Comparative Analysis of Methods for Obtaining Realistic Graphs from Kekule Cells

Aaron Germuth and Alex Aravind

Department of Computer Science
University of Northern British Columbia
E-mail: (germuth,csalex)@unbc.ca

February 25, 2015

Abstract

This is an abstract

1 Introduction

A certain class of molecules, the aromatic hydrocarbons, have been postulated to be able to conduct electricity. More specifically, alternating paths of single and double bonds resemble paths the electricity may follow. Upon conduction with a soliton, the single and double bonds are reversed, effectively opening and closing various paths. This switching behaviour has been described as a potential for molecular electronics.

Kekule Theory was introduced to systematize and classify this possibility. Each unique switching behaviour was given a classification. Hesselink proceeded in discovering a method for obtaining graphs for some classification. Each graph resembles an aromatic hydrocarbon, with nodes as atoms, and edges and bonds. This was the first attempt at a mapping from Kekule cell to graph. However, Hesselinks graphs made no attempt to resemble carbon chemistry.

We present and compare multiple methods for mapping a Kekule cell to a graph. However, unlike previous work, we only accept graphs which resemble realistic molecules according to a set of imposed restrictions. Section 2 outlines the rationale behind each restriction used. Section 3 outlines each of our methods for mapping a Kekule cell to a graph. Section 4 compares the performance of each method and concludes.

1.1 Overview

Section 2 consists of the restrictions used to obtain more realistic graphs, and the reasoning behind them. Section 3 explains the genetic algorithm, and how certain genetic operations are applied on graphs. It also discusses graph alterations used. Section 4 presents the results of the genetic algorithms and assesses its validity. This includes how realistic the obtained graphs are, and whether these structures can be synthesized chemically.

2 Restrictions

2.1 Degree

Carbon is tetravalent, meaning it has 4 valence electrons available to form covalent chemical bonds. This means (in most cases), carbon is only stable when bonded with 4 bonding electrons, in order to complete its octet. These 4 electrons may be provided from 4 distinct molecules, or there could be one or more molecules providing multiple electrons.

In π conjugated systems, every carbon atom contains precisely one double bond. Double bonds hold twice the amount of electrons and therefore each carbon can be at most connected to 3 distinct atoms. This is seen in all polycyclic polyunsaturated hydrocarbons. In this way, each node in our graphs has a maximum degree of 3.

The situation is slightly more complex for ports. Ports are not required to contain a double bond, and different resonance structures of the same graph may or may not contain double bonds at every port. However, all cells in Hesselink's classification, are 'flexible'. Flexible in this context, means that every port in the graph has atleast one port assignment where that port contains a double bond. And since any cell can be translated / port renamed to obtain one of the classified cells, it is safe to assume every port will contain a double bond in atleast one resonance structure, limiting ports to a degree of 3 as well.

Ports must also be able to interact with the outside world, by definition. If we assume this connection is through molecular wires, as in [9], or through an omniconjugated molecule [3], then each port must hold back one of its valence electrons to be able to connect to atoms outside the molecule. For this reason, we restrict ports to have a maximum degree of 2.

2.2 Cycles

When cycles within graphs are created, the concept of aromaticity is introduced. In Chemistry, conjugated ring formations are called aromatic, and the molecule must meet certain criteria or the ring will not be stable. This means any cycles formed within our graphs must also meet these criteria if they are to be considered stable. Aromatic molecules show increased stability compared to normal non-aromatic compounds. Anti-aromatic rings show decreased stability. In order to be considered aromatic a molecule must have:

1. A delocalized conjugated π system, resulting from alternating single and double bonds
2. Coplanar stucture, where all atoms of the ring are contained in the same plane
3. Contributing atoms arranged in one or more cycles
4. The number of contributing π delocalized electrons must be even, but not a multiple of 4 (4n + 2, where n = 0,1,2,...)

All of our graphs will by default meet the first condition. The second condition is difficult to determine. Aromatic rings are planar because it is comprised of multiple conjugated bonds. Each π conjugated bond results from the overlap in the nodal planes of each bonding atom. This creates a cloud of electron probability above and below the covalent bond. Each cloud generates a small electromagnetic field, which works to keep the ring flat (need cite here). Often non-aromatic molecules will align themselves in one plane to achieve aromatic status, where as anti-aromatic molecules will dis-align themselves to avoid anti-aromaticity (making them non-aromatic). A consequence of planarity is that any graph of the molecule should be able to visualized such that no edges are intersecting. (Currently we don't 100% detect whether graphs are planar or not).

2.2.1 Angle Strain

One factor attributing to whether a molecule can adopt a planar configuration is angle strain. When carbon contains only one double bond and two additional single bonds, its preferred conformation is trigonal planar. In this configuration, each atom is 120 degrees from each other and contained in the same plane. Smaller rings such as cyclobutadiene ([4] annulene) force angles of 90 between atoms if they are to remain planar. For this reason, cyclobutadiene adopts a non-planar configuration and is not aromatic. Larger rings can face the same difficulties, for example cyclodecapentene ([10] annulene) can't adopt a planar configuration as the bond angles would be 144 degrees. For this reason, cyclodecapentene is also not considered aromatic. Some larger rings such as [18] annulene can adopt a aromatic conformation, but rings of such sizes are quite rare. We limit cycles to size 5-7.

Cycles of length 5 and 7 have been shown to be stable in compounds such as azulene (see Figure 2). Cycles of length 6 represent benzene, and are generally accepted as the main contributor to aromatic rings. Any cycles of other sizes which appear can usually be artificially altered, through cycle expansion or contraction (see Section 3.4).

2.2.2 Cycle Connectivity

In regards to the third condition, groups of rings must also be considered. Instead of a single ring being aromatic, a group of rings may be deemed aromatic. In most ring systems, any two rings of a system share at most two atoms. This preserves the geometry of each cycle individually. This is seen in naphthalene, both hexagons share two common atoms. Even in complex polycyclic molecules such as pyracylene (see Figure 1), each cycle shares at most two nodes with any other cycle. Two cycles cannot share a single node (spiro-compound) as the shared node would have a degree of 4. Notice in Figure 1, the unrealistic compound has two cycles which share 4 nodes. To account for this, all cycles of each graph are found, and their intersections are computed. Unrealistic graphs are discarded.

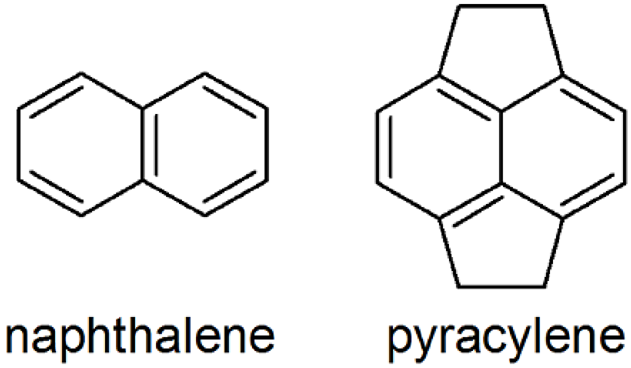


Figure 1: Napthalene has two benzene rings with 2 carbon atoms is common. Even in complex polycylic hydrocarbons such as pyracylene, every ring shares only 2 carbon atoms.

2.2.3 Huckel's Rule

The fourth condition is the most difficult. Due to the filling of quantum-mechanical electron orbitals, only aromatic molecules with 4n+2 electrons exhibit aromaticity. Compounds with 4n electrons exhibit anti-aromaticity. Groups of cycles add their delocalized electrons together to be one single π conjugated system (see Figure 2). This condition alone implies we must limit all ring systems to size of length 2, 6, or 10. However, Huckel's rule is mainly meant for monocyclic systems and has been shown to be inaccurate for many compounds containing more than three fused aromatic rings (see Figure 3) [25]. To account for this, in polycyclic systems, certain extensions to the rule have been added to better predict aromaticity [26].

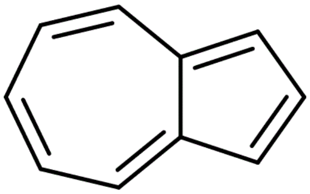


Figure 2: Azulene, a structure composed of an aromatic heptene and pentene ring attached together. The total amount of π contributing electrons is 10, 7 from heptene and 3 from pentene. Since 10 = 4n + 2 (where n = 2), Huckel's rule correctly predicts this structure to be aromatic.

We choose to ignore Huckel's rule and any extensions. This is because even anti-aromatic compounds such as pentalene (8 contributing electrons) can be stabilized by adding the appropriate substituent. Either electron-donating or electron-withdrawing groups can be attached to the ring

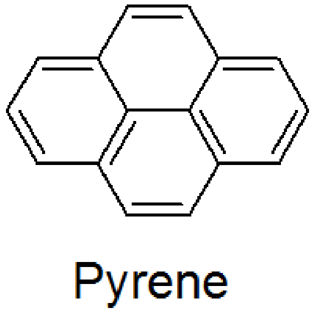


Figure 3: Although Pyrene has 16 conjugated electrons (which does not satisfy 4n + 2), it is still an aromatic molecule.

to allow aromaticity (see Figure 4). Additionally, it is possible to insert hereroatoms into the structure, such as nitrogen and sulphur. In addition to possibly making the molecules easier to synthesize, the nitrogen may add electrons to the conjugated system while leaving the switching behavior undisturbed [3] (see Figure, not existant yet

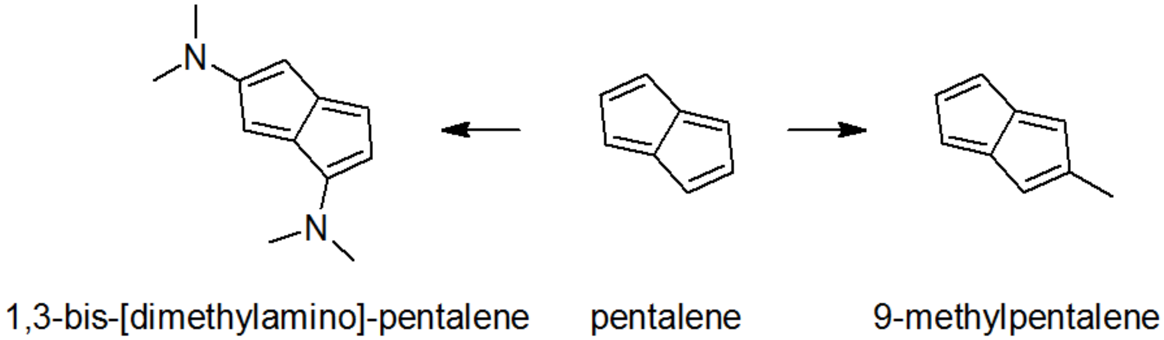


Figure 4: Pentalene can attach either the electron-donating Me [27] or $N(Me)_2$ [28] group to achieve aromatic structure.

Despite ignoring Huckel's rule, we cannot discard this problem completely. Since ports do not need to contain a double bond, in some resonance structures of the same molecule, the amount of contributing electrons may be different (see Figure 5 and 6). Unless every resonance structure is aromatic, or at least every structure can be made aromatic by the same substituent, the molecule will not be able to realize all port assignments of the cell and is therefore unfit.

To account for this, we must ensure that each ring contains the same amount of electrons in every resonance structure. It is a difficult problem to determine which port placements could cause this. We have yet to discover a general solution. In any ring system, a single port is allowed. Since a double bond spans two atoms, every atom in the ring will still always contain a double bond, leaving the total number of bonds to remain constant. However, when two ports are added into a ring, or two rings are conjoined, it becomes unclear. For example, see Figure 7. To ensure all results are manageable, all rings systems are restricted to one port. This means graphs such as napthalene, can only have a single port within the cycle. But if one graph contained two bridged

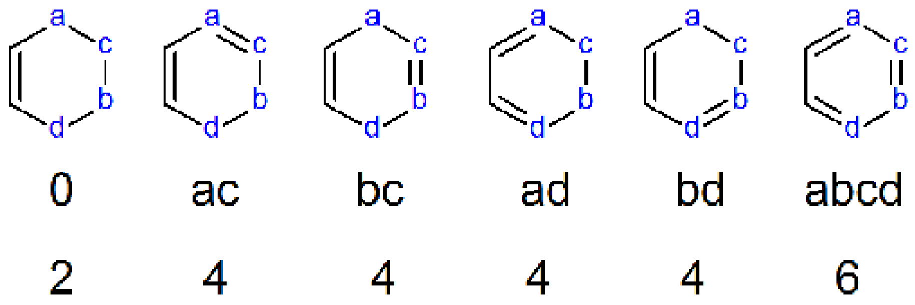


Figure 5: Below is a graph of benzene, which represents K5 of Hesselink's class 4 classification. The first line gives the structure, the second line the port assignment, and the third line the amount of electrons contributing to the ring. The problem here, is that only 2 and 6 are aromatic, and if we add substituents to stabilize the 4 electron structures, the 2 and 6 become unstable.

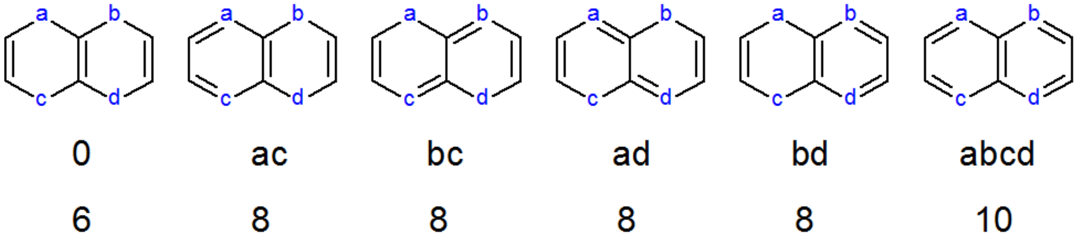


Figure 6: Napthalene also can represent K5 of Hesselink's class 4 classification. This graph has the same problem as before. In fact, most graphs with 2 or more ports within a cycle system have this problem.

napthalenes, each could contain a single port. This unnecessarily removes some graphs which could work, but it is a costly computation to determine for each graph whether it works. All other ports within the graph must be exo-cyclic.

Finally, graphs representing molecules must be connected. A disjoint graph more accurately represents two molecules rather than one. Although these restrictions do a good job of trimming out almost all unrealistic graphs, they are not absolute. Additionally, our restrictions make no attempt to ensure all graphs can be synthesized (more on this in Section 4).

3 Genetic Algorithm

A genetic algorithm is used to search for graphs which apply to a given cell. The algorithm begins by randomly generating a population of graphs. Each starting graph has anywhere from rank to 31 nodes, and rank - 1 to 20 edges. There must be at least rank - 1 edges in order to connect the the minimum amount of nodes together. Graphs with fitness 0 or lower (normally arises from disconnected ports) are deleted and a new graph will take its place. This makes up the first generation.

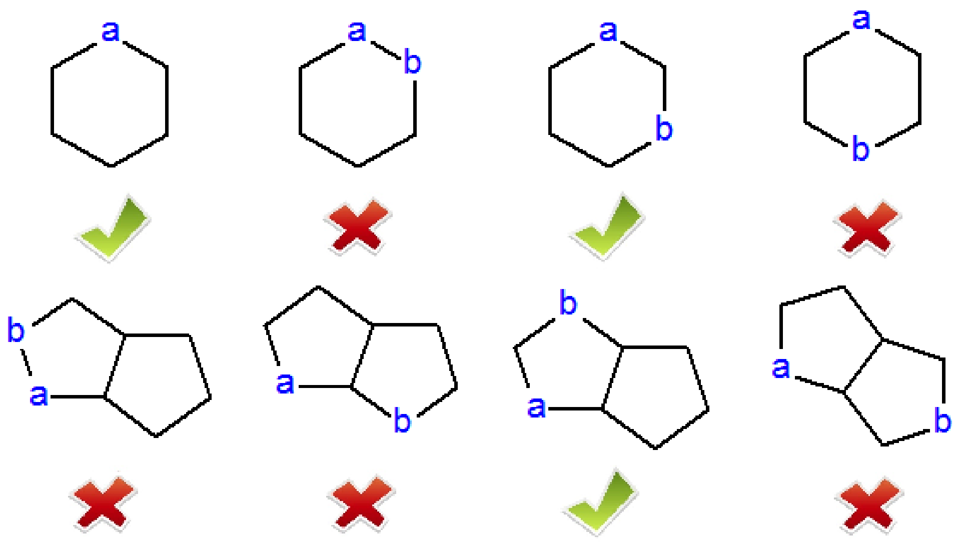


Figure 7: Benzene and Pentalene with two ports at various locations. Structures with a checkmark below indicate all resonance structures contain the same amount of double bonds. Structures with an x, indicate the opposite, and are unfit.

In each iteration, the survivors of last generation's population must be chosen. This is done by selecting the best performing graphs (fitness-wise), along with a smaller amount of random graphs. Random graphs are added to ensure genetic diversity in the population. This sub-population now undergoes genetic operations such as mutation and crossover.

The genetic algorithm has two terminating conditions. Either the maximum number of iterations has been reached, or some threshold of graphs which satisfy the cell has been reached. In some smaller cases, the genetic algorithm terminates immediately after generating the initial population, as it contained at least x graphs which satisfy the cell (where x is the threshold of graphs desired).

Resulting graphs are converted to SMILES (cite here) (small molecular input line entry system) in order to be visualized. Conversion first involves finding a spanning tree of the graph. All edges in the original graph which are not in the spanning tree must be edges which complete a cycle. Therefore each node on either side of the edge is labeled as a cycle point. We then preform depth first search over the spanning tree, and append each nodes label as we reach it.

The ports are labeled by atoms with a P (normally reserved for a phosphorous atom). We need not label each port individually since Kekulé cells are invariant across port renaming [1]. The open source Chemistry Development Kit (citation) is used to parse the SMILES and generate a two-dimensional structure. Other than the ports, the structure resembles the hydrogen-suppressed carbon skeleton of the molecule.

3.1 Mutation

Mutation involves randomly perturbing graphs to obtain new ones. We normally generate enough mutants so that it is likely each graph will be mutated more than once. Mutations include:

1. Addition of a node
2. Removal of a random node, and all edges to it
3. Addition of a random edge [1]
4. Removal of a random edge
5. Port Extension

Port extension involves adding a new node in the position of every port, and then adding the port connected to that node. This will always be viable since any port can have at most degree 2, and after port extension, the new nodes will have at most degree 3. This is often useful for spreading the ports out and allowing new edges without breaking our degree restrictions. Since each port will always have a degree of 1 after extension, each port can facilitate a new bond. See Figure 8.

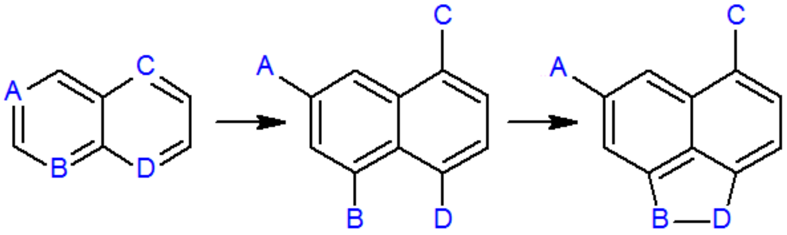


Figure 8: Napthalene (left) undergoing a port extension (center). Notice that after the port extension, it is now feasible to add an edge between B and D, which can then create Acenaphthene (right).

3.2 Crossover

Crossover begins by randomly selecting 2 graphs from the sub-population. These graphs are combined to create a new child graph. On average, each graph produces two children. Children receive all edges with both parents share, and have a 50% chance to get any other edge found in one of the parents. This creates children which are a hybrid of both their parents.

[1] Edges are randomly chosen until an edge is found that can be added without breaking our restrictions on the degree of vertices.

3.3 Fitness Function

Fitness values are integers given to a graph. Any graphs cell can be calculated using Hesselink's procedure here: (explain a lot). The fitness is calculated by iterating over the graphs cell compared to the cell we are evolving towards. For every port assignment they share, the fitness is incremented by one. For every port assignment that is missing, or is extra, fitness is decremented one point. If a cell is empty, with no port assignments [2], it is discarded.

3.4 Cycle Detection

Cycles are detected using an upper bounded implementation of the 'flooding' algorithm (need cite here). In the flooding algorithm, each node sends out a packet to all if it's neighbors. Any time a node receives a packet, it appends its name and sends the new packet to all of its neighbors. Packets are not sent back to the node who it was just received it from. If at any time a packet has been appended by the same node twice, we have found a cycle. As soon as a packet finds a cycle, the packet is discontinued. Packets of length > 12 are also discontinued. This introduces the error that a graph could contain a cycle of length 13. The amount of computation time saved is usually worth it, considering how unlikely a cycle of length 13 or greater is.

Some optimizations can be made considering the subset of all possible graphs we are dealing with. Packets only begin on nodes with degree of 3, as this is guaranteed to find all cycles. Every point within a cycle has a degree of 2, so if that cycle is connected to any other node, then at least one of that cycles nodes will have a degree of 3. If there are no nodes with degree of 3, the graph either contains no cycles, or the graph is one big cycle. Therefore, in this case, we only start a single packet at one random node. This is still slow.

3.5 Cycle Extension and Contraction

Cycle extension is done by inserting two additional carbon atoms into the cycle at the same position (see Figure 9). This has no effect on the Kekulé cell as the two atoms are conjugated with the rest of the ring and simply expand it. This procedure is nearly identical to operation v of [3]. Care must be taken to not select two nodes which also are a part of another cycle, as the resulting graph will contain unrealistic bonding distances and cycle connectivity. The opposite procedure can be used to contract cycles. Contraction is only possible if there are two neighboring nodes with no edges outside the ring. Graphs with un-alterable cycles outside our limits are excluded.

3.6 Connecting a Disjoint Graph

A disjoint graph is not an accurate representation of a molecule. Luckily, disjoint graphs can often be connected using the following procedure. If we add internal vertices which must always

[2] This can happen when the genetic algorithm creates a graph which has a (non-port) vertex with no edges to it (degree 0). There is no way for this graph to have a Kekulé structure since the lone vertex can never have a double bond.

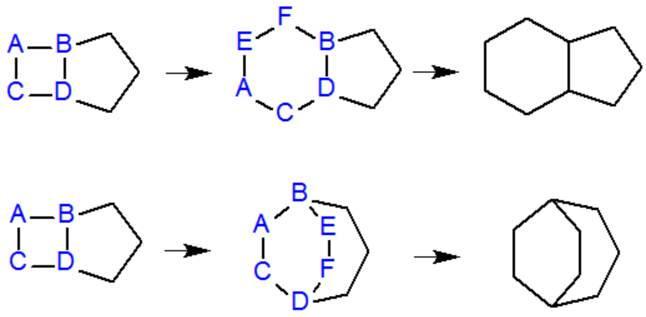


Figure 9: This shows the extension of the left 4 cycle into a proper 6 cycle. The top half shows a correct procedure, selecting to add nodes between A and B. The bottom half shows the incorrect procedure of adding two nodes in between B and D. Double bonds are excluded from this diagram for clarity.

conjugate within themselves, they will not interfere with the cell. For example, see Figure 10.

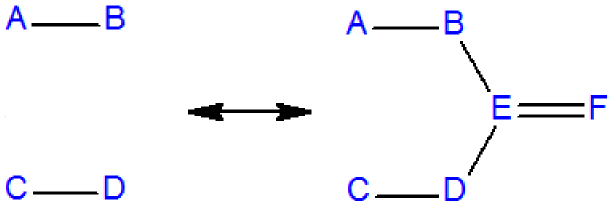


Figure 10: Process used to turn a disjoint graph to a connected graph with the same cell. Two new nodes E and F can be added which always conjugate to each other.

In the right-hand graph of Figure 10, F is not a port and so must always contain a double bond. The only other node which is available to form a double bond with F is E, so in every resonance structure of this molecule E and F will contain a double bond with each other (as shown in right diagram). Therefore d or b can never form a double bond with their new edge (that would result in two adjacent double bonds), and the Kekulé cell has not been changed. This now resembles a stable connected molecule. This procedure is similar to Operation iii and vi of [3].

3.7 Minor Implementation Details

3.7.1 The Bit Vector

A Bit Vector (or Bit array, bitmap, bitset, bitstring, etc) is an array data structure that compactly stores bits. It can be used to implement a simple set data structure. Although the Bit Vector is an array, the entire set can be contained with a single binary integer. Every 1 indicates the presence of something, and 0, the absence. For example, 0001 0101 (or 21) is a set of three elements. Node 1, node 3, and node 5.

Because the entire array is within an integer, it is very easy and computationally quick to perform certain operations. You can add or remove a node from your set by simple addition / subtraction. You can iterate through the set by starting at 1 and performing left bit shifts. You can use bitwise AND to take the intersection of two sets, or to check whether a set contains a given element (Figure 1). You can add two sets together with bitwise OR. You can even take the symmetric difference between two sets with XOR.

Whenever you can distinguish elements of your set with a single number, and your sets are not large enough to overflow an integer, you should consider using bit vectors. They are extremely compact, versatile, and quick.

In this application, bit vectors represent sets of nodes, as in the example above. They are also used to represent edges (a set of nodes with 2 elements) of graphs.

3.7.2 The Cell

A port assignment is the set of ports which have a double bond in a particular Kekulé structure. In this application, cells have two purposes.

1. A cell can hold an array of port assignments, one for every possible resonance structure of that molecule. It can be used to fully describe the electrical switching behavior of a molecule.
2. A cell can hold an array of edges. Each edge is represented using a Bit Vector. In this way, a cell con constitute an entire graph.

3.8 Additional Note

It should be noted that rather than using the genetic algorithm to approximate a solution (as seen in most applications), an answer generated by it is only useful if it found an actual solution. However, for all cells of rank <6 and most cells of rank 6, the genetic algorithm can converge to multiple solutions for a given cell.

4 Results

Most graphs produced by the algorithm can be feasibly realized in stable molecules. However, results are not perfect. As stated previously, rings of sizes other than 6 are not always stable.

Planarity? The second condition is another current drawback. Since the cycle detection algorithm needs an upper bound, in rare cases, cycles above this limit can appear. This can cause other smaller cycles which share more than two atoms with the large cycle to not be detected. Additionally, some bonding distances may be unrealistic. For examples of both, see Figure 11.

In most small cases (ports <6), the genetic algorithm is able to obtain atleast 10 different graphs which satisfy the cell immediately. For some of the later classifications of rank 5, and cells of rank

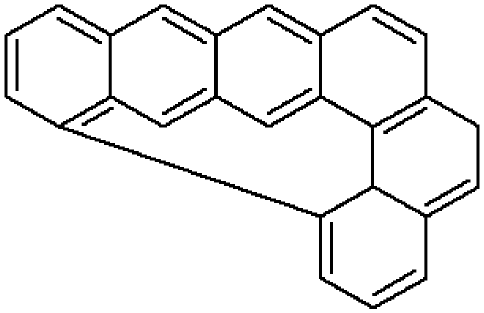


Figure 11: An example of an unrealistic molecule that could be generated by the genetic algorithm. There is a long stretched bond from the first benzene (top left) to the last benzene (bottom right). This stretched bond completes a cycle that has a length of 9. If the cycle detection algorithm had a limit of 8, this graph could be a result. Not only is the stretched bond unrealistic, but the cycle it creates shares up to 3 atoms with some rings.

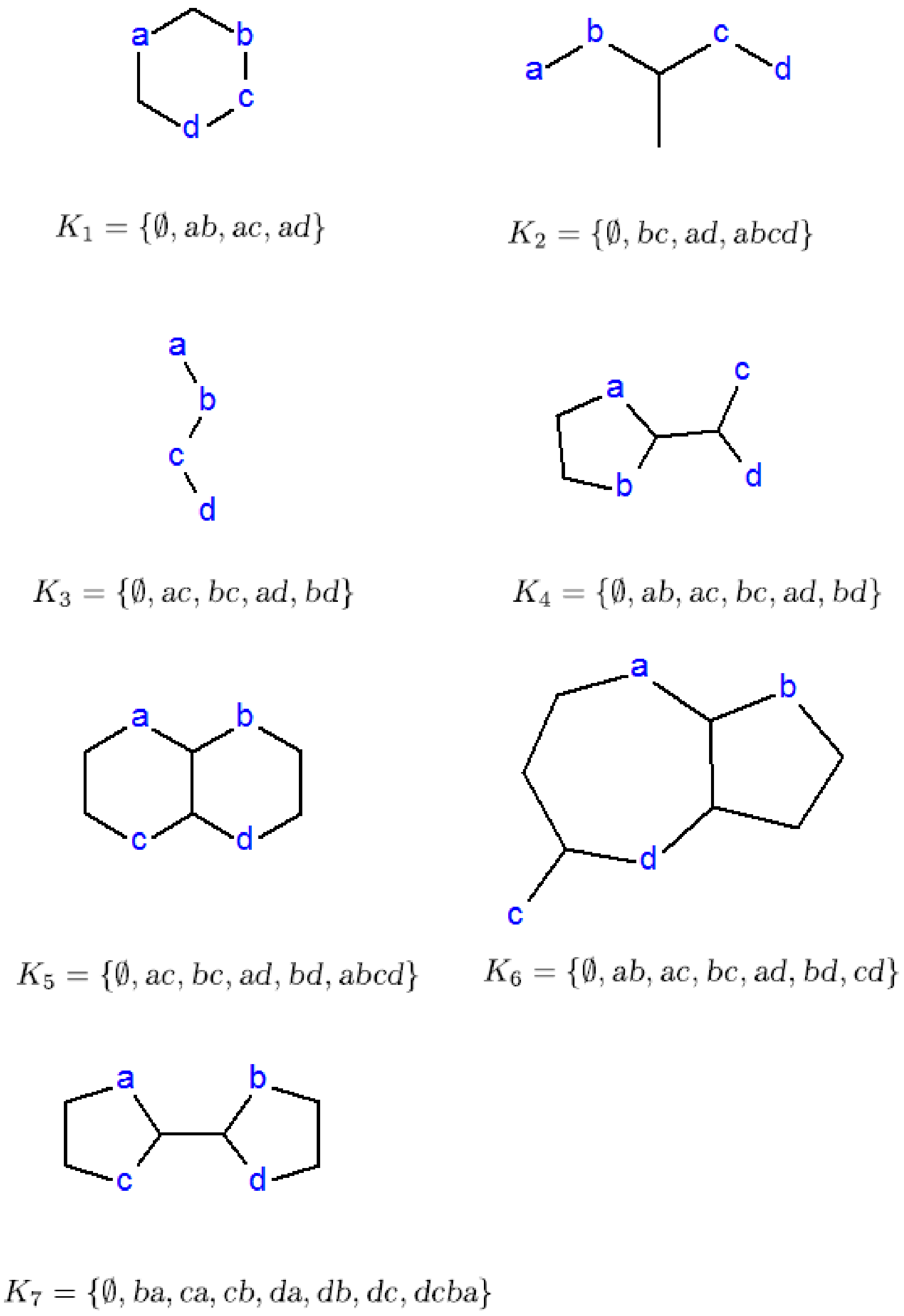
6 or higher, it can take longer. Some cells of rank 6 it is unable to find a graph for. As it is possible that no such realistic graph exists for some cells, it is difficult to determine whether a solution exists. The number of nodes is currently maxed at 31 nodes, as nodes are stored in a bit array, which would overflow the integer at 32.

Below is graphs for all cells of rank 4 and 5. Double bonds are excluded from the graphs shown below, as only one resonance structure could be seen at a time. Some graphs may have cells different than the title below them. This is because their cells can be port translated to achieve the classified cell. This is acceptable since switching behaviour is invariant across port translation [2]. As stated previously, the genetic algorithm is capable of coming to multiple solutions for each cell, and the ones shown below are simply ones selected by us. Notice the trend that as the Kekulé cell number increases, the Kekulé cell becomes longer. This means more resonance structures of each graph, and more connections between atoms. The higher amount of connectivity required in each cell makes it so the next graph is usually more difficult to find than the last one. Looking at Figures 13 and 14, it is easy to see that the graphs get bigger and more complex as the cell number increases.

Finding graphs for cells of higher ranks builds on this challenge. Not only is every graph larger, but there must be an additional atom with degree of 2. The number of combinations increases drastically, with 3 classes of rank 3, 7 classes of rank 4, 24 classes of rank 5, and 214 classes of rank 6. When attempting to find graphs for all cells of rank 6, we were able to find graphs for X out of 214. All graphs found can be viewed here: (website).

Increasing the number of nodes to above 32 will likely yield more results in some cases, however, it will lower performance. It is possible many graphs do not have a realistic graph which matches the cell.

Even if all graphs produced represent stable molecules, there are additional problems. Each hydrocarbon must be able to be easily synthesized for it to be of any use. Analysis on the synthesis of hydrocarbons here.



$K_1 = \{\emptyset, ab, ac, ad\}$

$K_2 = \{\emptyset, bc, ad, abcd\}$

$K_3 = \{\emptyset, ac, bc, ad, bd\}$

$K_4 = \{\emptyset, ab, ac, bc, ad, bd\}$

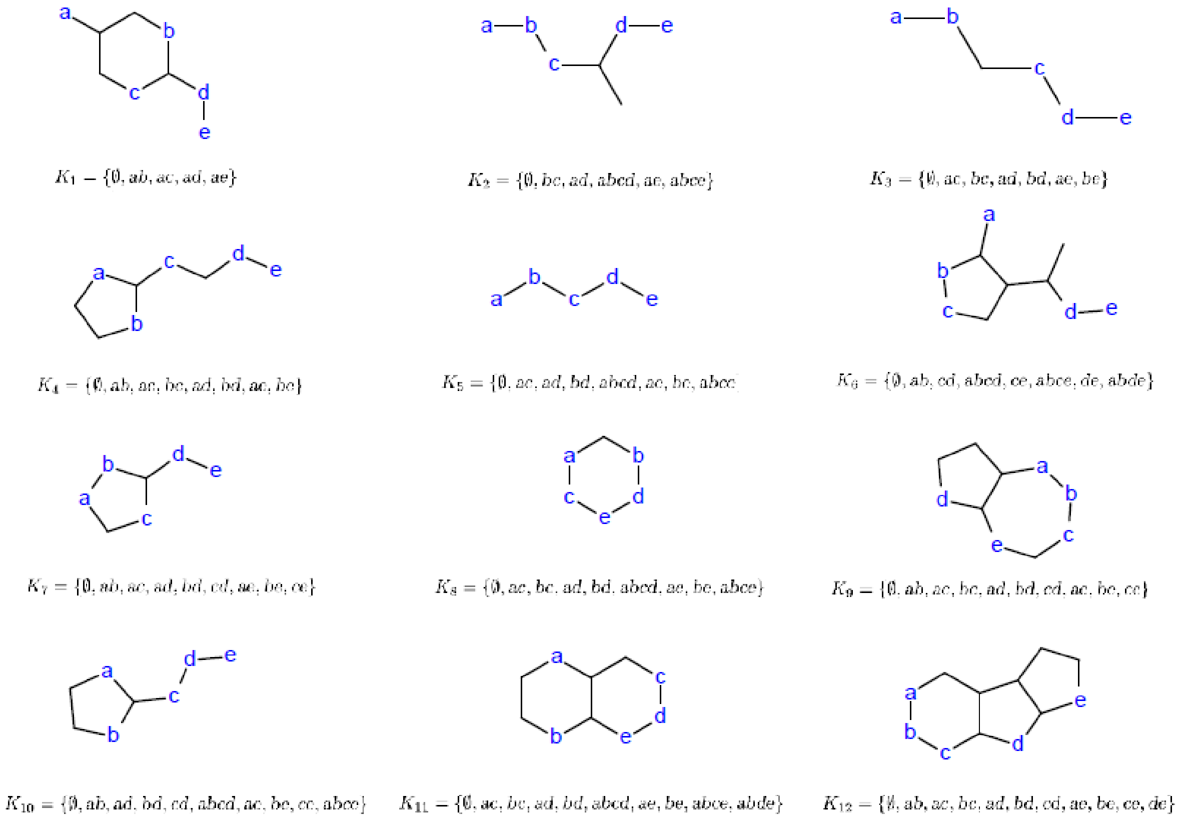
$K_5 = \{\emptyset, ac, bc, ad, bd, abcd\}$

$K_6 = \{\emptyset, ab, ac, bc, ad, bd, cd\}$

$K_7 = \{\emptyset, ba, ca, cb, da, db, dc, dcba\}$

Figure 12: Realistic Graphs for each cell in Hesselink's classification of all cells with rank 4. K_1 resembles benzene, K_5 napthalene, and K_6 azulene.

Polycylic aromatic hydrocarbons are commonly produced as byproducts of fuel burning, and have been classified as carcinogenic, mutagenic, and teratogenic. This could pose a challenging problem if we wish to use large aromatic hydrocarbons in circuitry.



$K_1 = \{\emptyset, ab, ac, ad, ae\}$ $K_2 = \{\emptyset, bc, ad, abcd, ae, abce\}$ $K_3 = \{\emptyset, ac, bc, ad, bd, ae, be\}$

$K_4 = \{\emptyset, ab, ac, bc, ad, bd, ae, be\}$ $K_5 = \{\emptyset, ac, ad, bd, abcd, ae, be, abce$ $K_6 = \{\emptyset, ab, cd, abcd, ce, abce, de, abde\}$

$K_7 = \{\emptyset, ab, ac, ad, bd, cd, ae, be, ce\}$ $K_8 = \{\emptyset, ac, bc, ad, bd, abcd, ae, be, abce\}$ $K_9 = \{\emptyset, ab, ac, bc, ad, bd, cd, ae, be, ce\}$

$K_{10} = \{\emptyset, ab, ad, bd, cd, abcd, ae, be, ce, abce\}$ $K_{11} = \{\emptyset, ac, bc, ad, bd, abcd, ae, be, abce, abde\}$ $K_{12} = \{\emptyset, ab, ac, bc, ad, bd, cd, ae, be, ce, de\}$

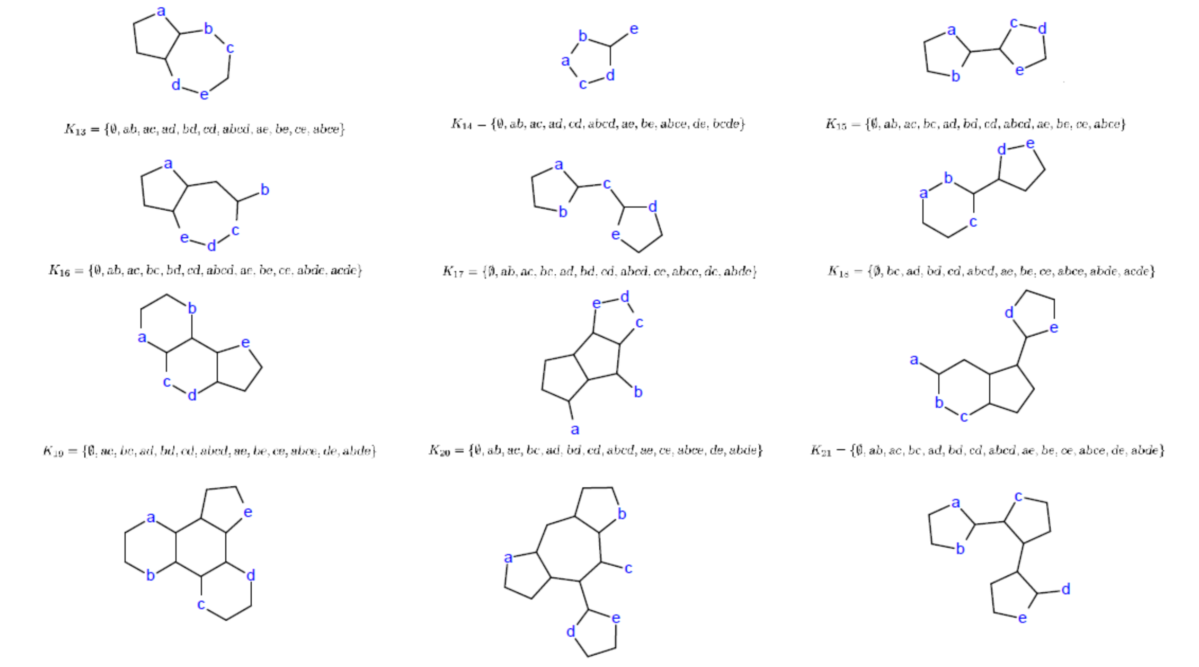
Figure 13: Realistic Graphs for the first 12 cells in Hesselink's classification of all cells with rank 5. K_1 and K_8 resembles benzene, K_{11} napthalene, and K_9 azulene.

4.1 Pending Work and Future Directions

Currently a genetic algorithm is implemented in Java as per Section 3 and has provided results for all the figures seen previously. However, the Restriction section needs review and completeness from a chemist, to ensure correctness. Other pending work includes research into the synthesis of hydrocarbons, obtaining a complete profile of the genetic algorithm's effectiveness, fact checking / citations, and the formatting of this document.

It would be interesting to see an approach using graph modification, rather than creation. Then one could use such graphs found by Hesselink, and mutate them until they resemble realistic graphs. Some methods include Hesselink et al's [2] process of splitting a node of high degree into two nodes of lower degree (or vice versa). It can also be used to shift edges between certain nodes. [3] also outlines many operations to edit graphs without changing the switching behavior. A formal comparison between such methods and our own in finding realistic graphs would be invaluable.

There are many possible improvements in our implementation of graph generation through a genetic algorithm. As of now, graphs have only been found for cells of small rank (<7). One technique to allow the discovery of more graphs would be to increase the number of maximum nodes.



$K_{13} = \{\emptyset, ab, ac, ad, bd, cd, abcd, ae, be, ce, abce\}$ $K_{14} - \{\emptyset, ab, ac, ad, cd, abcd, ae, be, abce, de, bcde\}$ $K_{15} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, ae, be, ce, abce\}$

$K_{16} = \{\emptyset, ab, ac, bc, bd, cd, abcd, ae, be, ce, abde, acde\}$ $K_{17} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, ce, abce, dc, abde\}$ $K_{18} = \{\emptyset, bc, ad, bd, cd, abcd, ae, be, ce, abce, abde, acde\}$

$K_{19} = \{\emptyset, ae, bc, ad, bd, cd, abcd, ae, be, ce, abce, de, abde\}$ $K_{20} = \{\emptyset, ab, ae, bc, ad, bd, cd, abcd, ae, ce, abce, de, abde\}$ $K_{21} - \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, ae, be, ce, abce, de, abde\}$

$K_{22} - \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, be, ce, abce, de, abde, acde\}$ $K_{23} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, ae, be, ce, abce, de, abde, acde\}$ $K_{24} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, ae, be, ce, abce, de, abde, acde, bcde\}$

Figure 14: Realistic Graphs for the second half of cells in Hesselink's classification of all cells with rank 5. K_{13}, K_{16} azulene both resemble azulene.

This could yield more results on some of the more difficult cells, but involve more computation, even in the average case. Perhaps it is possible to implement a more efficient cycle detection algorithm, or even calculate fitness on every 2nd iteration of the genetic algorithm. These changes may allow even more results to be yielded from a heuristic genetic algorithm based approach.

References

[1] W. H. Hesselink, Graph Theory for alternating hydrocarbons with attached ports. Indagationes Mathematicae, Elsevier, 24:115141, 2013.

[2] W.H. Hesselink, J.C. Hummelen, H.T. Jonkman, H.G. Reker, G.R. Renardel de Lavalette, M.H. van der Veen, Kekule Cells for Molecular Computation. Cornell University Library Online, 2013.

[3] M.H. van der Veen. π-Logic. PhD thesis, University of Groningen, May 2006.

[4] A.J. Heeger, S. Kivelson, J.R. Schrieffer, and W.-P Su. Solitons in conducting polymers, Rev. Mod. Phys.,60:781, 1988.

[5] G. E. Moore, Cramming more components onto integrated circuits. Electronics Magazine. p. 4. Retrieved 2006-11-11.

[6] C. A. Mack, Fifty Years of Moore's Law. IEEE Transactions on semiconductor manufacturing, 24:2 2011.

[7] M. H. van der Veen, M. T. Rispens, H. T. Jonkman, and J. C. Hummelen, Molecules with Linear π-Conjugated Pathways between all Substituents: Omniconjugation. Adv. Function Mater, 14:3, 2004.

[8] S.N. Yalirahi, M.A. Ratner, Interplay of topology and chemical stability on the electronic transport of molecular junctions, Ann. New York Acad. Sci. 960 (2002) 153.

[9] J. Reichert, R. Ochs, D. Beckmann, H. B. Weber, M. Mayor, H. von Lhneysen,Phys. Rev. Lett. 2002, 88, 176804.

[10] A. Aviram, M. A. Ratner, Chem. Phys. Lett. 1974, 29, 277.

[11] D. B. Strukov, K. K. Likharev, Nanotechnology 2005, 16, 137.

[12] Y. Karzazi, J. Cornil, J. L. Brdas, J. Am. Chem. Soc. 2001, 123, 10076.

[13] J. M. Tour, M. Kozaki, J. M. Seminario, J. Am. Chem. Soc. 1998, 120, 8486.

[14] A. Aviram, J. Am. Chem. Soc. 1988, 110, 5687.

[15] H. W. Ch. Postma, T. Teepen, Z. Yao, M. Grifoni, C. Dekker, Science 2001, 293, 76.

[16] C. Joachim, J. K. Gimzewski, Chem. Phys. Lett. 1997, 265, 353.

[17] T. D. Anthopoulos, C. Tanase, S. Setayesh, E. J. Meijer, J. C. Hummelen, P. W. M. Blom, D. M. de Leeuw, Adv. Funct. Mater. 2004, 16, 2174.

[18] G. M. Tsivgoulis, J.-M. Lehn, Chem. Eur. J. 1996, 2, 1399.

[19] J. J. D. de Jong, L. N. Lucas, R. M. Kellogg, J. H. van Esch, B. L. Feringa, Science 2004, 304, 278.

[20] A. P. de Silva, H. Q. N. Gunaratne, C. P. McCoy, Nature 1993, 364, 42.

[21] F. M. Raymo, S. Giordani, J. Org. Chem. 2003, 68, 4158.

[22] K. Rurack, A. Koval'chuck, J. L. Bricks, J. L. Slominskii, J. Am. Chem. Soc. 2001, 123, 6205.

[23] D. Parker, J. A. G. Williams, Chem. Commun. 1998, 245.

[24] A. P. de Silva, I. M. Dixon, H. Q. N. Gunaratne, T. Gunnlaugsson, P. R. S. Maxwell, T. E. Rice, J. Am. Chem. Soc. 1999, 121, 1393.

[25] Roberts, J. D., Streitwieser Jr, A., and Regan, C. M. Small-Ring Compounds. X. Molecular Orbital Calculations of Properties of Some Small-Ring Hydrocarbons and Free Radicals. Journal of the American Chemical Society, 74(18), 4579-4582. 1952.

[26] Kruszewski, J., and Krygowski, T. M. An Extension of the Hckel 4N+ 2 Rule to Polycyclic Non-alternant Conjugated Hydrocarbons. Canadian Journal of Chemistry, 53(6), 945-951. 1975.

[27] Hafner, K., Donges, R., Goedecke, E. and Kaiser, R. Concerning Pentalene, 2-Methylpentalane, and 1,3- Dimethylpentalene. Angew. Chem. Int. Ed. Engl., 12: 337-339. 1973.

[28] Hafner, K., Bangert, K. F. and Orfanos, V 1,3-Bis(dimethylamino)pentalene. Angew. Chem. Int. Ed. Engl., 6: 451452. 1967.