Comparison of Graph Invariants in Predicting Fullerene Stability

Thiago H. de A. Lemos, André Luiz Pires Guedes*

Federal University of Paraná, Curitiba/PR, Brasil
thalemos@inf.ufpr.br, andre@inf.ufpr.br

Abstract

Fullerenes are molecules constituted entirely by tertiary carbon atoms, forming a surface that contains only pentagonal and hexagonal faces. This paper presents a comparison of some graph theoretic invariants in the context of predicting the stability of fullerenes. In particular: the Fowler-Manolopoulos criterion, the diameter, the Wiener index, the bipartite edge frustration, the independence number, the number of perfect matchings, the Fries number, and the Taylor number. With three exceptions, their values were computed for each fullerene isomer with up to 130 vertices, and each isolated pentagon isomer with up to 160 vertices.

1 Introduction

For most of the twentieth century, the only known substances constituted entirely by carbon atoms were diamond, graphite, and amorphous carbon. That changed in 1985, when Kroto, Heath, O'Brien, Curl, and Smalley provided the first hint of the existence of an aggregate with 60 carbon atoms [52], later confirmed to be shaped as a truncated icosahedron [50]. Designated *Buckminsterfullerene*, in honour of Richard Buckminster Fuller's geodesic domes, the structure was later found to appear naturally in soot, along with other similar molecules [14]. Since

then, the simplified term fullerene is used to designate any molecule comprised wholly of tertiary carbon atoms, forming polyhedral structures in which all faces are either pentagons or hexagons. An extensive, in-depth treatment of fullerenes and their properties can be found in Fowler and Manolopoulos' "An Atlas of Fullerenes" [27].

One of the main problems encountered by chemists when dealing with fullerenes is the fact that the C_n formula does not serve as a unique identifier for a fullerene with n atoms. For example, although the Buckminsterfullerene is denoted by C_{60}, its structure is only one among the 1812 combinatorially distinct ways of arranging 60 carbon atoms without violating the definition of fullerene – the others lead to *isomers* of Buckminsterfullerene (molecules with the same chemical formula, but different structures). Actually, certain results from the area of mathematics that studies polyhedric surfaces (mathematical abstractions useful to model the structure of fullerenes) show that there should be fullerenes C_n for each even value of $n \geq 20$ (with the exception of 22) [35], and the number of isomers of C_n is in $\Theta(n^9)$ [14, 57].

However, even though the number of isomers for a particular C_n can be significantly high, only a fraction of these structures is stable enough to be successfully isolated in experiment, existing for long enough to be detected. The difficulty, then, is how to know *which* isomers are more relevant (i.e., *stable*) among the mass of possibilities for each C_n. One of the ways researchers found to answer this question is based on the observation that the chemical properties of a molecule (and, consequently, its expected stability) are closely related to certain invariants in its molecular graph – a graph in which the vertices correspond to the atoms and the edges to the bonds between the atoms. Such technique has proven particularly adequate in predicting the stability of fullerenes, since the presence of no atoms other than carbon simplifies the representation via molecular graph.

So far, several invariants have been investigated in the literature, though it is hard to determine how effective they actually are. Generally, when a group of authors propose a new invariant to predict the stability of fullerenes, there is little concern in presenting experimental results comparing the performance of the new

method with that of other invariants. Even in those cases in which comparisons are performed, they tend to take into account only one or two more traditional invariants.

Ergo, the main contribution of this work consists in presenting some evidence of the consistency (or discrepancy) between expected stabilities according to each of several invariants selected from literature – that is, investigate if and how much these invariants "agree" as to which are the most stable isomers. Furthermore, the experiments documented in the literature are *usually* restricted to values of n around 100 or 120, given the computational difficulty.

Thus, a second objective of this work is to raise this limit in a way as to encompass the studied invariants as uniformly as possible. In particular, calculations were made for each isomer with up to 130 vertices, and each isolated pentagon isomer with up to 160 vertices – with three exceptions due to lack of time and computational power: the independence number (up to 148), the Fries number (up to 124 and from 132 to 142), and the Taylor number (up to 130). These exceptions are properly signaled from now on.

2   Definitions and Notation

A *graph* is an ordered pair $G = (V, E)$, where V is a non-empty set of *vertices* and $E \subseteq \binom{V}{2}$ is a set of *edges* disjoint from V. The *order* and *size* of a graph G correspond, respectively, to the values $n = |V|$ and $m = |E|$. Given a vertex $v \in V$, $N_G(v)$ denotes the set of all neighbours of v in G, and $deg_G(v) = |N_G(v)|$ denotes the *degree* of v. Each edge $e = \{u, v\}$ in a graph G may optionally be associated to a non-negative number denominated *weight of e* and denoted by the symbol $w(e)$. Note that this work deals only with finite and simple graphs, so, from now on, it is assumed that the term "graph" is always used with this meaning.

A graph G is said to be *complete* if, and only if, all its vertices are pairwise adjacent. A complete graph with n vertices is denoted by K_n. A *bipartite graph*, G, is a graph in which the vertex set, V, can be partitioned in two disjoint subsets X and Y, such that each edge $e \in E$ has one endpoint in X and another in Y. This (X, Y) partition is denominated a *bipartition* of G. A *matching* $M \subseteq E$ in

a graph G is a subset of edges from E such that each vertex $v \in V$ is incident on, at most, one edge in M. Vertices incident on some edge of M are said to be *matched*, and the others are said to be *unmatched*. A matching M in G is said to be a *perfect matching* if all vertices of G are matched. Note also that, when G has weights associated with its edges, the *weight* of a matching M in G is defined as the sum of the weights of all edges in M.

Given these preliminary concepts, it is now possible to explain how to represent fullerenes using graphs. As defined in [14], a *fullerene graph* with n vertices, denoted by C_n, is a 3-regular, planar graph, where each face is formed by 5 or 6 edges. Note that, due respectively to 3-regularity and 3-connectivity, C_n must contain exactly $3n/2$ edges and have a unique immersion in the plane (preserving face boundaries). From this definition it is possible to use Euler's formula [18] and conclude that fullerenes have $n/2 - 10$ hexagonal faces and exactly 12 pentagonal faces (regardless of the size) [27]. Also according to Fowler and Manolopoulos, fullerenes C_n exist only for even values of n, with the exception of $n = 22$.

However, the notation C_n works as a unique identifier only for the three simplest fullerenes (those with n equal to 20, 24, or 26). For values other than those, the notation is ambiguous, as there are multiple C_n isomers, each one with a distinct fullerene graph. One popular solution to this problem involves ordering the isomers lexicographically using their smallest *face spiral sequences* [14,27]. For a given value of n, the notation for each isomer is combined with a number k (corresponding to its position in this order), resulting in $C_n{:}k$. For example, the Buckminsterfullerene can be identified as $C_{60}{:}1812$, because its smallest face spiral is the last among the 1812 isomers.

It is also important to make a brief comment regarding the use of the word *stability* when talking about a fullerene isomer. In the literature about fullerenes, even if the possible meanings of thermodynamic and kinetic stability are somewhat related, the term is usually applied in an ambiguous way. So, in order to avoid any confusion, from now on the expression *stable isomer* will be used specifically in its less formal sense (similarly to [21]): to designate isomers that were or that *theoretically* can be produced in sufficient quantity to be observed in laboratory.

The first criterion used successfully to determine the stability of fullerenes was the *isolated pentagon rule*, proposed by Kroto and Schmalz [51, 60] and later formalized by Raghavachari [53] and Fowler and Manolopoulos [27]. This rule is based on the observation that adjacencies between pentagons cause an increase in what is known as *steric strain* of the molecule, which negatively impacts stability. For $n \geq 70$ all fullerenes have at least one isomer with no pentagon adjacencies and each one of them receives the special denomination of *IPR isomer*. Sadly, even if this criterion has shown itself consistent with experimental observations regarding stability, beyond a certain point even the number of IPR isomers becomes too great to be efficiently manageable at present. Given that the isolated pentagon rule is unable to distinguish between IPR isomers, this problem motivated the search for new evaluation schemes.

3 Project and Implementation

The first step to accomplish the aforementioned goals for this work was obtaining a set of fullerene graphs on which the analysis could be performed. An obvious approach would involve using the *buckygen* [7] or *fullgen* [9–11] programs, but the files generated by these programs do not order isomers using the lexicographically smallest face spiral[1] (something indispensable for a correct analysis here, since several other works use this same referencing scheme). The alternative solution was using the *House of Graphs* [8] database as the source of fullerene graphs for this work.

After this was settled, we choose to compare the following invariants: i) the Fowler-Manolopoulos criterion, ii) the diameter, iii) the Wiener index, iv) the bipartite edge frustration, v) the independence number, vi) the number of perfect matchings (also known as the number of Kekulé structures), vii) the Fries number, viii) the Taylor number. These choices were made trying to strike a balance between the importance of the invariant in the literature, its expected predicting potential, and the difficulty in its implementation. Programs were written to

---

[1]One of the authors was able to successfully contact (via email) researchers Gunnar Brinkmann, Brendan McKay, and Jan Goedgebeur, who confirmed those programs define their own custom order for the isomers.

compute the value of these invariants on the selected set of fullerene graphs. The following sections, after a short review of the literature for each invariant, briefly discuss implementation details. All programs were written in C++, and are available (publicly and free for academic activities) on the web[2]. Please note that at certain points some third party libraries had to be used, and those may be subject to specific license agreements.

3.1 Fowler-Manolopoulos Criterion

In the same paper treating the isolated pentagon rule, Raghavachari also proposes an adaptation that allows one to deal with fullerenes for which there are multiple IPR isomers. According to Raghavachari [53], one should consider not only *if* two pentagonal faces are separate, but *how* separate they are. Theoretically, the steric tension (which interferes in stability) is minimized when the curvature of the pentagonal regions is distributed in the most uniform possible way through the surface of the molecule [27].

The first step in quantifying this distribution is defining a way to deal with hexagons, instead of pentagons. For each hexagon, the *hexagonal neighborhood index* is the number of other hexagons to which it is adjacent. This way it is possible to assign to each isomer a signature $\langle h_0, h_1, h_2, h_3, h_4, h_5, h_6 \rangle$, where h_k is the number of hexagons with neighborhood index k. The signatures can be summarized into a single number, which represents the value of the invariant for a given fullerene graph. Fowler and Manolopoulos defined this number as the standard deviation, σ_h, of the distribution of the hexagonal neighborhood index [27]:

$$\sigma_h = \sqrt{\langle k^2 \rangle - \langle k \rangle^2} \qquad \text{where} \quad \langle k \rangle = \frac{\sum_{k=0}^{6} k h_k}{\sum_{k=0}^{6} h_k} \quad \text{and} \quad \langle k^2 \rangle = \frac{\sum_{k=0}^{6} k^2 h_k}{\sum_{k=0}^{6} h_k}$$

The program to compute the value of this invariant was divided in three stages: i) obtaining a set of dual graphs from the set of fullerene graphs (which spawned a subprogram, denominated *planar2dual*, implementing an algorithm adapted from

---

[2]http://www.inf.ufpr.br/arg/fullerene

Guibas, Stolfi and Guedes [36,38]), ii) iterating over all neighbours of vertices with degree 6 in the dual, computing hexagon neighborhood indices, and iii) applying directly the formula above.

3.2   Diameter and Wiener Index

The Wiener Index, denoted by W, is a topological index presented in 1947 by the North American chemist Harry Wiener [66] that has a close relation with a great number of physical and chemical properties of the alkanes (a type of hydrocarbon). Although Wiener defined W only for the alkanes, the Japanese chemist Haruo Hosoya [43] was the first to propose a relationship between W and the distances in a general molecular graph. Hosoya showed that the value of W in a molecule corresponds to exactly half the sum of all the elements in the distance matrix of its corresponding molecular graph (the matrix M containing one line and one column for each vertex in the graph, where position $M_{i,j}$ corresponds to the distance between the vertices identified by i and j). In other words, for a graph $G = (V, E)$, the index $W(G)$ is merely the sum of the distances between all pairs of vertices of G [56] (considering only unique pairs, that is).

The Wiener index is a way to measure the ramification of the framework formed by the carbon atoms in the molecule and, consequently, to measure how compact it really is [6,55]. This means that W represents the reason between surface and volume in a molecule formed by carbon atoms, and serves as a sign of the presence of certain intermolecular forces [39] which, in turn, affect stability. The expected relationship is that, as W increases, the stability decreases, but there are already studies making comments about the low precision of this relationship (for some fullerenes), such as the work of Slanina et al. [61].

An invariant that is conceptually related to the Wiener index, and that has already been studied for some classes of fullerenes [25], is the *diameter*. The diameter of a graph G is the largest distance between two vertices of G, considering all possible pairs of vertices. Even if this invariant is not considered particularly relevant by the literature in predicting the stability of fullerenes, investigating the use of the diameter for this purpose is not entirely without justification. Andova

at al. [3] defined an upper bound for the diameter of a fullerene graph, G, as being $diam(G) \leq \frac{1}{5}n+1$. In the same paper it is highlighted that, even if the diameter is relatively small in fullerene graphs with symmetry similar to that of the Buckminsterfullerene (icosahedral), the diameter of graphs modeling nanotubes is linear on the number of vertices. Furthermore, this work presents a good opportunity to compare the diameter (obtained "for free" when calculating the Wiener index) with the other invariants. Both invariants were easily computed from the distance matrices of each graph, constructed using an implementation of the Floyd-Warshal algorithm adapted from Cormen et al. [12].

3.3   Bipartite Edge Frustration

One rather intuitive measure of the "bipartivity" of a graph can be defined through counting the number of edges violating the most remarkable characteristic of a bipartite graph: the fact that the two endpoints of each edge are not part of the same bipartition class (edges violating a bipartition in this way are said to be *frustrated* with respect to the bipartition). This invariant was first introduced in the study of complex networks [42] and, although not computable efficiently in the general case, it can be calculated in polinomial time in the specific case of fullerene graphs [17]. For each bipartition (X, Y) of the vertex set, V, of a given graph, G, denote by F_{XY} the set formed by all edges frustrating bipartition (X, Y). The *bipartite edge frustration* of G, denoted by $\varphi(G)$, is the cardinality of the smallest F_{XY} considering all possible bipartitions (X, Y) of V. Equivalently, $\varphi(G)$ can be defined as the size of the smallest set of edges that need to be removed from G to obtain a bipartite spanning subgraph. More formally:

$$\varphi(G) = \text{mín}\left\{|F_{XY}| \mid (X, Y) \text{ is a bipartition of } V\right\}$$

Došlić and Vukčević show [17] the relationship that exists between $\varphi(G)$ and a particular set of edges in the dual dual graph of G. This set, denoted by H and denominated *obstacle*, is the set formed by edges that need to be removed from G' (the dual of G) in order to obtain a spanning subgraph without vertices of odd degree. Let $K_{12}(G)$ be the complete graph having one vertex for each

pentagonal face of G, where each edge is associated to a weight corresponding to the distance between the faces represented by its endpoints (this graph receives the special denomination of *pentagon distance graph* of G). Again according to Došlić and Vukčević, each obstacle of minimum cardinality will be mapped directly to a minimum weight perfect matching in $K_{12}(G)$. The weight of this matching (i.e., the sum of the weights of its edges) corresponds to the value of the bipartite edge frustration in G.

Similarly to the program calculating the Fowler-Manolopoulos criterion, the program corresponding the the bipartite edge frustration receives the duals of the fullerene graphs as input. The list of vertices is iterated over, and a breadth-first search is performed at each vertex with degree 5. The distances between these vertices are used to create the K_{12} graph, and a special library (implementing the Blossom V algorithm [48]) is used to calculate a minimum weight perfect matching.

3.4 Independence Number

The application of this invariant as a criterion for predicting fullerene stability was initially motivated by one of the conjectures made automatically by the *Graffiti* program [19, 20]. Despite originating from extremely limited data, the fortuitous observation made by the program, that stable isomers had smaller independence number than the others, seemed to be a promising criterion in selecting the more relevant isomers (at least in the limited range of values of n considered at the time). To define formally the independence number, it is first necessary to revise the concept of *independent set*. An independent set, $S \subseteq V$, of a graph $G = (V, E)$ is a (maximal) subset of the vertices of G such that the induced subgraph $G[S]$ has no edges. An independent set with maximum cardinality is called a *maximum independent set*, and its size corresponds to the *independence number* of graph G, denoted by $\alpha(G)$. More formally [14]:

$$\alpha(G) = \text{máx}\{|S| \mid S \text{ is an independent set of } G\}$$

Although the statistical observation made by the program *Graffiti* has motivated the proposition of the independence number as a criterion to predict stability in fullerenes [21], there was no coherent chemical explanations justifying

this proposition. Actually, for the specific case of carbon nanotubes, the independence number is specially imprecise, given that this type of fullerene has an independence number far greater than other isomers with the same number of atoms (what doesn't necessarily mean that they are any less stable) [21]. According to recent studies [26, 28], there doesn't seem to be much correlation between the stability of fullerenes and the independence number.

Given the extremely high number of input graphs, and the fact that the problem is known to be *NP-Hard* [33], once again the choice was made to use a specialized library, but one relying on a slightly different definition of the independence number, based on the concept of *clique*. A clique of a given graph, G, is a (maximal) subset of vertices of G such that the induced subgraph $G[C]$ is complete. In this case, the *mcqd* [49] library was used to calculate a maximum clique on the *complement* of each input graph. The size of this clique corresponds exactly to the independence number on the original graph.

3.5   Number of Kekulé Structures

In Chemistry, the concept equivalent to that of a perfect matching in a carbon framework is known as *Kekulé Structure* (in honor of the German chemist Friedrich August Kekulé), and the number of structures of this type supported by the molecule (denoted here by $K(G)$ for a fullerene graph G) provides a measure of its electronic stability. This number, for the specific case of planar graphs, can be computed in polynomial time by the *FKT* algorithm − an acronym formed by the initials of the researchers Fisher, Kasteleyn and Temperley [45, 63], who independently discovered a result (posteriorly generalized by Kasteleyn for all planar graphs [46, 47]) which would lead to the algorithm. The pseudocode, adapted from [65], can be seen in Algorithm 1.

Given the need of calculating determinants and of representing the graphs using adjacency matrices, the implementation of this algorithm involved using a widely known library for linear algebra: the *Eigen* library [37]. It is also important to mention that, as noted in [4], the measure of electronic stability provided by the number of Kekulé structures seem to be slightly rough. Even in the limited

---

Algorithm 1 FKT(G)

Require: A planar graph $G = (V, E)$, immersed in the plane
Ensure: Number of perfect matchings, $K(G)$, of G
  Let $T = (V, F)$ be a spanning tree of G
  for each $e \in F$ **do**
    Give e an arbitrary orientation
  end for
  Let H be a tree with one vertex for each face in G (including the outer face) and an edge between two vertices if their corresponding faces in G share an edge not in T. The root can be chosen arbitrarily, for example, the vertex corresponding to the outer face.
  for each leaf v in H that is not the root **do**
    Let f be the face G corresponding to v
    Let e be the (only) edge of G in f without an orientation
    Orient e such that the number of clockwise-oriented edges in f is odd
    Remove v from H
  end for
  Let M be the oriented adjacency matrix of G (filled only with $1, -1$, and 0)
  return $|\sqrt{det(M)}|$

---

analysis in this paper, the pure counting of Kekulé structures fails to isolate the Buckminsterfullerene as the most stable isomer [59]. The invariant also appears to favor non-spheric structures with high steric strain (probably owing to the fact that a high value of K is associated with more pentagon adjacencies), unlikely candidates for stable molecules. This suggests that some matchings in particular may be more important than others, or than counting the pure total.

3.6   Fries Number

One of the first works to take this notion into account was that of Fries, who proposed the number of *benzenoid hexagons* as a criterion for stability [30, 31]. A *benzenoid hexagon* (with respect to a given Kekulé structure of matching M) is nothing more than a hexagonal face such that the bonds represented by its edges are alternatingly simple (σ) and double (π), or, equivalently, such that its edges are alternatingly part of M. The notation and terminology used here will be that of [4]. That is, the *Fries number* for a perfect matching M, denoted by $n_F(M)$, is the number of benzenoid hexagons in M. On the other hand, the *Fries number* for a given fullerene $G = (V, E)$, denoted by $F(G)$, is the maximum value of $n_F(M)$

considering all possible perfect matchings in G. The problem of finding $F(G)$ can be modeled as an integer programming problem, as described in [40], which was the strategy followed here. In particular, after experimenting with the $GLPK$[3], $SCIP$ [1], and lp_solve [5] libraries, the preference was for the GLPK, mostly because of its ease of integration with pre-existing code.

3.7 Taylor Number

Even though the Fries number is an improvement over the pure Kekulé count, rigidly following the criterion of maximizing the number of benzenoid hexagons might not be a very "refined" technique, specially for the IPR isomers (which are already known to be more prone to stability). In particular, those perfect matchings having largest n_F end up positioning their edges along pentagonal faces of the IPR fullerene, something that should be avoided, according to Taylor [62], with greater priority than maximizing the number of benzenoid hexagons.

For a perfect matching M, let $n_T(M)$ be the number of edges of M belonging to pentagonal faces in G, and let $n_{P2}(M)$ be the number of pentagonal faces containing exactly 2 edges of M. Two variants of the *Taylor number*, named in honor of Taylor, are proposed in [4], both defined (similarly to the Fries number $F(G)$) as being the maximum number of benzenoid hexagons $n_F(M)$ among all perfect matchings M of a given set of perfect matchings.

The difference is that the first version, $T_a(G)$, first selects those perfect matchings *minimizing* the value of $n_T(M)$, and then selects those *maximizing* $n_F(M)$. Whereas the second version, $T_b(G)$, begins by selecting those perfect matchings minimizing $n_T(M)$ *and* $n_{P2}(M)$, and then selects those maximizing $n_F(M)$. Unlike what happened with the Fries number, no explanation was found detailing how to model the calculation of the Taylor number using integer programming, only a statement in [4] that it had been done. Therefore, after selecting one of the versions to implement ($T_a(G)$), the method in [40] was adapted to include some new variables and restrictions, and the GLPK library was used once more.

---

[3]Access the link https://www.gnu.org/software/glpk/

4   Experiments and Results

As mentioned previously, one of the objectives of this work involves calculating
the values of the invariants for all isomers with n between 20 and 130, and all IPR
isomers with n between 132 and 160. Therefore, for each invariant the experiments
were executed in this order, following the values of n sequentially. Several machines
were used in this computation, including three shared machines from UFPR's
Department of Informatics (all running operating system Linux Mint 17 Qiana)
and a Macbook Pro with processor Intel Core 2 Duo 2.4GHz, 4GB of RAM, and
Mac OS X Mountain Lion. For almost all experiments the execution time was
no more than a few hours or days (even with the amount of graphs to consider).
Only for three invariants the execution was so long that the planned objective
could not be completed: the independence number, the Fries number, and the
Taylor number. One attempted strategy to mitigate this problem was changing
the code to allow parallel execution of the programs for different values of n and,
although it had some effect, it was not enough to complete all experiments in time.
In particular, the independence number was calculated for the values of n between
20 and 148, the Fries number for the values between 20 and 124 and between 132
and 142, and the Taylor number for the values between 20 and 130.

The output of all experiments was a set of approximately 500 *.csv* files, occu-
pying over 4GB of disk space. There is one file for each invariant and each value
of n in the designated range (with the exception of the diameter and the Wiener
index, which are stored in the same files). Inside each file there is one line for each
isomer, containing its identifier and the invariant value. Three main approaches
were chosen to analyze and extract conclusions from such a large data set: com-
parison with values from other works (Subsection 4.1), verification of performance
in predicting isomers known to be stable (Subsection 4.2), and statistical analysis
(Subsection 4.3).

4.1   Comparison with Values from Other Works

Comparing the values calculated for this work with those calculated in others
provides an independent measure of the correctness of the written programs and

gives a certain degree of confidence to all "new" values (those not yet calculated by anyone else or not found by these authors among the literature). It is important to mention that, given the libraries used for some invariants and the nature of *floating point* values, the results of some computations might be subject to minute rounding errors, which will be ignored in this comparison (i.e., the rounded value will be considered). Following is a list of values from other works that match those computed here (no useful reference was found for the diameter):

- Fowler-Manolopoulos criterion: Tables 4.3, 4.4, and 4.5, from [27], show the value of the invariant for the IPR isomers with 76, 78, and 84 atoms;

- Wiener index: [22] shows the value of this invariant for isomers of C_{40};

- Bipartite edge frustration: Table 1 from [17] shows the maximum value of $\varphi(G)$ for all IPR isomers with n between 94 and 120, and also for those with $n = 140$ or 180. This paper also states that the maximum value of $\varphi(G)$ for $n = 40$ is achieved only by two isomers, and that all IPR isomers with n between 60 and 92 have $\varphi(G) = 12$ (both statements corroborated here);

- Independence number: Table 3 from [21] shows the maximum and minimum values of the invariant for all values of n corresponding to the isomers known to be stable, along with the value for these isomers in particular;

- Number of Kekulé Structures, Fries Number, and Taylor Number: Table 4 from [4] shows the values of these invariants for some isomers, and Table 5 shows, for the Fries number, the minimum, maximum, and number of isomers to achieve these limits considering isomers of C_{60}. Another paper [64] (Table 1) also shows a selection of numbers of Kekulé structures for values of n between 20 and 60.

4.2   Stable Isomers

In [21] there is a list of isomers known to be stable that, although limited, can be used to give an ideia of the effectiveness of each invariant. The list is enumerated based on spiral codes for the IPR fullerenes only, and they correspond to the

following isomers on the more general notation: C_{60} : 1812, C_{70} : 8149, C_{76} : 19150, C_{78} : 24105, C_{78} : 24106, C_{78} : 24107, C_{84} : 51590, C_{84} : 51591. Tables 1 to 8, which can be seen below, document, for each invariant, the situation of each of those special isomers.

Each table is comprised of five columns: *Isomer*, *Max.?*, *Min.?*, *% abv.*, *% blw.*. They inform, respectively, the code identifying the isomer, if it reaches the maximum value (in which case there is a number between parenthesis indicating how many other isomers reach the maximum), if it reaches the minimum value (again with a number if it does), the percentage of isomers (for the same value of n) with value strictly above it, and the percentage of isomers with value strictly below it. Note that, when there are multiple isomers with the same value, the sum of percentages might be nowhere near close to 100 (since repeated values do not count as strictly above or below). When interpreting the content of these tables, a sign of good performance is the capacity of the invariant to isolate each of these special isomers in the most precise way possible. In this sense, there are two "ideal" scenarios: all isomers on the table are among the few to reach maximum value, or all of them are among the few to reach minimum value. In the first case, the second column would be filled with "Y" (and the value between parenthesis would be the smallest possible), the third column with "N", the fourth column would be as close to zero as possible, and the fifth column as close to 100 as possible. In the second case, these values would be inverted.

Despite the fact the Fowler-Manolopoulos criterion was expected (according to the literature) to outperform all others as the most precise invariant, and even though its performance is quite reasonable, what is seen here is that Tables 4 (bipartite edge frustration) and 5 (independence number) clearly illustrate the two ideal scenarios. At least for this limited range, both invariants seem capable of filtering more than 99% of possible isomers. Another interesting detail on those two cases is that, as the value of n increases, there is a slight increase in the number of isomers reaching the maximum (or minimum). In order to prevent this from becoming a problem for higher values of n, a good strategy seems to be the one adopted by [17]: considering only IPR isomers, which by themselves are already a

small fraction of the total number of isomers.

On the other end of the spectrum, the diameter seems to be one of the least useful invariants in predicting the stability of fullerenes. As all isomers for a given n are very similar, the diameters of the corresponding graphs end up repeating themselves over and over again, meaning that, even when a stable isomer reaches the minimum, the number of other isomers also doing that is so high that defeats any advantage of using the invariant in the first place. For the Wiener index the situation is only a little better, as it is off to a very nice start (precisely filtering several of the stable isomers) but later has its efficacy drastically reduced as n goes beyond 78. If this behavior confirms itself for even higher values, the invariant will end up being of as little use as the diameter.

As for the number of Kekulé structures, although faring better than the diameter (filtering out a significant part of the total), the invariant seems to be easily replaceable by the Fries number or the Taylor number, showing a clearly superior (and nearly identical, with a small advantage for the Taylor number) performance. Perhaps surprisingly, the Fries number and the Taylor number seem to perform as good as or even better than the Fowler-Manolopoulos criterion. It remains to be seen if this advantage maintains itself for higher values of n.

Table 1. F.-M. Crit. – stable isomers.

Isomer	Max.?	Min.?	% abv.	% blw.
$C_{60} : 1812$	N	Y(1)	99,94	0,00
$C_{70} : 8149$	N	N	99,21	0,64
$C_{76} : 19150$	N	N	95,97	4,03
$C_{78} : 24105$	N	N	88,63	11,36
$C_{78} : 24106$	N	N	94,48	5,31
$C_{78} : 24107$	N	N	99,87	0,07
$C_{84} : 51590$	N	Y(3)	99,99	0,00
$C_{84} : 51591$	N	Y(3)	99,99	0,00

Table 2. Diameter – stable isomers.

Isomer	Max.?	Min.?	% abv.	% blw.
$C_{60} : 1812$	N	Y(1261)	30,41	0,00
$C_{70} : 8149$	N	N	26,20	0,02
$C_{76} : 19150$	N	Y(4845)	74,70	0,00
$C_{78} : 24105$	N	Y(2619)	89,14	0,00
$C_{78} : 24106$	N	N	8,51	10,86
$C_{78} : 24107$	N	N	8,51	10,86
$C_{84} : 51590$	N	N	33,59	0,02
$C_{84} : 51591$	N	N	33,59	0,02

Table 3. Wiener Index – stable isomers.

Isomer	Max.?	Min.?	% abv.	% blw.
$C_{60} : 1812$	N	Y(1)	99,94	0,00
$C_{70} : 8149$	N	Y(1)	99,99	0,00
$C_{76} : 19150$	N	Y(3)	99,98	0,00
$C_{78} : 24105$	N	Y(1)	99,99	0,00
$C_{78} : 24106$	N	N	99,78	0,18
$C_{78} : 24107$	N	N	87,45	11,17
$C_{84} : 51590$	N	N	37,88	60,55
$C_{84} : 51591$	N	N	41,16	57,11

Table 4. Bip. Edge Frust. - stable isomers

Isomer	Max.?	Min.?	% abv.	% blw.
$C_{60} : 1812$	Y(1)	N	0,00	99,94
$C_{70} : 8149$	Y(1)	N	0,00	99,99
$C_{76} : 19150$	Y(2)	N	0,00	99,99
$C_{78} : 24105$	Y(5)	N	0,00	99,98
$C_{78} : 24106$	Y(5)	N	0,00	99,98
$C_{78} : 24107$	Y(5)	N	0,00	99,98
$C_{84} : 51590$	Y(33)	N	0,00	99,94
$C_{84} : 51591$	Y(33)	N	0,00	99,94

Table 5. Ind. Number − stable isomers.

Isomer	Max.?	Min.?	% abv.	% blw.
C_{60} : 1812	N	Y(1)	99,94	0,00
C_{70} : 8149	N	Y(1)	99,99	0,00
C_{76} : 19150	N	Y(1)	99,99	0,00
C_{78} : 24105	N	Y(3)	99,99	0,00
C_{78} : 24106	N	Y(3)	99,99	0,00
C_{78} : 24107	N	N	95,30	0,01
C_{84} : 51590	N	Y(17)	99,97	0,00
C_{84} : 51591	N	Y(17)	99,97	0,00

Table 6. Num. of K. Str. − stable isomers.

Isomer	Max.?	Min.?	% abv.	% blw.
C_{60} : 1812	N	N	1,10	98,84
C_{70} : 8149	N	N	3,77	96,22
C_{76} : 19150	N	N	9,69	90,31
C_{78} : 24105	N	N	13,80	86,20
C_{78} : 24106	N	N	4,46	95,54
C_{78} : 24107	N	N	30,39	69,60
C_{84} : 51590	N	N	28,08	71,92
C_{84} : 51591	N	N	29,99	70,01

Table 7. F. Number − stable isomers.

Isomer	Max.?	Min.?	% abv.	% blw.
C_{60} : 1812	Y(1)	N	0,00	99,94
C_{70} : 8149	Y(8)	N	0,00	99,90
C_{76} : 19150	N	N	0,08	97,52
C_{78} : 24105	N	N	0,53	96,50
C_{78} : 24106	N	N	0,00	99,97
C_{78} : 24107	N	N	0,03	99,47
C_{84} : 51590	N	N	0,53	97,02
C_{84} : 51591	N	N	0,03	99,47

Table 8. T. Number − stable isomers.

Isomer	Max.?	Min.?	% abv.	% blw.
C_{60} : 1812	Y(1)	N	0,00	99,94
C_{70} : 8149	Y(8)	N	0,00	99,94
C_{76} : 19150	N	N	0,07	99,12
C_{78} : 24105	N	N	0,22	99,11
C_{78} : 24106	N	N	0,00	99,98
C_{78} : 24107	N	N	0,02	99,81
C_{84} : 51590	N	N	1,34	94,98
C_{84} : 51591	N	N	0,02	99,85

4.3 Statistical Analysis

The statistical analysis of the computed values allows to measure, independent from the literature, how much the different invariants "agree" or "disagree" among themselves. In particular, the chosen method for the analysis was *Spearman's rank correlation coefficient* [41], calculated for each pair of invariants and taking into account the *order* of the data instead of its intrinsic value. By convention, the coefficient ranges from −1 (strongest negative correlation) to 1 (strongest positive correlation), with zero representing the absence of correlation between variables.

In this work, a positive correlation indicates that the two invariants "order" the isomers approximately the same way, whereas a negative correlation indicates that the isomers are ordered more or less inversely by the two invariants. Values closer to 1 (or −1) indicate that the two invariants *agree* among themselves, and values closer to 0 show the opposite. Note that negative correlations are expected, since in some invariants the objective is to minimize the value, while in others the objective is to maximize it. What matters, more than the sign, is that the correlation is far from zero.

The following pages contain 28 graphs, which can be seen in Figures 1 through 28, one for each pair of invariants. In each graph, the X axis corresponds to the value of n and is limited to the range $[20, 160]$, while the Y axis corresponds to Spear-

man's coefficient. The points, obviously, correspond to the coefficients between the two invariants, calculated for each even value of n in the range with the exception of the special cases mentioned before (independence number, Fries number, and Taylor number), of the values $20, 24$, and 26 (for which there are only one isomer) and of the value 22 (since there are no fullerenes with 22 vertices).

First, it is interesting to note the relationship between the Fowler-Manolopoulos criterion, appointed by the literature as the most promising invariant, with other invariants. In each of the graphs of Figures 1 to 7 there is a similar pattern, characterized by a medium correlation for values of n up to 80 or 100, followed by a few alternating highs and lows. This correlation is particularly close to zero for the Fries and Taylor numbers, whereas for the diameter, the Wiener index, and the independence number the value is closer to 0.5, indicating a good degree of agreement.

Second, in at least five of those graphs (Figures 1, 3, 4, 6, 7), the lowest and highest points seem to repeat themselves at the same values of n $(100, 120, 140, 160)$. Although these limited observations are not enough to determine the existence of an alternation for multiples of 20 above 100 (especially for the Fries, Taylor, and independence numbers), there are strong clues of a tendency. If this was confirmed, it could be an indication of characteristics particularly prone to stability on these isomers multiple of 20 − something equally detectable by all invariants.

A third case deserving attention is that of the bipartite edge frustration and the independence number, which, at least according to the potential to filter isomers known to be stable, seem to be the most promising invariants. On Figure 19 it is shown that the correlation between them is not particularly strong for small values of n. However, as n approaches and surpasses 100, the situation changes and the correlation jumps to 1. This nearly complete agreement maintains itself at least up to $n = 130$, when the dots lower again. Perhaps with more observations in this range it would be possible to see clearly some highs and lows just like on the Fowler-Manolopoulos criterion.

The diameter, even being the least relevant invariant in practice, shows a positive correlation with most other invariants. In at least three cases (all for n above

100) this correlation is significant: on Figures 9 and 10 (bipartite edge frustration and independence number), and on Figure 11 (number of Kekulé structures, although with a negative correlation). As a matter of fact, with the exception of the Wiener index (which, generally, shows medium or weak correlations) and, partially, of the Fowler-Manolopoulos criterion, all invariants seem to agree rather strongly for the range of n starting at approximately 100 and going up to 130, with coefficients very close to 1 or -1.

Finally, Figure 28 illustrates the case of the Fries and Taylor numbers, in which the correlation is consistently positive. Something expected, since one is nothing more than a modified version of the other. This similarity is reinforced by the fact that all other invariants correlate in almost exactly the same way with both, as can be seen on Figures 6, 7, 12, 13, 17, 18, 21, 22, 24, 25, 26, and 27.
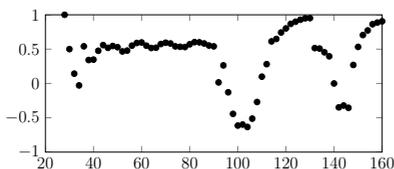
Figure 1. Spearman's Coefficient between the Fowler-Manolopoulos Criterion and the Diameter.
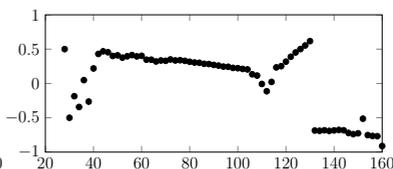
Figure 2. Spearman's Coefficient between the Fowler-Manolopoulos Criterion and the Wiener Index.
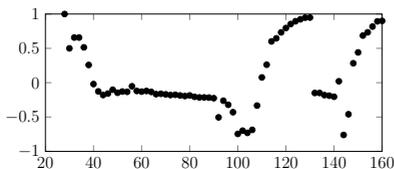
Figure 3. Spearman's Coefficient between the Fowler-Manolopoulos Criterion and the Bipartite Edge Frustration.
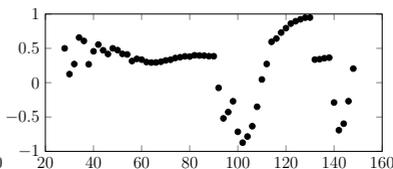
Figure 4. Spearman's Coefficient between the Fowler-Manolopoulos Criterion and the Independence Number∗.
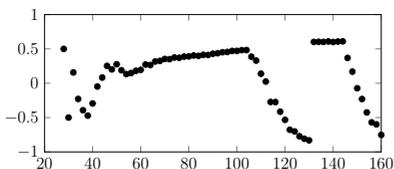
Figure 5. Spearman's Coefficient between the Fowler-Manolopoulos Criterion and the Number of Kekulé Structures.
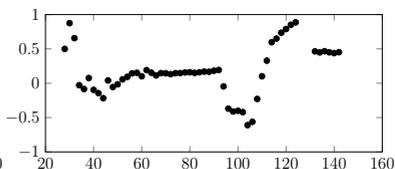
Figure 6. Spearman's Coefficient between the Fowler-Manolopoulos Criterion and the Fries Number∗.
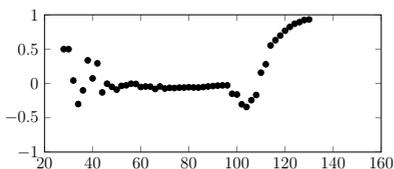
Figure 7. Spearman's Coefficient between the Fowler-Manolopoulos Criterion and the Taylor Number∗.
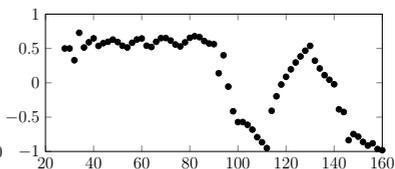
Figure 8. Spearman's Coefficient between the Diameter and the Wiener Index.
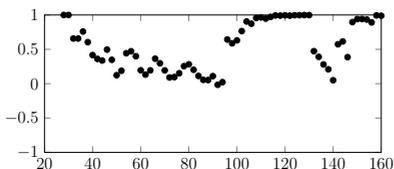
Figure 9. Spearman's Coefficient between the Diameter and the Bipartite Edge Frustration.
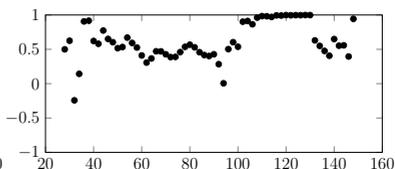
Figure 10. Spearman's Coefficient between the Diameter and the Independence Number∗.
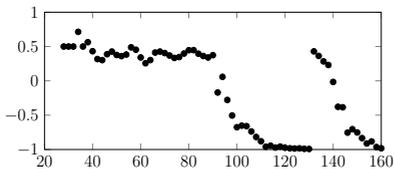
Figure 11. Spearman's Coefficient between the Diameter and the Number of Kekulé Structures.
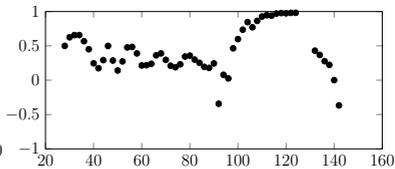
Figure 12. Spearman's Coefficient between the Diameter and the Fries Number∗.
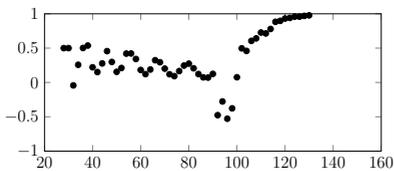
Figure 13. Spearman's Coefficient between the Diameter and the Taylor Number*.
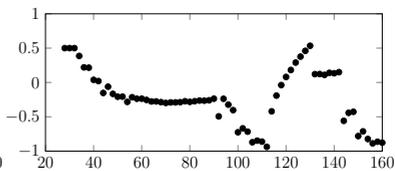
Figure 14. Spearman's Coefficient between the Wiener Index and the Bipartite Edge Frustration.
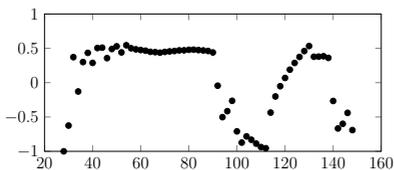
Figure 15. Spearman's Coefficient between the Wiener Index and the Independence Number*.
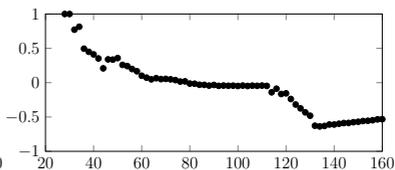
Figure 16. Spearman's Coefficient between the Wiener Index and the Number of Kekulé Structures.
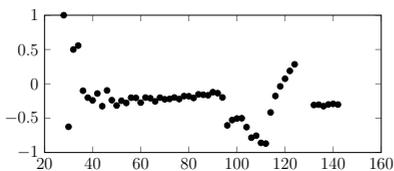
Figure 17. Spearman's Coefficient between the Wiener Index and the Fries Number*.
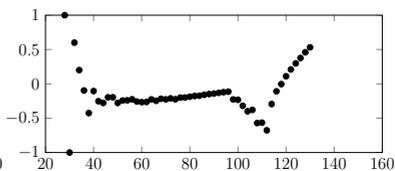
Figure 18. Spearman's Coefficient between the Wiener Index and the Taylor Number*.
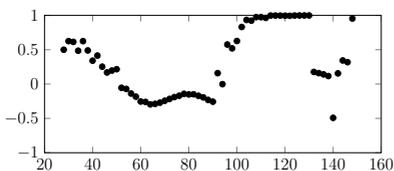
Figure 19. Spearman's Coefficient between the Bipartite Edge Frustration and the Independence Number*.
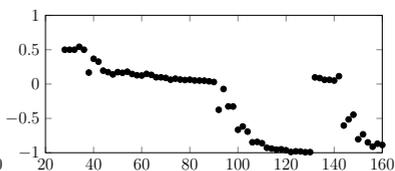
Figure 20. Spearman's Coefficient between the Bipartite Edge Frustration and the Number of Kekulé Strucutres.
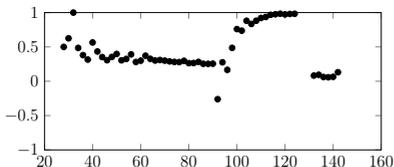
Figure 21. Spearman's Coefficient between the Bipartite Edge Frustration and the Fries Number∗.
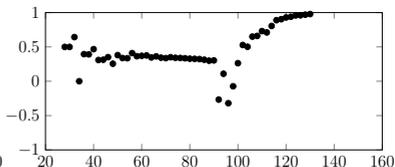
Figure 22. Spearman's Coefficient between the Bipartite Edge Frustration and the Taylor Number∗.
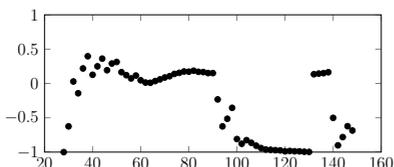
Figure 23. Spearman's Coefficient between the Independence Number∗ and the Number of Kekulé Structures.
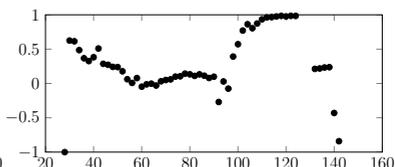
Figure 24. Spearman's Coefficient between the Independence Number and the Fries Number∗.
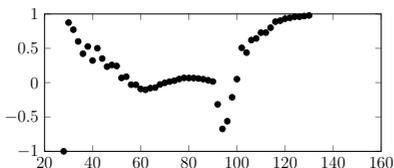
Figure 25. Spearman's Coefficient between the Independence Number and the Taylor Number∗.
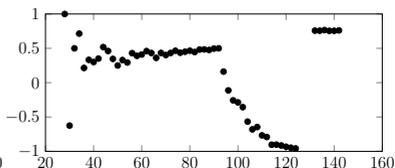
Figure 26. Spearman's Coefficient between the Number of Kekulé Structures and the Fries Number∗.
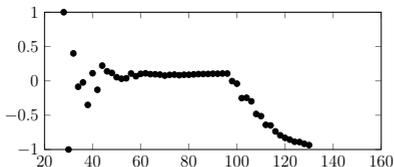
Figure 27. Spearman's Coefficient between the Number of Kekulé Structures and the Taylor Number∗.
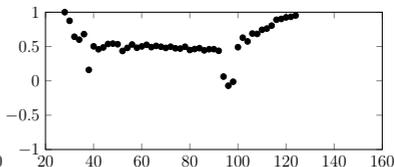
Figure 28. Spearman's Coefficient between the Fries Number and the Taylor Number∗.

5 Final Remarks and Future Work

This paper has presented a comparison of several graph invariants in the context of predicting the stability of fullerene molecules, which can be represented using the fullerene graphs defined in Section 2. The invariants in question were selected considering their relevance in the literature, their alleged potential for predicting stability, and the difficulty involved in their implementation.

For each one of these invariants, after careful study and consideration of available information in the literature, an algorithm to calculate it was implemented and executed for each fullerene isomer with up to 130 vertices, and for each IPR fullerene with at least 132 and up to 160 vertices (the three exceptions being the independence number, the Fries number, and the Taylor number, for which only part of this range was completed). Although most invariants mentioned before have already been studied in such a context, as far as the authors know this is the first comparative study taking all these invariants into account.

In order to extract useful conclusions from all these values, three approaches were chosen. The first is comparing the values with those calculated in other works, which serves as an independent verification of the correctness of the written programs and gives a greater degree of confidence to the data. The second was evaluating the performance of the invariants in predicting the stability of fullerenes already known to have been previously observed in laboratory. Two invariants were clearly more successful than the others in this task: the bipartite edge frustration and the independence number. The third approach was to perform a statistical analysis of the calculated values for each invariant, quantifying how much they agree or disagree among themselves through Spearman's rank correlation coefficient.

Three conclusions in particular deserve special attention: i) almost all invariants seem to agree rather strongly for values of n between 100 and 130, ii) the Fowler-Manolopoulos criterion, initially having an average correlation with other invariants in the range of n between 80 and 100, shows a singular behavior for values of n greater than 100, with a sequence of highs and lows that seems to

repeat itself always at the same multiples of 20, and iii) despite indications in literature that the independence number is not strongly related to stability, this invariant was capable of isolating with high precision all isomers known to be stable − strong evidence that it is more specialized version, the closed-shell independence number [13, 14] (not analyzed here), could perform even better.

Finally, all that remains is to mention some possible extensions of this work. The most direct continuations would be filling up the "gaps" left in the independence number, the Fries number, and the Taylor number (calculating their values as initially planned, to all isomers with up to 160 vertices) or expanding the limits even further − for example, considering all IPR isomers with up to 200 vertices. Another alternative would be the inclusion of other invariants, which could not be included here for lack of time. Such is the case of the closed-shell independence number [13, 14], the *resistance distance* (similar to the Wiener index, but based on an analogy with electricity, considering each edge of the fullerene graph as a wire with unit resistance [22]), the number of spanning trees (a concept also known as *complexity*, which is inversely proportional to the number of adjacencies between pentagons and, consequently, proportional to the relative stability [23]), the topological descriptor Ψ from Réti and László (based in the traditional pentagon indices, adding the new concept of *pentagon arm indices* [54]), the smallest eigenvalue of the distance matrix [3, 24, 34], the saturation number [3, 15, 16], the Cheeger constant [29, 44], and many others [2, 32, 58, 61].

References

[1] T. Achterberg, SCIP: Solving constraint integer programs, *Math. Prog. Comput.* **1** (2009) 1–41.

[2] M. Alcami, G. Sachez, S. Diaz−Tendero, Y. Wang, F. Martin, Structural patterns in fullerenes showing adjacent pentagons: C_{20} to C_{72}, *J. Nanosci. Nanotech.* **7** (2007) 1329–1338.

[3] V. Andova, T. Došlić, M. Krnc, B. Lužar, R. Škrekovski, On the diameter and some related invariants of fullerene graphs, *MATCH Commun. Math. Comput. Chem.* **68** (2012) 109–130.

[4] S. J. Austin, P. W. Fowler, P. Hansen, D. E. Manolopoulos, M. Zheng, Fullerene isomers of C_{60}. Kekulé counts versus stability, *Chem. Phys. Lett.* **228** (1994) 478–484.

[5] M. Berkelaar, K. Eikland, P. Notebaert, lp_solve 5.5, open source (mixed-integer) linear programming system, Software, 2004, Available at http://lpsolve.sourceforge.net/5.5/.

[6] D. Bonchev, N. Trinajstić, Information theory, distance matrix, and molecular branching, *J. Chem. Phys.* **67** (1977) 4517–4533.

[7] G. Brinkman, J. Goedgebeur, B. D. McKay, Generation of fullerenes, *J. Chem. Inf. Model.* **52** (2012) 2910–2918.

[8] G. Brinkman, J. Goedgebeur, H. Mélot, K. Coolsaet, House of graphs: a database of interesting graphs, *Discr. Appl. Math.* **161** (2013) 311–314. Available at http://hog.grinvin.org.

[9] G. Brinkman, S. Greenberg, C. Greenhil, B. D. McKay, R. Thomas, P. Wollan, Generation of simples quadrangulations of the sphere, *Discr. Math.* **305** (2005) 33–54.

[10] G. Brinkman, B. D. McKay, Construction of planar triangulations with minimum degree 5, *Discr. Math.* **301** (2005) 147–163.

[11] G. Brinkman, B. D. McKay, Fast generation of planar graphs, *MATCH Commun. Math. Comput. Chem.* **58** (2007) 323–357.

[12] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, 2009.

[13] S. Daugherty, W. Myrvold, P. W. Fowler, Backtracking to compute the closed–shell independence number of a fullerene, *MATCH Commun. Math. Comput. Chem.* **58** (2007) 385–401.

[14] S. M. Daugherty, *Independent Sets and Closed–Shell Independent Sets of Fullerenes*, PhD thesis, Univ. Victoria, 2009.

[15] T. Došlić, On some structural properties of fullerene graphs, *J. Math. Chem.* **31** (2002) 187–195.

[16] T. Došlić, Saturation number of fullerene graphs, *J. Math. Chem.* **43** (2008) 647–657.

[17] T. Došlić, D. Vukičević, Computing the bipartite edge frustration of fullerene graphs, *Discr. Appl. Math.* **155** (2007) 1294–1301.

[18] L. Euler, Demonstratio nonnullarum insignium proprietatum, quibus solida hedris planis inclusa sunt praedita, *Novi Commentarii Academiae Scientiarum Petropolitanae* **4** (1758) 140–160.

[19] S. Fajtlowicz, On conjectures of graffiti, *Discr. Math.* **72** (1988) 113–118.

[20] S. Fajtlowicz, *On conjecture of Graffiti V*, Western Michigan Univ., Kalamazoo, 1995, pp. 367–376.

[21] S. Fajtlowicz, C. E. Larson, Graph–theoretic independence as a predictor of fullerene stability, *Chem. Phys. Lett.* **377** (2003) 485–490.

[22] P. W. Fowler, Resistance distances in fullerene graphs, *Croat. Chem. Acta* **75** (2002) 401–408.

[23] P. W. Fowler, Complexity, spanning trees and relative energies in fullerene isomers, *MATCH Commun. Math. Comput. Chem.* **48** (2003) 87–96.

[24] P. W. Fowler, A note on the smallest eigenvalue of fullerenes, *MATCH Commun. Math. Comput. Chem.* **48** (2003) 37–48.

[25] P. W. Fowler, G. Caporossi, P. Hansen, Distance matrices, Wiener indices and related invariants of fullerenes, *J. Phys. Chem. A* **105** (2001) 6232–6242.

[26] P. W. Fowler, S. Daugherty, W. Myrvold, Independence number and fullerene stability, *Chem. Phys. Lett.* **448** (2007) 75–82.

[27] P. W. Fowler, D. E. Manolopoulos, *An Atlas of Fullerenes*, Dover Pub., Mineola, 2006.

[28] P. W. Fowler, W. Myrvold, Some chemical applications of independent sets, in: P. Hansen, N. Mladenović, J. A. M. Perez, B. M. Battista, J. M. Moreno-Vega (Eds.), *Proceedings of the 18th Mini Euro Conference on Variable Neighborhood Search*, Tenerife, Spain, 2005, pp. 16–37.

[29] P. W. Fowler, K. M. Rogers, S. Fajtlowicz, P. Hansen, G. Caporossi, Facts and conjectures about fullerene graphs: Leapfrog, cylinder and Ramanujan fullerenes, in: A. Betten, A. Kohnert, R. Laue, A. Wassermann (Eds.), *Algebraic Combinatorics and Applications*, Springer, Berlin, 2000, pp. 134–146.

[30] K. Fries, Über bicyclische verbindungen und ihren vergleich mit dem naphtalin, *Liebigs Ann.* **454** (1927) 121–324.

[31] K. Fries, R. Walter, K. Schilling, Über tricyclische verbindungen, in denen naphtalin mit einem heterocyclus anelliert ist, *Liebigs Ann.* **516** (1935) 248–285.

[32] L. H. Gan, J. Liu, Q. Hui, S. Q. Shao, Z. H. Liu, General geometric rule for stability of carbon polyhedra, *Chem. Phys. Lett.* **472** (2009) 224–227.

[33] M. R. Garey, D. S. Johson, *Computers and Intractability; A Guide to the Theory of NP-Completeness*, Freeman, New York, 1990.

[34] C. Godsil, G. Royle, *Algebraic Graph Theory*, Springer, New York, 2001.

[35] B. Grünbaum, T. S. Motzkin, The number of hexagons and the simplicity of geodesics on certain polyhedra, *Can. J. Math.* **15** (1963) 744–751.

[36] A. L. P. Guedes, Representação de variedades combinatórias de dimensão n, Master's thesis, COPPE/Engenharia de Sistemas e Computação - UFRJ, Rio de janeiro, RJ, Brazil, 1995.

[37] Eigen v3, http://eigen.tuxfamily.org, 2010.

[38] L. Guibas, J. Stolfi, Primitives for the manipulation of general subdivisions and the computation of Voronoi diagrams, *ACM Trans. Graph.* **4** (1985) 74–123.

[39] I. Gutman, Y. N. Yeh, S. L. Lee, Y. L. Luo, Some recent results in the theory of the Wiener number, *Indian J. Chem.* **32A** (1993) 651–661.

[40] P. Hansen, M. Zheng, Numerical bounds for the perfect matching vectors of a polyhex, *J. Chem. Inf. Comp. Sci.* **34** (1994) 305–308.

[41] R. V. Hogg, J. McKean, A. T. Craig, *Introduction to Mathematical Statistics*, Pearson, Harlow, 2012.

[42] P. Holme, F. Liljeros, G. R. Edling, B. J. Kim, Network bipartivity, *Phys. Rev. E* **68** (2003) #056107.

[43] H. Hosoya, A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, *Bull. Chem. Soc. Jpn.* **44** (1971) 2332–2339.

[44] C. Justus, *Boundaries of Triangle–Patches and the Expander Constant of Fullerenes*, PhD thesis, Univ. Bielefeld, 2007.

[45] P. Kasteleyn, The statistics of dimers on a lattice: I. The number of dimer arrangements on a quadratic lattice, *Physica* **27** (1961) 1209–1225.

[46] P. W. Kasteleyn, Dimer statistics and phase transitions, *J. Math. Phys.* **4** (1963) 287–293.

[47] P. W. Kasteleyn, Graph theory and crystal physics, in: F. Harary (Ed.), *Graph Theory and Theoretical Physics*, Acad. Press, New York, 1967, pp. 43–110.

[48] V. Kolmogorov, Blossom V: a new implementation of a minimum cost perfect matching algorithm, *Math. Prog. Comput.* **1** (2009) 43–67.

[49] J. Konc, D. Janežič, An improved branch and bound algorithm for the maximum clique problem, *MATCH Commun. Math. Comput. Chem.* **58** (2007) 569–590.

[50] W. Krätschmer, L. D. Lamb, K. Fostiropoulos, D. R. Huffman, Solid C_{60}: a new form of carbon, *Nature* **347** (1990) 354–358.

[51] H. W. Kroto, The stability of the fullerenes C_n, with $n = 24, 28, 32, 36, 50, 60$ and 70, *Nature* **329** (1987) 529–531.

[52] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, R. E. Smalley, C_{60}: Buckminsterfullerene, *Nature* **318** (1985) 162–163.

[53] K. Raghavachari, Ground state of C_{84}: Two almost isoenergetic isomers, *Chem. Phys. Lett.* **190** (1992) 397–400.

[54] T. Réti, I. László, On the combinatorial characterization of fullerene graphs, *Acta Polytech. Hung.* **6** (2000) 85–93.

[55] D. H. Rouvray, The modeling of chemical phenomena using topological indices, *J. Comput. Chem.* **8** (1987) 470–480.

[56] D. H. Rouvray, The rich legacy of half a century of the Wiener index, in: D. H. Rouvray, R. B. King (Eds.), *Topology in Chemistry: Discrete Mathematics of Molecules*, Horwood Pub., Chichester, 2002, pp. 16–37.

[57] C. H. Sah, A generalized leapfrog for fullerene structures, *Fullerene Sci. Tech.* **2** (1994) 445–458.

[58] S. Schein, M. Sands–Kidner, A geometric principle may guide self assembly of fullerene cages from Clathrin Triskella and from carbon atoms, *Biophys. J.* **94** (2008) 958–976.

[59] T. G. Schmalz, W. A. Seitz, D. J. Klein, G. E. Hite, C_{60} carbon cages, *Chem. Phys. Lett.* **130** (1986) 203–207.

[60] T. G. Schmalz, W. A. Seitz, D. J. Klein, G. E. Hite, Elemental carbon cages, *J. Am. Chem. Soc.* **110** (1988) 1113–1127.

[61] Z. Slanina, F. Uhlk, S. L. Lee, E. Ōsawa, Geometrical and thermodynamic approaches to the relative stabilities of fullerene isomers, *MATCH Commun. Math. Comput. Chem.* **44** (2001) 335–348.

[62] R. Taylor, Rationalisation of the most stable isomer of a fullerene C_n, *J. Chem. Soc. Perkin Trans. 2* **1** (1992) 3–4.

[63] H. N. V. Temperley, M. E. Fisher, Dimer problem in statistical mechanics-an exact result, *Philos. Mag.* **6** (1961) 1061–1063.

[64] F. Torrens, Computing the permanent of the adjacency matrix for fullerenes, *Int. El. J. Mol. Des.* **1** (2002) 351–359.

[65] V. V. Vazirani, NC algorithms for computing the number of perfect matchings in k3,3-free graphs and related problems, *Inf. Comput.* **80** (1989) 152–164.

[66] H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.