2004), 93 – 95.

[236] ZIEGLER, G. M.. *Lectures on Polytopes*, revised ed. Springer-Verlag, Berlin, 1998.

Index

Notations