Interchangeability of Relevant Cycles in Graphs

Petra M. Gleiss[a], Josef Leydold[b,*] and Peter F. Stadler[a,c]

[a]Institute for Theoretical Chemistry and Molecular Structural Biology,
University of Vienna, Währingerstrasse 17, A-1090 Vienna, Austria
Phone: **43 1 4277-52737 Fax: **43 1 4277-52793
E-Mail: {pmg,studla}@tbi.univie.ac.at
URL: http://www.tbi.univie.ac.at/~{pmg,studla}

[b]Dept. for Applied Statistics and Data Processing
University of Economics and Business Administration
Augasse 2-6, A-1090 Wien, Austria
Phone: **43 1 31336-4695 Fax: **43 1 31336-738
E-Mail: Josef.Leydold@statistik.wu-wien.ac.at
URL: http://statistik.wu-wien.ac.at/staff/leydold
*Address for correspondence

[c]The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
Phone: (505) 984 8800 Fax: (505) 982 0565
E-Mail: stadler@santafe.edu
URL: http://www.tbi.univie.ac.at/~studla

Submitted: July 26, 1999; Accepted: March 12, 2000.

Abstract

The set $\mathcal{R}$ of relevant cycles of a graph G is the union of its minimum cycle bases. We introduce a partition of $\mathcal{R}$ such that each cycle in a class $\mathcal{W}$ can be expressed as a sum of other cycles in $\mathcal{W}$ and shorter cycles. It is shown that each minimum cycle basis contains the same number of representatives of a given class $\mathcal{W}$. This result is used to derive upper and lower bounds on the number of distinct minimum cycle bases. Finally, we give a polynomial-time algorithm to compute this partition.

Keywords: Minimum Cycle Basis, Relevant Cycles

AMS Subject Classification: Primary 05C38. Secondary 05C85, 92D20, 92E10.

1. Introduction

Cycle bases of graphs have a variety of applications in science and engineering. For example, applications occur in structural flexibility analysis [9], electrical networks [3], and in chemical structure storage and retrieval systems [5]. Brief surveys and extensive references can be found in [8, 7].

The set $\mathcal{R}$ of relevant cycles of a graph G is the union of its minimum cycle bases [12, 15]. We define an equivalence relation "interchangeability" on $\mathcal{R}$ such that the cycles in a class $\mathcal{W} \in \mathfrak{P}$ of the associated partition $\mathfrak{P}$ can be expressed as a sum of a linearly independent set consisting of other cycles in $\mathcal{W}$ and shorter cycles. The motivation for introducing $\mathfrak{P}$ originated in the context of RNA folding; a brief sketch is included as appendix.

The main result is that every class $\mathcal{W}$ in $\mathfrak{P}$ has the following property: the cardinality of the intersection of $\mathcal{W}$ with every minimal cycle basis is the same. This result is used to prove upper and lower bounds on the number of distinct minimal cycle bases.

The partition $\mathfrak{P}$ can be obtained in polynomial time from $\mathcal{R}$. While the number of relevant cycles may grow exponentially with the number $|V|$ of vertices [15], there are typically only $\mathcal{O}(|V|^3)$ relevant cycles [6].

2. Preliminaries

Let $G(V, E)$ be a simple, connected, unweighted, undirected graph with vertex set V and edge set E. The set $\mathcal{E}$ of all subsets of E forms an m-dimensional vector space over GF(2) with vector addition $X \oplus Y := (X \cup Y) \setminus (X \cap Y)$ and scalar multiplication $1 \cdot X = X$, $0 \cdot X = \emptyset$ for all $X, Y \in \mathcal{E}$. In order to simplify the notation we shall write

$$\bigoplus_{\mathcal{X}} = \bigoplus_{C \in \mathcal{X}} C \tag{1}$$

for $\mathcal{X} \subseteq \mathcal{E}$. A *generalized cycle* is a subgraph such that any vertex degree is even. A *cycle* is a connected subgraph such that every vertex has degree 2. We represent a (generalized) cycle by its edge set C.

The set $\mathcal{C}$ of all generalized cycles forms a subspace of $(\mathcal{E}, \oplus, \cdot)$ which is called the *cycle space* of G. A basis $\mathcal{B}$ of the cycle space $\mathcal{C}$ is called a *cycle basis* of $G(V, E)$ [1]. The dimension of the cycle space is the *cyclomatic number* or *first Betti number* $\nu(G) = |E| - |V| + 1$. It is obvious that the cycle space of a graph is the direct sum of the cycle spaces of its 2-connected components. It will be sufficient therefore to consider only 2-connected graphs in this contribution.

The length $|C|$ of a generalized cycle C is the number of its edges. The length $\ell(\mathcal{B})$ of a cycle basis $\mathcal{B}$ is the sum of the lengths of its generalized cycles: $\ell(\mathcal{B}) = \sum_{C \in \mathcal{B}} |C|$. A *minimum cycle basis* $\mathcal{M}$ is a cycle basis with minimum length. The generalized cycles in $\mathcal{M}$ are chord-less cycles (see [8]). Hence we may consider cycles instead of generalized cycles from here on. For the sake of completeness we note that a minimum cycle basis is a cycle basis in which the longest cycle has the minimum possible length [2].

Definition 1. [12] *A cycle C is* relevant *if it cannot be represented as an $\oplus$-sum of shorter cycles. We denote the set of all relevant cycles by $\mathcal{R}$.*

Proposition 2. [15] *A cycle C is relevant if and only if it is contained in a minimum cycle basis.*

Definition 3. [6] *A cycle C in G is* essential *if it is contained in every minimum cycle basis of G.*

The set of all cycles of a graph G forms a matroid, see e.g. [11, 18]. We restate this fact in the following form:

Proposition 4 (Matroid Property)**.** *Let $\mathcal{Q}$ be a set of cycles containing a minimum cycle basis. Then a minimum cycle basis $\mathcal{B}$ can be extracted from $\mathcal{Q}$ by a* greedy procedure *in the following way: (i) Sort $\mathcal{Q}$ by cycle length and set $\mathcal{B} = \emptyset$. (ii) Transversing $\mathcal{Q}$ in the established order, set $\mathcal{B} \leftarrow \mathcal{B} \cup \{C\}$ whenever $\mathcal{B} \cup \{C\}$ is linearly independent.*

3. A Partition of $\mathcal{R}$

Lemma 5. *For each relevant cycle $C \in \mathcal{R}$, exactly one of the following holds:*

(i) C is essential, or

(ii) There is a cycle $C' \in \mathcal{R}$, $C' \neq C$, and a set of relevant cycles $\mathcal{X} \subseteq \mathcal{R} \setminus \{C, C'\}$ such that $\mathcal{X} \cup \{C'\}$ is linearly independent, $|C| = |C'|$, $|C''| \leq |C|$ for all $C'' \in \mathcal{X}$, and $C = C' \oplus \bigoplus_{\mathcal{X}}$.

Proof. Let $\mathcal{Y} = \{C'' \in \mathcal{R} \big| \, |C''| \leq |C|\}$. If $\operatorname{rank}(\mathcal{Y}) > \operatorname{rank}(\mathcal{Y} \setminus \{C\})$, then C is contained in every minimum cycle basis as an immediate consequence of the matroid property. In other words, C is essential.

Now assume $\operatorname{rank}(\mathcal{Y}) = \operatorname{rank}(\mathcal{Y} \setminus \{C\})$. Hence $C = \bigoplus_{\mathcal{Z}}$ for some $\mathcal{Z} \subseteq \mathcal{Y} \setminus \{C\}$. Without loss of generality we may assume that $\mathcal{Z}$ is an independent set of cycles. By the relevance of C, $\mathcal{Z}$ cannot consist only of cycles that are all strictly shorter than C, thus there is $C' \in \mathcal{Z}$ such that $|C'| = |C|$, and we can write

$$C = C' \oplus \bigoplus_{\mathcal{Z} \setminus \{C'\}}. \tag{2}$$

It remains to show that C is not essential in this case: Adding $C \oplus C'$ to both sides of equ.(2) yields $C' = C \oplus \bigoplus_{\mathcal{Z}\setminus\{C'\}}$. Thus we may extract two different minimum cycle bases from $\mathcal{R}$ one of which contains C but not C', while the other contains C' but not C, simply by ranking C before or after C' when sorting $\mathcal{R}$. Thus neither C nor C' is essential. $\square$

Definition 6. *Two relevant cycles $C, C' \in \mathcal{R}$ are* interchangeable*, $C \leftrightarrow C'$, if (i) $|C| = |C'|$ and (ii) there is a set $\mathcal{X} \subset \mathcal{R}$, $C, C' \notin \mathcal{X}$, $\mathcal{X} \cup \{C'\}$ is a linearly independent subset of relevant cycles, such that $C = C' \oplus \bigoplus_{\mathcal{X}}$ and $|C''| \le |C|$ for all $C'' \in \mathcal{X}$.*

Lemma 7. *Interchangeability is an equivalence relation on $\mathcal{R}$.*

Proof. Trivially, we have $C \leftrightarrow C$; symmetry follows immediately from the proof of lemma 5. In order to verify transitivity, assume $C \leftrightarrow C'$, $C' \leftrightarrow C''$ and set $C' = C \oplus \bigoplus_{\mathcal{X}}$ and $C'' = C' \oplus \bigoplus_{\mathcal{X}'}$. We have to distinguish two cases:

(i) $C'' \in \mathcal{X}$. Then $C' = C \oplus C'' \oplus \bigoplus_{\mathcal{X}\setminus\{C''\}}$. Adding $C \oplus C'$ on both sides yields $C = C'' \oplus \bigoplus_{\mathcal{X}\setminus\{C''\}\cup\{C'\}}$. By assumption, $\mathcal{X} \cup \{C'\}$ does not contain C and is an independent subset of relevant cycles, i.e., $C \leftrightarrow C''$. The case $C \in \mathcal{X}'$ is treated analogously.

(ii) $C'' \notin \mathcal{X}$ and $C \notin \mathcal{X}'$. We have

$$C'' = \Big(C \oplus \bigoplus_{\mathcal{X}}\Big) \oplus \bigoplus_{\mathcal{X}'} = C \oplus \bigoplus_{\mathcal{X}\triangle\mathcal{X}'} = C \oplus \bigoplus_{\mathcal{Z}}$$

where $\mathcal{X}\triangle\mathcal{X}'$ denotes the symmetric difference, and $\mathcal{Z} \subseteq \mathcal{X}\triangle\mathcal{X}'$ is a non-empty independent set of cycles that does not contain C or C''. Thus $C'' \leftrightarrow C$. $\square$

Corollary 8. *A relevant cycle C is essential if and only if it is not $\leftrightarrow$-interchangeable with any other cycle.*

Remark. We cannot assume that for the set $\mathcal{X} \subset \mathcal{R}$ in definition 6, $\mathcal{X} \cup \{C'\}$ is a subset of a minimum cycle basis. Figure 1 gives a counter example. In what follows let C_F, C'_F, C_G and C'_G denote the relevant cycles of length 6 through F and G, respectively, and let C_O be the cycle $\{O_1, \dots, O_6\}$. $\mathcal{Z}$ always denotes an independent subset of $\mathcal{R} \setminus \{C_F, C'_F, C_G, C'_F, C_O\}$. Then $C_F = C_G \oplus (C'_G \oplus C'_F \oplus \bigoplus_{\mathcal{Z}})$, where the right hand side is linearly independent, i.e., $C_F \leftrightarrow C_G$. However, the r.h.s. contains both C_G and C'_G and hence it is not a subset of a minimum cycle basis. Moreover, C_F cannot be expressed as an $\oplus$-sum of an independent subset of relevant cycles that contains C_G but not C'_G.

The graph in figure 1 demonstrates also that we cannot define a "stronger" interchangeability relation, $\leftrightarrow_s$, by replacing the condition that $\mathcal{X} \cup \{C'\}$ is independent by "$\mathcal{X} \cup \{C'\}$ is a subset of a minimum cycle basis" in definition 6. The relation $\leftrightarrow_s$ is not symmetric: We find $C_F = C_O \oplus (C'_F \oplus \bigoplus_{\mathcal{Z}})$, where the r.h.s. is a subset of a minimum cycle basis, i.e., $C_F \leftrightarrow_s C_O$. However, we always have $C_O = C_F \oplus (C'_F \oplus \bigoplus_{\mathcal{Z}})$ where the r.h.s is not a subset of a minimum cycle basis.

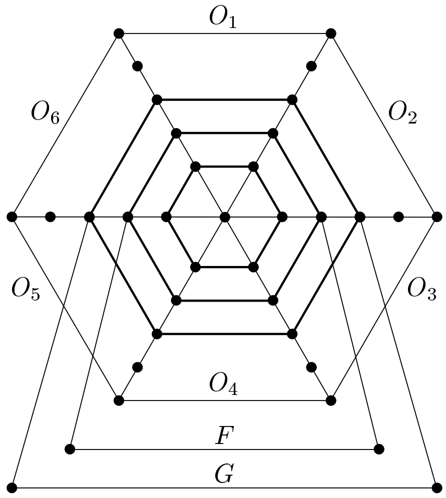


Figure 1. The set of relevant cycles of this graph consists of all triangles, all 4-cycles, two 6-cycles through F, two 6-cycles through G and the seven 6-cycles through at least one of the edges O_i. The three inner hexagons (thick lines) are not relevant, because they are the sum of triangles and 4-cycles. Notice that all 3- and 4-cycles are essential. Moreover, every minimum cycle basis contains exactly one 6-cycle through F and G, respectively, and six of the seven 6-cycles through at least one of the edges O_i. Moreover, no 6-cycle is essential.

Lemma 9. *Let C be a relevant cycle such that $C = \bigoplus_{\mathcal{X}}$ for a linearly independent set $\mathcal{X}$ of cycles with length less or equal $|C|$. Set $\mathcal{X}_= = \{C' \in \mathcal{X} \big| |C'| = |C|\}$. Then $C' \leftrightarrow C$ for each cycle $C' \in \mathcal{X}_=$.*

Proof. By lemma 7 $C \leftrightarrow C$. Assume there exists a $C' \in \mathcal{X}_= \setminus \{C\}$. Then $C' = C \oplus \bigoplus_{\mathcal{X}_= \setminus \{C\}}$, i.e., $C' \leftrightarrow C$ as proposed. □

Lemma 10. *Let $\mathcal{B}$ be a minimum cycle basis and let $\mathcal{W}$ be an $\leftrightarrow$-equivalence class of $\mathcal{R}$. Then $\mathcal{B} \cap \mathcal{W} \neq \emptyset$.*

Proof. Suppose there is a minimum cycle basis $\mathcal{B}$ and an $\leftrightarrow$-equivalence class $\mathcal{W}$ such that $\mathcal{W} \cap \mathcal{B} = \emptyset$. Choose $C \in \mathcal{W}$. By the matroid property there is an independent set of cycles $\mathcal{Q} = \mathcal{Q}_= \cup \mathcal{Q}_< \subseteq \mathcal{B}$ such that $C = \bigoplus_{\mathcal{Q}}$. By lemma 9 we have $\mathcal{Q}_= \subseteq \mathcal{W}$ which contradicts $\mathcal{B} \cap \mathcal{W} = \emptyset$ unless $\mathcal{Q}_= = \emptyset$. Thus $C = \bigoplus_{\mathcal{Q}_<}$ and hence $C \notin \mathcal{B}$ by proposition 2. □

Theorem 11. *Let $\mathcal{B}$ and $\mathcal{B}'$ be two minimum cycle bases and let $\mathcal{W}$ be an $\leftrightarrow$-equivalence class of $\mathcal{R}$. Then $|\mathcal{B}\cap\mathcal{W}| = |\mathcal{B}'\cap\mathcal{W}|$.*

Proof. Consider an $\leftrightarrow$-equivalence class $\mathcal{W}$ consisting of cycles of length l. Define $\mathcal{B}_= = \{C \in \mathcal{B} \big| |C| = l\}$, $\mathcal{B}_< = \{C \in \mathcal{B} \big| |C| < l\}$, and analogously for the second basis $\mathcal{B}'$. Assume $|\mathcal{B}'\cap\mathcal{W}| > |\mathcal{B}\cap\mathcal{W}|$ and set $\mathcal{W}\cap\mathcal{B} = \{C_1,\ldots,C_j\}$, $\mathcal{W}\cap\mathcal{B}' = \{D_1,\ldots,D_j,\ldots,D_k\}$. By lemma 10, $j > 0$. As a consequence of the matroid property we may assume $\mathcal{B}'_< = \mathcal{B}_<$ and we may write each D_i as a linear combination of cycles from $\mathcal{B}_< \cup \mathcal{B}_=$. Moreover by lemma 9 this linear combination cannot contain any cycles from $\mathcal{B}_= \setminus \mathcal{W}$. Since there are more than j cycles D_i there is a non-trivial linear combination

$$F = \bigoplus_{i\in I} D_i = \left[\bigoplus_{i\in J} C_i\right] \oplus \bigoplus\nolimits_{\mathcal{X}\subseteq\mathcal{B}'_<}$$

with $I \subseteq \{1,\ldots,k\}$ and $J \subseteq \{1,\ldots,j\}$ such that $\bigoplus_{i\in J} C_i = \mathsf{0}$. Thus

$$\left[\bigoplus_{i\in I} D_i\right] \oplus \bigoplus\nolimits_{\mathcal{X}\subseteq\mathcal{B}'_<} = \mathsf{0}$$

and hence $\{D_i| \in I\} \cup \mathcal{X} \subseteq \mathcal{B}'_= \cup \mathcal{B}'_<$ is linearly dependent, contradicting the assumption that $\mathcal{B}'$ is a basis. ☐

As an immediate consequence of theorem 11 we recover the well known fact [2, Thm. 3], that any two minimum cycle bases contain the same number of cycles with given length.

Definition 12. *Let $\mathcal{B}$ be a minimum cycle basis and let $\mathcal{W}$ be an $\leftrightarrow$-equivalence class of $\mathcal{R}$. We call* $\operatorname{knar}(\mathcal{W}) = |\mathcal{B}\cap\mathcal{W}|$ *the* relative rank *of $\mathcal{W}$ in $\mathcal{R}$.*

Corollary 13. *Let $\mathcal{W}$ be an $\leftrightarrow$-equivalence class such that* $\operatorname{knar}(\mathcal{W}) = k$. *Then each $C \in \mathcal{W}$ can be written as $C = \bigoplus_{\mathcal{Y}} \oplus \bigoplus_{\mathcal{Z}}$ where $\mathcal{Z}$ consists only of cycles shorter than $|C|$ and $\mathcal{Y} \subseteq \mathcal{W}\setminus\{C\}$ has cardinality $|\mathcal{Y}| \le \operatorname{knar}(\mathcal{W})$.*

We close this section with a few examples:

Complete graphs. The relevant cycles of a K_n, $n \ge 3$, are its triangles. It follows immediately that all triangles are $\leftrightarrow$-equivalent.

Outerplanar graphs. Outerplanar graphs have a unique minimal cycle basis [10], i.e., each relevant cycle is essential. Thus there are $\nu(G)$ interchangeability classes consisting of a single cycle.

Triangulations. For each triangulation of the sphere all relevant cycles of the corresponding graph are triangles. Moreover, The $\oplus$-sum of all triangles equals $\mathsf{0}$, while any proper subset is independent. Thus there is a single $\leftrightarrow$-equivalence class with $\operatorname{knar}(\mathcal{W}) = |\mathcal{R}| - 1$.

If we change the situation a little bit, such that there is exactly one face cycle C of length $l > 3$, i.e., the graphs corresponds to a triangulation of the plane but not the sphere, then C is the $\oplus$-sum of all triangles and hence not relevant. Thus all triangles are essential, i.e., we have $|\mathcal{R}|$ $\leftrightarrow$-equivalence classes, all of $\operatorname{knar}(\mathcal{W}) = 1$. This example demonstrates that partitioning into $\leftrightarrow$-equivalence classes — similar to number and length of minimum cycle bases — can be very unstable against small changes in the geometry of graphs.

Chordal graphs. The next example shows that there are rather "irregular-looking" examples for which all relevant cycles are contained in the same $\leftrightarrow$-equivalence class. A graph is *chordal* (also called triangulated or rigid circuit) if all cycles of length $|C| \geq 4$ contain a chord, i.e., an edge connecting two of its non-adjacent vertices.

Let G be connected and let A be a minimal separating vertex set. Then there are two connected graphs $G_i = (V_i, E_i)$, $i = 1, 2$ such that $V = V_1 \cup V_2$, $E = E_1 \cup E_2$, and $A = V_1 \cap V_2$. If $\Sigma = (A, E_1 \cap E_2)$ is a complete graph, $G_1 \cup G_2$ is called a *simplicial decomposition* of G at A. This procedure can be repeated until no further separating complete graphs can be found. It can be shown that the resulting indecomposable subgraphs are independent of the order of the decomposition [14, Prop.4.1]. The resulting components are the *simplicial summands* of G. A graph is chordal if and only if all its simplicial summands are complete graphs [4].

Lemma 14. *If G is a 3-connected chordal graph then $\mathcal{R}$ consists of a single $\leftrightarrow$-equivalence class.*

Proof. Since $C \in \mathcal{R}$ only if it is chord-less, it follows that all relevant cycles of a chordal graph are triangles. If G is 3-connected, the minimum separating clique Σ contains a triangle. Let G_1 and G_2 be the two adjacent simplicial summands. Then all triangles in G_1 are contained in a single $\leftrightarrow$-equivalence class; the same is true for all triangles in G_2. Since the intersection of G_1 and G_2 contains at least one triangle by assumption, all triangles of their union are contained in the same $\leftrightarrow$-equivalence class, and the lemma follows by induction. □

4. The Number of Minimal Cycle Bases

As an application of the $\leftrightarrow$-partition of $\mathcal{R}$ we derive bounds on the number of distinct minimal cycle bases of G.

Theorem 15. *Let $\mathcal{R} = \bigcup_{i=1}^{m} \mathcal{W}_i$ be the partition of the set of relevant cycles into $\leftrightarrow$-equivalence classes. Then the number M of distinct minimum cycle bases satisfies*

$$\prod_{i=1}^{m} |\mathcal{W}_i| \leq M \leq \prod_{i=1}^{m} \binom{|\mathcal{W}_i|}{\operatorname{knar}(\mathcal{W}_i)}. \tag{3}$$

Proof. The lower bound follows from the fact that, by lemma 10, each minimum cycle basis contains at least one element from each $\leftrightarrow$-equivalence class, and the fact that, by the matroid property, each element of $\mathcal{W}_i$ can be chosen. The upper bound follows directly from theorem 11 by assuming that the $\operatorname{knar}(\mathcal{W}_i)$ basis elements from $\mathcal{W}_i$ can be chosen freely. □

There even exists a universal bound that depends only on the number of relevant cycles and the cyclomatic number.

Corollary 16. *The number M of distinct minimum cycle bases satisfies*

$$M \leq \binom{|\mathcal{R}|}{\nu(G)}. \tag{4}$$

Proof. This upper bound follows immediately if we neglect any restrictions for the choice of $\nu(G)$ relevant cycles for a minimum cycle basis. □

Corollary 17. *Upper and lower bound coincide in equ.(3) if all $\leftrightarrow$-equivalence classes satisfy* $\operatorname{knar}(\mathcal{W}) = 1$ *or* $\operatorname{knar}(\mathcal{W}) = |\mathcal{W}| - 1$.

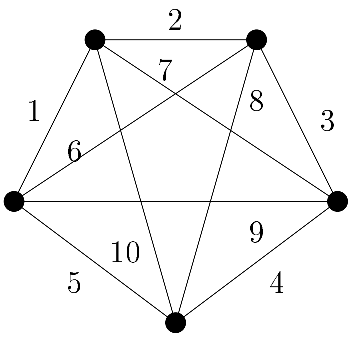
It is tempting to speculate that the upper bound might be attained by all graphs. Equivalently, then we could choose $\operatorname{knar}(\mathcal{W})$ cycles from $\mathcal{W}$ without restrictions when extracting a minimum cycle basis from $\mathcal{R}$. Unfortunately, this is not the case as the following examples show.

The triangles of K_5. Figure 2 lists the 10 triangles of K_5. Each triangle is contained in two of the five induced K_4-subgraphs a to e. Thus there are 5 dependent four-sets of cycles:

$$A \oplus B \oplus G \oplus J = 0 \qquad B \oplus C \oplus F \oplus H = 0 \qquad A \oplus E \oplus F \oplus I = 0$$
$$C \oplus D \oplus G \oplus I = 0 \qquad D \oplus E \oplus H \oplus J = 0$$

It is clear that all 10 cycles A through J are $\leftrightarrow$-equivalent forming a single equivalence class with $\operatorname{knar}(\text{triangles}) = \nu(K_5) = 6$. In general, it is clear that all triangles of a complete graph K_n, $n \geq 3$, belong to a single $\leftrightarrow$-equivalence class.

More importantly, however, 5 of the $\binom{10}{4} = 210$ combinations of 4 cycles and hence at least $5\binom{6}{2} = 75$ of the $\binom{10}{6} = 210$ sets of six triangles are dependent. As a consequence, neither the upper nor the lower bound in equ.(3) is an equality for K_5.



	1	2	3	4	5	6	7	8	9	10	a	b	c	d	e
A	*	*				*					*	*			
B		*	*				*				*				*
C			*	*				*						*	*
D				*	*				*				*	*	
E	*				*					*		*	*		
F		*						*		*		*			*
G			*			*			*		*			*	
H				*			*			*			*		*
I					*	*		*				*		*	
J	*						*		*		*		*		

Figure 2. The 10 triangles of K_5 cover the five sub-K_4s a through e twice.

Small relative ranks. The final example shows that Corollary 17 cannot be improved even if we restrict ourselves to graphs in which all $\leftrightarrow$-classes have small relative rank, or when only a single $\leftrightarrow$-class has $\mathrm{knar}\,(W) \geq 1$. The family of graphs in figure 3 shows that linearly dependent subsets $\mathcal{V} \subset \mathcal{W}$ with $|\mathcal{V}| \leq \mathrm{knar}\,(\mathcal{W})$ can be found even for $\mathrm{knar}\,(\mathcal{W}) = 2$.

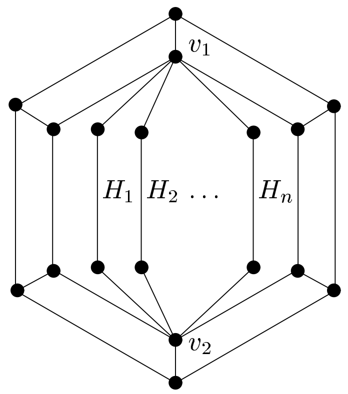


Figure 3. The 4-cycles are all essential. All 6-cycles are in one equivalence class $\mathcal{W}$ with $\mathrm{knar}\,(\mathcal{W}) = n+1$ and $|\mathcal{W}| = n^2/2+3n/2+2$. The outer cycle (of length 6) can be expressed as $\oplus$-sum of all 4-cycles and the inner 6-cycle that does not contain any path H_i. Thus no minimum cycle basis can contain these two 6-cycles.

5. A Connection with Vismara's Prototypes

Phillipe Vismara [15] describes an algorithm for constructing the set of relevant cycles $\mathcal{R}$ that makes use of a partitioning of $\mathcal{R}$ into *cycle families.* Let $\preceq$ be an arbitrary ordering of the vertex set V of G. Set $V_r = \{x \in V, x \preceq r\}$.

Proposition 18. [15] *Let C be a relevant cycle, and let r be the vertex of C that is maximal w.r.t. the order $\preceq$. Then there are vertices $p, q \in V_r$ such that C consists of two disjoint shortest paths $(r \dots p)$ and $(r \dots q)$ of the same lengths linked by the edge $\{p, q\}$ if $|C|$ is odd or a path (p, x, q), $x \in V_r$, if $|C|$ is even.*

Remark. A subgraph H of G is *isometric* if $d_H(x, y) = d_G(x, y)$ for all vertices $x, y \in V_H$. It is easy to verify that a relevant cycle must be isometric. The converse is not true, however: Horton [8] gives a counter-example.

Definition 19. [15] *Let C^r_{pqx} be a cycle as described in proposition 18. The cycle family $\mathcal{F}^r_{pqx}$ consists of all cycles C satisfying the following conditions:*

(i) $|C| = |C^r_{pqx}|$;

(ii) C contains the vertex r as well as the edge $\{p, q\}$ or the path (p, x, q);

(iii) There are two shortest paths $(p \dots r)$ and $(q \dots r)$ in C that pass only through vertices $\preceq$-smaller than r, i.e., that are contained in V_r.

Note that the cycle families $\mathcal{F}^r_{pqx}$ explicitly depend of the order $\preceq$ on V. Vismara shows that $\{\mathcal{F}^r_{pqx} | C^r_{pqx}$ is relevant$\}$ forms a partition of $\mathcal{R}$ for any order $\preceq$ on V.

Lemma 20. *Let $\mathcal{F} \subseteq \mathcal{R}$ be a relevant cycle family, and let $\mathcal{B}$ be a minimum cycle basis. Then for all $C, C' \in \mathcal{F}$ there is an independent set $\mathcal{Y} \subseteq \mathcal{B}$ such that $C \oplus C' = \bigoplus_{\mathcal{Y}}$ and $|C''| < |C| = |C'|$ for all $C'' \in \mathcal{Y}$.*

Proof. Let P, P' and Q, Q' be the paths connecting (r, p) and (r, q) in C and C', respectively. Then each of the combinations of paths $\{P, Q\}$, $\{P', Q\}$, $\{P, Q'\}$, $\{P', Q'\}$ belongs to a (possibly generalized) cycle in $\mathcal{F}$, which we denote by $C = C_{PQ}$, $C_{P'Q}$, $C_{PQ'}$, and $C' = C_{P'Q'}$ as outlined in [15]. Explicitly we have $C_{PQ} = P \oplus Q \oplus \{p, q\}$ if $|C|$ is odd and $C_{PQ} = P \oplus Q \oplus \{p, x\} \oplus \{x, q\}$ if $|C|$ is odd, etc. Note that the cycles $C_{P'Q}$ and $C_{PQ'}$ are not necessarily connected. Since P and P' have the same end points, their sum $P \oplus P'$ is an edge-disjoint union of cycles, which we denote by $\mathcal{A}$. Thus $C = C_{P'Q} \oplus \bigoplus_{\mathcal{A}}$ and analogously we obtain $C' = C_{P'Q'} = C_{P'Q} \oplus \bigoplus_{\mathcal{A}'}$, and thus $C' = C \oplus \bigoplus_{\mathcal{A} \triangle \mathcal{A}'}$. Since each cycle $C'' \in \mathcal{A} \triangle \mathcal{A}'$ satisfies $|C''| \leq 2d(r, p) = 2d(r, q) < |C|$, it follows from the matroid property that C'' can be written as an $\oplus$-sum of basis elements taken from $\mathcal{Y}$. $\square$

Corollary 21. *For each relevant cycle family $\mathcal{F}$ there is an $\leftrightarrow$-equivalence class $\mathcal{W}$ such that $\mathcal{F} \subseteq \mathcal{W}$.*

6. Computing Interchangeability Classes

Clearly, we may treat each length-class separately, since $\leftrightarrow$-equivalent cycles have the same length k. For a given length l we set $\mathcal{B}_< = \{C \in \mathcal{B} \big| |C| < l\} = \{C_1, \ldots, C_{n_0}\}$ where $\mathcal{B}$ is an arbitrary minimal cycle basis, and $\mathcal{R}_= = \{C \in \mathcal{R} \big| |C| = l\} = \{C_{n_0+1}, \ldots, C_{n_1}\}$. By lemma 9, two cycles $C_{j'}$, $C_{j''}$ of length l are $\leftrightarrow$-equivalent if and only if the linear equation

$$\left(\bigoplus_{C_{k'} \in \mathcal{B}_<} a_{k'} C_{k'} \right) \oplus \left(\bigoplus_{C_k \in \mathcal{R}_=} a_k C_k \right) = 0 \tag{5}$$

has a solution with $a_{j'} = a_{j''} = 1$ which is minimal in the following sense: if we take any strict subset of the coefficients with $a_k = 1$ and $a_{k'} = 1$ then there is no solution with exactly these coefficients being nonzero.

Let $A = (C_1, \ldots, C_{n_0}, C_{n_0+1}, \ldots, C_{n_1})$ be the $(|E| \times n_1)$-matrix with these cycles as its column vectors. It can be transformed into the reduced row echelon form $\tilde{A}$ by Gauß-Jordan elimination. Then exactly the first n rows of $\tilde{A}$ are nonzero, where $n = \operatorname{rank}(A)$. Denote the column of the first nonzero element in row i $(i = 1, \ldots, n)$ by p_i, i.e., $\tilde{A}_{i,p_i} = 1$ and $\tilde{A}_{i,j} = 0$ for all $j < p_i$. By construction, $i > i'$ implies $p_i > p_{i'}$; the columns p_i have a single nonzero entry $\tilde{A}_{i,p_i}$. Note that the upper-left $n_0 \times n_0$ submatrix of $\tilde{A}$ is the identity matrix since $\mathcal{B}_<$ is linearly independent.

Now we color the columns $n_0 + 1, \ldots, n_1$ of $\tilde{A}$ in the following way: (1) Two columns j' and j'' $(> n_0)$ have the same color if there exists a row i such that $\tilde{A}_{ij'} = \tilde{A}_{ij''} = 1$. (2) Use as many colors as possible. Notice that all such colorings are homomorphic.

Lemma 22. *Two cycles $C_{j'}, C_{j''} \in \mathcal{R}_=$ are $\leftrightarrow$-equivalent if and only if the columns j' and j'' have the same color in the above coloring of $\tilde{A}$.*

Proof. Let $P = \{p_1, \ldots, p_n\}$ and $\bar{P} = \{j | j = 1, \ldots, n_1,\ j \notin P\}$. If $j \in \bar{P}$ then $j > n_0$ since the upper left $n_0 \times n_0$ block is the identity matrix. When solving equ.(5) we can arbitrarily set $a_j = 0$ or $a_j = 1$ for each $j \in \bar{P}$, while the values a_{p_i} are then determined by this choice.

(i) If $p_i > n_0$ then C_{p_i} is essential if and only if $\tilde{A}_{ij} = 0$ for all $j \neq p_i$.

(ii) Set $a_j = 1$ for exactly one $j \in \bar{P}$, $j > n_0$. Then there is a row i such that $\tilde{A}_{ij} = 1$. This implies $\tilde{A}_{ip_i} = 1$. Moreover, the corresponding solution of equ.(5) is minimal. Thus $C_j \leftrightarrow C_{p_i}$.

(iii) Set $a_{j'} = a_{j''} = 1$ for exactly two columns $j', j'' \in \bar{P}$. We claim that the corresponding solution $\vec{a}$ of equ.(5) is minimal if and only if there exists a row i such that $\tilde{A}_{ij'} = \tilde{A}_{ij''} = 1$. If such a row exists, we have $a_{p_i} = a_{j'} \oplus a_{j''} = 0$. Note that $a_{j'}$ and $a_{j''}$ are the only coefficients that can be freely chosen; hence changing one of

Table 1. Algorithm: Compute $\leftrightarrow$-partition $\mathfrak{P}$.

Input: $\mathcal{R}, \mathcal{B}$ /* relevant cycles and a minimal cycle basis */
Output: $\mathfrak{P}$ /* interchangeability partition */

Sort minimal cycle basis by length: $\{B_1, \ldots, B_\nu\}$.
Sort relevant cycles by length: $\{C_1, \ldots, C_n\}$.
$\mathfrak{P} \leftarrow \emptyset$.
for each cycle length l **do**
  $\mathcal{B}_< \leftarrow \{B \in \mathcal{B} \big| |B| < l\}$.
  $\mathcal{R}_= \leftarrow \{C \in \mathcal{R} \big| |C| = l\}$.
  $A \leftarrow (\mathcal{B}_<, \mathcal{R}_=)$. /* matrix of cycles */
  $\tilde{A} \leftarrow$ reduced row echelon form of A. /* Gauß-Jordan elimination */
  /* Color columns in submatrix $\tilde{B} = (\tilde{A}_{ij})$, $j > |\mathcal{B}_<|$ as in lemma 22. */
  Assign each column $j > |\mathcal{B}_<|$ a different color.
  for each row $i = 1, \ldots, \mathrm{rank}\,(A)$ **do**
    if $\tilde{A}_{ij'} = \tilde{A}_{ij''} = 1$ **then**
      Identify the colors of j' and j''.
  for each color c **do**
    $\mathfrak{P} \leftarrow \mathfrak{P} \cup \{$all cycles with color $c\}$.

them to 0 would require us to change a_{p_i} to 1, i.e., $\vec{a}$ is indeed minimal. Conversely if no such row i exists, set $a_{j'} = 0$. This produces a new solution where all coefficients a_{p_i} with $\tilde{A}_{ip_i} = 1$ and $A_{ij'} = 1$ become zero whereas all others remain unchanged. This contradicts minimality.

Paragraphs (ii) and (iii) together imply that two cycles $C_{j'}, C_{j''} \in \mathcal{R}_=$ are $\leftrightarrow$-equivalent if they have the same color.

(iv) Now suppose two cycles $C_{j'}, C_{j''} \in \mathcal{R}_=$ are $\leftrightarrow$-equivalent but the corresponding columns have different colors. By paragraphs (i) and (ii) we may assume that $j', j'' \in \bar{P}$. There exists a minimal solution of equ.(5) with $a_{j'} = a_{j''} = 1$. Let $\bar{P}_{j'} = \{j \in \bar{P} | a_j = 1, j$ has same color as $j'\}$ and define $\bar{P}_{j''}$ analogously. Then there is no row i, such that $\tilde{A}_{ik'} = \tilde{A}_{ik''} = 1$ for a $k' \in \bar{P}_{j'}$ and $k'' \in \bar{P}_{j''}$. Analogously to paragraph (iii), however, this solution cannot be minimal, a contradiction. $\square$

Table 1 compiles a procedure that uses lemma 22.

Theorem 23. *The $\leftrightarrow$-equivalence classes of the set $\mathcal{R}$ can be computed by algorithm 1 in $\mathcal{O}(|\mathcal{B}|\,|\mathcal{R}|^2\,|E|)$ operations.*

Proof. Sorting $\mathcal{R}$ and $\mathcal{B} \subseteq \mathcal{R}$ by length requires $\mathcal{O}(|\mathcal{R}| \ln |\mathcal{R}|)$ operations. The Gauß-Jordan elimination requires at most $\mathcal{O}(|\mathcal{R}|^2\,|E|)$ operations. Coloring all the $\tilde{B}$'s needs at most $\mathcal{O}(|\mathcal{R}|\,|E|)$ comparisons. Note that this is only a rather crude upper bound for the worst case. The actual requirements are by far smaller for most graphs. $\square$

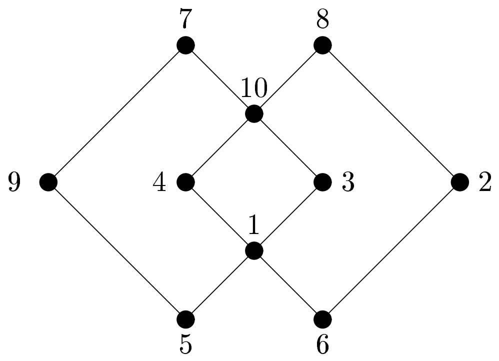


Figure 4. The set of relevant cycles consists of $C_0 = (1, 3, 10, 4)$, $C_1 = (4, 1, 6, 2, 8, 10)$, $C_1' = (3, 1, 6, 2, 8, 10)$, $C_2 = (4, 1, 5, 9, 7, 10)$, and $C_2' = (3, 1, 5, 9, 7, 10)$. We observe $C_0 = C_1 \oplus C_1' = C_2 \oplus C_2'$. Thus $\{C_1, C_1', C_2, C_2'\}$ is an $\leftrightarrow$-equivalence class. Vismara's algorithm identifies C_0,C_1 and C_2 as cycle prototypes. Thus we cannot write $C_1 = C_2 \oplus \bigoplus_{\mathcal{Z}}$ such that $\mathcal{Z}$ contains only cycle prototypes.

We have assumed in algorithm 1 that a minimum cycle basis $\mathcal{B}$ is supplied as input since Vismara's algorithm for computing $\mathcal{R}$ also produces a minimal cycle basis. Of course it could be extracted from the Gauß-Jordan eliminations at virtually no extra cost. Algorithm 1 may require exponential time in terms of $|V|$ since the number of relevant cycles may grow exponentially [15]. Typically, however, there are only $\mathcal{O}(|V|^3)$ relevant cycles [6].

It is not possible to determine $\leftrightarrow$-equivalence with the set $\widehat{\mathcal{R}}$ of "cycle prototypes" that is computed in the first step of Vismara's algorithm. A counterexample is shown in Figure 4.

Appendix

The motivation for the present contribution arises from the search for a suitable energy model for RNA secondary structure computations in the presence of so-called pseudo-knots. A (pseudo-knot free) secondary structure is an outerplanar graph G; hence it has a unique minimum cycle basis $\mathcal{B}(G)$ [10]. The standard energy model for this class of molecules assigns additive energy contributions $E(C)$ to the cycles $C \in \mathcal{B}(G)$, depending on the type of the nucleotides that are represented by the vertices of G [16].

The RNA secondary structure prediction problem can be rephrased as minimizing the energy function $E(G) = \sum_{C\in\mathcal{B}(G)} E(C)$ over the class of secondary structure

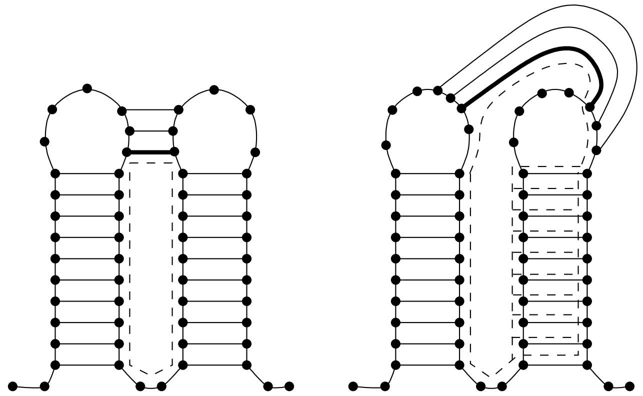


Figure 5. Two RNA secondary structures with a pseudo-knot. The only difference between the two structures is the exact location of the "middle stem" consisting of the three base pairs that connect the two hairpin loops. The energy contribution for the "pseudo-knot formation" should be attributed to the relevant cycle(s) associated with the "closing pair" of the "middle stem", indicated by a thick line.
On the l.h.s. there is a unique relevant cycle (indicated by the dashed line) associated with the "closing pair" of the middle "stem". In the example on the r.h.s. we find ten relevant cycles that differ by the $\oplus$-addition of one or more of the 4-cycles of the rightmost "stem". It seems natural therefore to associate an energy contribution not with an individual relevant cycle, but rather with an equivalence class $\mathcal{W}$ of cycles, in this case with the class of equal length cycles indicated by the dashed lines on the r.h.s.

graphs (which are sub-cubic outerplanar graphs satisfying a few further restrictions, see e.g. [13, 17] for details). Pseudo-knots violate outerplanarity and in general lead to graphs with a non-unique minimum cycle basis, Figure 5. The set $\mathcal{R}$ of relevant cycles, i.e., the union of all minimum cycle bases [15] seems to be a good candidate for extending the energy model. However, as the example in Figure 5 shows, sometimes there is a large class of relevant cycles associated with what biophysically is a single structural element. These are exactly the interchangeable cycles.

It seems natural therefore to average over contributions of interchangeable cycles or to define the energy parameters in such a way that all interchangeable cycles contribute the same energy. Hence we suggest that

$$E(G) = \sum_{\mathcal{W}\in\mathfrak{P}} \frac{\mathrm{knar}\,(\mathcal{W})}{|\mathcal{W}|} \sum_{C\in\mathcal{W}} E(C) \tag{6}$$

serves as a suitable generalization of the standard energy model for nucleic acid structures.

Acknowledgements

This work was supported in part by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung* Proj. 12591-INF and by the *European Commission* within the framework of the Biotechnology Program (BIO-4-98-0189). Input from two anonymous referees is gratefully acknowledged.

References

[1] W.-K. Chen. On vector spaces associated with a graph. *SIAM J. Appl. Math.*, 20:525–529, 1971.

[2] D. M. Chickering, D. Geiger, and D. Heckerman. On finding a cycle basis with a shortest maximal cycle. *Inform. Processing Let.*, 54:55–58, 1994.

[3] L. O. Chua and L. Chen. On optimally sparse cycle and coboundary basis for a linear graph. *IEEE Trans. Circuit Theory*, 20:54–76, 1973.

[4] G. A. Dirac. On rigid circuit graphs. *Abh. Math. Sem. Hamburg*, 25:71–76, 1961.

[5] G. M. Downs, V. J. Gillet, J. D. Holliday, and M. F. Lynch. Review of ring perception algorithms for chemical graphs. *J. Chem. Inf. Comput. Sci.*, 29:172–187, 1989.

[6] P. M. Gleiss and P. F. Stadler. Relevant cycles in biopolymers and random graphs. Presented at the Fourth Slovene International Conference in Graph Theory, Bled, Slovenia, 1999.

[7] D. Hartvigsen and R. Mardon. The all-pair min-cut problem and the minimum cycle basis problem on planar graphs. *SIAM J. Discr. Math.*, 7:403–418, 1994.

[8] J. D. Horton. A polynomial-time algorithm to find the shortest cycle basis of a graph. *SIAM J. Comput.*, 16:359–366, 1987.

[9] A. Kaveh. *Structural Mechanics: Graph and Matrix Methods*. Research Studies Press, Exeter, UK, 1992.

[10] J. Leydold and P. F. Stadler. Minimal cycle basis of outerplanar graphs. *Electr. J. Comb.*, 5:R16, 1998. See `http://www.combinatorics.org`.

[11] J. Oxley. Graphs and series-parallel networks. In N. White, editor, *Theory of Matroids*, volume 26 of *Encyclopedia of Mathematics and Applications*, pages 97–126. Cambridge University Press, Cambridge, UK, 1986.

[12] M. Plotkin. Mathematical basis of ring-finding algorithms in CIDS. *J. Chem. Doc.*, 11:60–63, 1971.

[13] P. F. Stadler and C. Haslinger. RNA structures with pseudo-knots: Graph-theoretical and combinatorial properties. *Bull. Math. Biol.*, 61:437–467, 1999.

[14] C. Thomassen. Embeddings and minors. In R. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics*, pages 301–349. North-Holland, Amsterdam, 1995.

[15] P. Vismara. Union of all the minimum cycle bases of a graph. *Electr. J. Comb.*, 4:73–87, 1997. Paper No. #R9 (15 pages).

[16] A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Müller, D. H. Mathews, and M. Zuker. Co-axial stacking of helixes enhances binding of oligoribonucleotides and improves predicions of RNA folding. *Proc. Natl. Acad. Sci. USA*, 91:9218–9222, 1994.

[17] M. S. Waterman. Secondary structure of single-stranded nucleic acids. *Adv. Math. Suppl. Studies*, 1:167–212, 1978.

[18] D. J. A. Welsh. Matroids: Fundamental concepts. In R. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics*, pages 483–526. North-Holland, Amsterdam, 1995.