Binary coding of algebraic Kekulé structures of catacondensed benzenoid graphs

Damir Vukičević\textsuperscript{a}, Petra Žigert\textsuperscript{b,∗}

\textsuperscript{a} Faculty of Science, University of Split, Nikole Tesle 12, 21000 Split, Croatia
\textsuperscript{b} Department of Mathematics and Computer Sciences, PEF, University of Maribor, Koroška cesta 160, SI-2000 Maribor, Slovenia

Received 8 December 2004; received in revised form 17 July 2007; accepted 27 July 2007

Abstract

The algebraic Kekulé structure of a finite catacondensed benzenoid graph with $h$ hexagons is described by a binary code of length $h$. The procedure can be reversed, any (algebraic) Kekulé structure can be recovered from its binary code.

Keywords: Benzenoid hydrocarbons; Benzenoid graph; Kekulé structures; Algebraic Kekulé structures; Randić structures; 1-factor; Binary coding

1. Introduction

Benzenoid hydrocarbons are polycyclic conjugated molecules composed of six-membered rings. Each pair of carbon atoms is connected by a single or a double bond. Double bonds in a single ring represent the so-called π-electrons. The distribution of π-electrons is represented by the Kekulé structures. For the computer-aided manipulation with Kekulé structures a short representation of such a structure is very welcome. The possibility of storing a complete information on a Kekulé structure by means of a short binary string is another asset that may become indispensable when dealing with benzenoid hydrocarbons possessing thousands and hundreds of thousands of Kekulé structures \cite{1}. Such benzenoid hydrocarbons have recently been both synthesized \cite{2} and detected in a flame \cite{3}. The concept of algebraic Kekulé structures for several classes of polycyclic conjugated molecules has been presented in \cite{4}. For the case of catacondensed benzenoid hydrocarbons the binary coding procedure can be found in \cite{5}. Our aim is to present a new way of the binary coding using the algebraic Kekulé structures.

In the mathematical literature we usually refer to benzenoid hydrocarbons as benzenoid graphs. Benzenoid graphs are 2-connected subgraphs of the hexagonal lattice so that every bounded face is a hexagon. If all vertices of a benzenoid graph $B$ lie on its perimeter, then $B$ is said to be catacondensed; otherwise it is pericondensed. If an edge of a benzenoid graph $B$ does not belong to the perimeter of $B$, it is said to be an inner edge. A hexagon with exactly one inner edge is a terminal hexagon or a leaf. We will restrict ourselves to the finite case of catacondensed benzenoid graphs. So, each finite catacondensed benzenoid graph contains at least two terminal hexagons, except

\textsuperscript{∗} Corresponding author.
E-mail addresses: vukicevi@pmfst.hr (D. Vukičević), petra.zigert@uni-mb.si (P. Žigert).

0893-9659/5 - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aml.2007.07.027
benzene, which will be considered as a terminal hexagon as well. Let the set of all hexagons of $B$ be $\mathcal{H}(B)$, and let the set of all terminal hexagons of $B$ be $\mathcal{L}(B)$.

A matching of a graph $B$ is a set of pairwise independent edges. A matching is perfect or a 1-factor, if it covers all vertices of $B$. There is a natural one-to-one correspondence between the Kekulé structures of a benzenoid hydrocarbon and the 1-factors of a benzenoid graph. For details see [1].

Randić introduced in [6] the so-called algebraic Kekulé structures for benzenoid graphs, also called the Randić structures. In [7] Gutman et al. proved the one-to-one correspondence between ordinary and algebraic Kekulé structures of a catacondensed benzenoid graph, other than benzene.

For a catacondensed case the algebraic Kekulé structures can be defined as follows. Given a 1-factor of a catacondensed benzenoid graph $B$ we assign to each hexagon of $B$ an integer called the Randić hexagon number, which represents that 1-factor’s contribution of $\pi$-electrons to that particular hexagon. If $\pi$-electrons of an edge in the 1-factor are on the perimeter of $B$, then the contribution to the corresponding hexagon is 2, otherwise such $\pi$-electrons are shared by two hexagons and they contribute 1 to each hexagon. In order to obtain the Randić hexagon number of a hexagon $H$ of $B$ one must sum up all contributions of $\pi$-electrons of double bonds around $H$, see Fig. 1(a).

Let $B$ be a finite catacondensed benzenoid graph and let $\rho : \mathcal{H}(B) \to \{0, 1, 2, 3, 4, 5, 6\}$ be the function which assigns to each hexagon of $B$ its Randić hexagon number. Further, let $L \in \mathcal{L}(B)$. Then we define the labeling function $f_{B,L} : \mathcal{H}(B) \to \{0, 1\}$ by

\[
 f_{B,L}(L) = \begin{cases} 
 1, & \rho(L) > 4, \\
 0, & \text{otherwise},
\end{cases}
\]

and for $H \in \mathcal{H}(B) \setminus \{L\}$ we set

\[
 f_{B,L}(H) = \rho(H)(\text{mod } 2).
\]

The labeling function $f_{B,L}(L)$ for the benzenoid graph $B$ from Fig. 1(a) is shown in Fig. 1(b), where $L$ is a randomly chosen terminal hexagon.

In this note we prove the following result:

**Theorem 1.** Let $B$ be a finite catacondensed benzenoid graph other than benzene with a terminal hexagon $L$. Then the labeling function $f_{B,L}$ uniquely determines the 1-factor of $B$. 

---

Fig. 1. (a) The algebraic Kekulé structure of $B$, (b) a terminal hexagon $L$ and the labeling function $f_{B,L}$. 

---
2. Proof of the theorem

Lemma 2. Let $B$ be a finite catacondensed benzenoid graph other than benzene with a terminal hexagon $L$. If the labeling function $f_{B,L}$ is given, then one can determine for each inner edge of $B$, whether it belongs to the 1-factor or not.

Proof. Let us proceed by induction on $|\mathcal{H}(B)|$. Suppose that $|\mathcal{H}(B)| = 2$. Let $\mathcal{H}(B) = \{L, L'\}$ and let $e$ be the inner edge of $B$. It is straightforward to see, that $e$ belongs to the 1-factor if and only if $f_{B,L}(L') = 1$.

Now, let $|\mathcal{H}(B)| > 2$. Let $L' \in \mathcal{L}(B) \setminus \{L\}$, let $e$ be the inner edge of $L'$ and $H'$ the neighbouring hexagon of $L'$ in $B$. Further, let $B'$ be a catacondensed benzenoid graph obtained from $B$ by eliminating the terminal hexagon $L'$ from $B$. If $f_{B,L}(L')$ is an even number, then $e$ does not belong to the 1-factor and $f_{B,L}(H) = f_{B',L}(H)$ for each hexagon $H \in \mathcal{H}(B')$. Hence the lemma follows by induction in this case.

Now, suppose $f_{B,L}(L')$ is an odd number. Then edge $e$ belongs to the 1-factor of $B$. Also $f_{B,L}(H) = f_{B',L}(H)$ for each hexagon $H \in \mathcal{H}(B') \setminus \{H'\}$ and $f_{B,L}(H') = |1 - f_{B',L}(H')|$. Therefore, the lemma follows by induction in this case, too.

We are now ready for the proof of Theorem 1. Considering the Lemma 2 the 1-factor is uniquely determined on the hexagon $L$. Suppose the labeling function $f_{B,L}$ does not uniquely determine the 1-factor on the catacondensed benzenoid graph $B$. Therefore two hexagons from $\mathcal{H}(B)$ exist such that, on one, let it be $D$ the 1-factor is uniquely determined and on another one, say $U$, the 1-factor is not uniquely determined. Further, these two hexagons share an inner edge of $B$. Considering the symmetry, one of the 12 cases from Fig. 2 occurs. It is not difficult to check that in each case, using the Lemma 2 the 1-factor is uniquely determined on the hexagon $U$ which leads us to a contradiction.

3. The binary code

Using the labeling function $f_{B,L}(B)$ of a catacondensed benzenoid graph we can easily assign a binary code of length $|\mathcal{H}(B)|$ to the algebraic Kekulé structure of $B$. The inner dual of a catacondensed benzenoid graph $B$ is a $|\mathcal{H}(B)|$ vertex tree $T$. An arbitrary terminal hexagon is selected as the root of this tree. The hexagons
of $B$ are then numbered such that the hexagon $H_i$ is a predecessor of the hexagon $H_j$ in $T$ if and only if $i < j$. Such a numbering can be obtained by the Depth-First Search algorithm (DFS) or by the Breadth-First Search algorithm (BFS), see [8]. The bits of a binary code are then the labels of hexagons. For example, on Fig. 3(a) we see the DFS numbering of the hexagons of a benzenoid graph $B$ from Fig. 1(b) with $L$ as a rooted vertex. Further, the binary code is assigned according to the algebraic Kekulé structure from the same figure.

Naturally, the procedure can be reversed. Given the binary code one can reconstruct the algebraic Kekulé structure from it. Of course, in order to be able to do that, the selected ordering of the hexagons must be known. Again, similarly as in the proof of the Theorem 1, the Lemma 2 is essential in the reconstruction procedure, too. In Fig. 3 we see two steps of the reconstruction procedure of the algebraic Kekulé structure from Fig. 1(a). In the first step (a) the order of hexagons is determined and the 1-factor on inner edges is provided by Lemma 2. In the second step (b) the 1-factor on the perimeter is constructed and the Randić hexagon numbers are assigned.

4. Application

Here we propose one of the possible applications of our result. Suppose that we have to solve the following problem. We are faced with a large family of Randić structures and we need to find the largest subset of this family in which no structure repeats twice (we assume that the number of structures is much larger than the number of hexagons). Of course, we could compare all pairs of structures and solve the problem in a quadratic amount of time.

The alternative approach is the following:

1. Form an array of $2^{|\mathcal{H}(B)|}$ bits (each bit set to 0) and address each of them by a binary code of length $|\mathcal{H}(B)|$.
2. For each structure
   2.1 find the code of the observed structure,
   2.2 if there is number 0 on the location of this code replace it by 1,
   2.3 if there is number 1 on the location of its code eliminate this structure.

It can be easily seen that this algorithm solves our problem at a linear (depending on the number its structures) amount of time. The speed of this algorithm is critically determined by the speed of the execution of the step 2.1. Therefore, it is very important to have a fast coding algorithm. Note that our approach to coding Randić structures is extremely fast and hence it results in a very efficient and good algorithm.
Acknowledgements

The authors would like to thank the referees for all the comments and remarks, specially for those about the finiteness. The second author was supported by the Ministry of Education, Science and Sport of Slovenia under the grant J1-6150.

References