A Note On Minimum Path Bases

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

[a]Institute for Theoretical Chemistry and 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

[c]The Santa Fe Institute,
1399 Hyde Park Road, Santa Fe, NM 87501, USA

Abstract

Given an undirected graph $G(V, E)$ and a vertex subset $U \subseteq V$ the U-space is the vector space over GF(2) spanned by the paths with end-points in U and the cycles in $G(V, E)$. We extend Vismara's algorithm to the computation of the union of all minimum length bases of the U-space.

Keywords: Cycle Space, Relevant Cycles and Paths

AMS Subject Classification: 05C38 [paths and cycles].

1. Introduction

Let $G(V, E)$ be a simple graph. A cycle in G is a subgraph of G in which each vertex has even degree. A cycle is *elementary* if all its vertices have degree 2. Thus a cycle is an edge-disjoint union of elementary cycles.

A *uv-path* P, $u \neq v$, in G is a connected subgraph that has exactly two vertices of odd degree, u and v, called its *endnodes*, while all other vertices, called the *interior vertices* of P have even degree. A uv-path is elementary if all its interior vertices have degree 2. One easily checks that a uv-path is an edge-disjoint union of an elementary uv-path and a collection of elementary cycles.

The incidence vector $\vec{H}$ of a subgraph H of G is indexed by the edges of G and has coordinates $H_e = 1$ if e is an edge of H and $H_e = 0$ otherwise. By abuse of notation we henceforth use the same symbol H for a subgraph of G, its edge set, and the corresponding incidence vector. It is customary to consider the vectors H over $GF(2)$. Hence vector addition, $C \oplus D$, corresponds to the symmetric difference of the edge sets of subgraphs C and D of G. The incidence vectors of the cycles span the well-known cycle space $\mathfrak{C}(G)$ of G, see e.g. [1].

The weight $|H|$ of a subgraph H is simply the number of edges in H. The length of a basis $\mathcal{B}$ of a vector space $\mathfrak{V}$ of subgraphs of G is

$$\ell(\mathcal{B}) = \sum_{H \in \mathcal{B}} |H|. \tag{1}$$

We remark that all of the discussion below remains valid when we set $|H| = \sum_{e \in H} \omega(e)$ for arbitrary edge weights $\omega(e) > 0$.

Let $U \subseteq V$ be a nonempty set of vertices and consider the vector space $\mathfrak{U}^*$ generated by the incidence vectors of the uv-paths with $u, v \in U$. This construction is of interest for example in the context of chemical reaction networks, where a subset U of all chemical species V is fed into the system from the outside or is harvested from the system. The uv-paths hence correspond to productive pathways [2, 4, 7]. Hartvigsen [5] introduced the U-space $\mathfrak{U}(G)$ as the union of $\mathfrak{U}^*$ and the cycle space $\mathfrak{C}(G)$. He gives an algorithm for computing a minimum length basis of $\mathfrak{U}(G)$, a *minimum* $\mathfrak{U}$-*basis* for short, in polynomial time that extends a previous algorithm by Horton [6] for minimum length bases of the $\mathfrak{C}(G)$.

More recently, Vismara [8] showed how to compute the set of relevant cycles, i.e., the union of all minimum length bases of $\mathfrak{C}(G)$, using a method that is based on Horton's algorithm. It is the purpose of this note is to extend Vismara's approach to the U-space $\mathfrak{U}(G)$. In addition we briefly describe an implementation of this algorithm.

2. Dimension of the U-space $\mathfrak{U}(G)$

Lemma 1. *If G is biconnected, then $\mathfrak{U}^* = \mathfrak{U}(G)$.*

Proof. Since $\mathfrak{C}(G)$ is spanned by the elementary cycles, it is sufficient to show that any elementary cycle C is the sum of some uv-paths. Let $u, v \in U$. We show that there exist two vertices x and y in C, not necessarily distinct, and two paths D_1 and

D_2 from u to x and from v to y, respectively, such that D_1, D_2, and C are edge-disjoint. We also can split the elementary cycle C into two paths C_1 and C_2 with x and y as their endnodes, such that $C = C_1 \oplus C_2$. In case that some of the points u, v, x, and y coincide, then the corresponding paths above are empty paths, i.e. they have no edges. Now $P_1 = D_1 \oplus C_1 \oplus D_2$ and $P_2 = D_1 \oplus C_2 \oplus D_2$ are two uv-paths with $C = P_1 \oplus P_2$, as proposed.
To construct the paths D_1 and D_2 we start with an elementary cycle D that contains u and a vertex on C. Such a cycle exists by the biconnectedness of G. Obviously, C contains two vertices x and y such that the paths $D_1 = D[u, x]$ and $D_2 = D[u, y]$ are edge-disjoint and have no edge in common with C. Furthermore, since G is connected, there must be a path H from v to some vertex y' in C that also has no edge in common with C. If D_1 and H are edge-disjoint we can replace y by y' and D_2 by H and we are done. Analogously we replace D_1 by H if D_2 and H are edge-disjoint but D_1 and H are not. Otherwise D_2 and H must have a vertex h in common such that the subpath $H[v, h]$ has no edge in common with the subpath $D_2[h, y]$ and D_1 (otherwise we change the rôle of D_1 and D_2). Thus $D_2' = H[v, h] \oplus D_2[h, y]$ is a path from v to y that has no edge in common with D_1 and C. Hence we replace D_2 by D_2' and we are done. □

Remark. Notice that in the above proof u and v need not be distinct. As a consequence $\mathfrak{C}(G)$ is spanned by all cycles through a given vertex $u \in U$, provide that G is biconnected. It is therefore meaningful to extend the definition of $\mathfrak{U}^*$ to the special case $|U| = 1$ where $\mathfrak{U}^*$ is the cycle space of the biconnected component that contains $u \in U$. One could therefore define a uu-path as a connected cycle through the vertex u. However this would complicate the notation in the following.

Lemma 1 is not necessarily true for general connected graphs. Extending the argument of the proof above it is easy to see, however, that for each biconnected component H of G we have either $\mathfrak{C}(H) \subseteq \mathfrak{U}^*$ or $\mathfrak{C}(H) \cap \mathfrak{U}^* = \emptyset$, depending on whether a U-path passes through H.

The dimension of the cycle space is the cyclomatic number $\mu(G) = |E| - |V| + 1$ (for connected graphs). The dimension of the U-space $\dim(\mathfrak{U})$ can be given as following:

Theorem 2. *If G is connected then* $\dim(\mathfrak{U}(G)) = \mu(G) + |U| - 1$,

Proof. Let $C = C_1 \oplus C_2 \oplus \cdots \oplus C_k$, with $C_i \in \mathfrak{U}(G)$. Then for any vertex $x \in V$ the degree of x in C is even if and only if $\sum_{i=1}^{k} \deg_{C_i}(x)$ is even. In particular, the $\oplus$-sum of two paths between two vertices x and y is a cycle.
We proceed by induction on the number of vertices in U. Assume U contains only the two vertices x and y. To construct a basis for $\mathfrak{U}$, we need a path $P(x, y)$ in addition to the cycle basis, since all paths between x and y are obtained as $\oplus$-sums of the path $P(x, y)$ and some cycles. Hence $\dim(\mathfrak{U}) = \mu + 1$.
Now assume the proposition holds for $U \subset V$ and consider $U' = U \cup \{v\}$ for some $v \in V \setminus U$. Since there is no path with endpoint v in the basis of $\mathfrak{U}$ the degree of v is even for every $\oplus$-sum of elements in $\mathfrak{U}$. Thus $\dim(\mathfrak{U}') > \dim(\mathfrak{U})$. To obtain a basis for $\mathfrak{U}'$ we have to add a path $P(v, x)$ for some $x \in U$ to the basis of $\mathfrak{U}$. Clearly, for any $y \in U$, $P(v, x) \oplus P(x, y)$ is the edge-disjoint union of a path $P(v, y)$ and a (possibly

empty) collection of elementary cycles. All other paths from v to $y \in U$ can now be obtained as the $\oplus$-sum of the path $P(v, y)$ and an appropriate set of cycles. Hence $\dim(\mathfrak{U}') = \dim(\mathfrak{U}) + 1$ and the proposition follows. □

We immediately find the following

Corollary 3. *If G is a simple connected graph G and $U \subseteq V$ is non-empty, then* $\dim(\mathfrak{U}) = |E| - |V| + |U|$.

Notice that this result also holds for graphs G that are not connected provided that each component of G contains at least one vertex of U.

3. Minimal $\mathfrak{U}$-Bases and Relevant $\mathfrak{U}$-Elements

Definition 4. *Let $\mathfrak{V}$ be a vector space of subgraphs of G. We say that A is* relevant *in $\mathfrak{V}$ (for short $\mathfrak{V}$-relevant) if there is a minimum length basis $\mathcal{B}$ of $\mathfrak{V}$ such that $A \in \mathcal{B}$.*

In other words, the set $\mathcal{R}_{\mathfrak{V}}$ of $\mathfrak{V}$-relevant subgraphs is the union of all minimum length bases of $\mathfrak{V}$.

Lemma 5. *$A \in \mathfrak{V}$ is relevant if and only if A cannot be written as the $\oplus$-sum of strictly shorter elements of $\mathfrak{V}$.*

Proof. The proof of Vismara's [8] Lemma 1 is valid for arbitrary vector spaces of subgraphs. □

Horton's [6] mimimal cycle basis algorithm is based on an easy-to-check necessary condition for relevance: A cycle is *edge-short* if it contains an edge $e = \{x, y\}$ and a vertex z such that $C^{xy,z} = \{x, y\} \cup P_{xz} \cup P_{yz}$ where P_{xz} and P_{yz} are shortest paths[1]. Hartvigsen [5] generalized this notion to paths: A uv-path P is *edge-short* if there is an edge $e = \{x, y\}$ such that both $P[u, x]$ is a shortest ux-path and $P[y, v]$ is shortest yv-path. Here we write $P[p, q]$ for the subpath of P connecting p and q. Horton and Hartvigsen furthermore showed that it is sufficient to consider the cycles $C^{xy,z}$ and paths $P^{x,y}_{uv} = P_{ux} \cup \{x, y\} \cup P_{yv}$ for a *fixed* choice of the shortest paths P_{xy} between any two vertices of G. Thus a minimum cycle basis and a minimum $\mathfrak{U}$-basis can be obtained in polynomial time by means of the greedy algorithm, see [5, 6].

The related problem of computing all relevant cycles or $\mathfrak{U}$-elements can in general not be solved in polynomial time because the number of relevant cycles may grow exponentially with $|V|$ in some graph families, for an example see [8, Fig.2]. It is possible, however, to define a set of *prototypes* for the relevant cycles such that each relevant cycle C can be represented in the form

$$C = C^p \oplus S_1 \oplus S_2 \oplus \cdots \oplus S_k \tag{2}$$

where C^p is a prototype cycle, with $|C^p| = |C|$ and cycles S_i that are strictly shorter than C. One easily verifies that either both C and C^p are relevant or neither cycle is relevant. Furthermore, a minimal cycle basis contains at most one of these two cycles. We briefly recall Vismara's construction of prototypes for cycles and then extend it to U-paths.

[1] We reserve the symbol P_{xy} for **a** shortest path between x and y, while $P(x, y)$ may be any path between x and y, and $P[x, y]$ denotes the sub-path from x to y of a given path or cycle P.

We fix an arbitrary ordering of the vertex set of G. Consider an edge-short cycle C such that r is the largest vertex in C. If there is a vertex x in C such that C consists of two different shortest paths from x to r we say that C is even-balanced. The two vertices adjacent to x will be denoted by p and q. We write C_r^{pxq} for this situation. If C contains an edge $\{p, q\}$ such that $|P_{pr}|, |P_{qr}| < |C|/2$ then C is *odd-balanced* and we write C_r^{pq}. The cycle family associated with a balanced cycle C consists of all those cycles that share with C the vertex r, the edge $\{p, q\}$ or the path (p, x, q), respectively, and that contains shortest paths P_{pr} and P_{qr} such that each vertex in P_{pr} and P_{qr} preceeds r in the given ordering. Vismara shows that the members of a cycle family are related by equ.(2) and that the relevant cycle families form a partition of the set of relevant cycles. Any minimal cycle basis contains at most one representative from each cycle family.

Analogously, we now introduce balanced uv-paths in the following way: An edge-short uv-path P is *even-balanced* if there is a vertex w in P such that $|P[u, w]| = |P[w, v]|$ and $P[u, w]$ and $P[v, w]$ are shortest uw- and wv-paths, respectively. P is *odd-balanced* if there is an edge $e = \{x, y\} \in P$ such that $|P[u, x]| < \frac{1}{2}|P|$ and $|P[v, y]| < \frac{1}{2}|P|$, and $P[u, x]$ and $P[v, y]$ are shortest ux- and vy-paths respectively.

Theorem 6. *Any relevant U-path P consists of two disjoint shortest paths $P[u, x]$ and $P[v, y]$ linked by the edge $\{x, y\}$ if P is odd-balanced or by the path (x, w, y) if P is even-balanced.*

Proof. We know that P must be edge-short if it is relevant. Suppose P is edge-short but not balanced.

In the even case, let w be the vertex in P such that $|P[u, w]| = |P[w, v]|$. Since P is not balanced either $P[u, w]$ or $P[v, w]$ is not a shortest path. W.l.o.g. we assume that $P[u, w]$ is not uw-shortest. Let Q be a uw-shortest path. Then $|Q| < |P[u, w]| = \frac{1}{2}|P|$. Set $C = Q \oplus P[u, w]$. Clearly C is a cycle or an edge disjoint union of cycles and $|C| \leq |Q| + |P[u, w]| < |P|$. Now consider the path $P' = Q \oplus P[w, v]$; it is an edge-short U-path satisfying $|P'| \leq |Q| + |P[w, v]| < |P[u, w]| + |P[w, v]| = |P|$. We have

$$P = P[u, w] \oplus P[w, v] = P[u, w] \oplus Q \oplus Q \oplus P[w, v] = C \oplus P' \quad (3)$$

Hence P can be written as an $\oplus$-sum of strictly shorter elements of the U-space, and therefore it cannot be relevant.

In the odd case, there is an edge $\{x, y\}$ in P such that both $|P[u, x]| < \frac{1}{2}|P|$ and $|P[y, v]| < \frac{1}{2}|P|$. Since P is not balanced either $P[u, x]$ or $P[v, y]$ is not shortest. W.l.o.g. we assume that $P[u, x]$ is not ux-shortest, and consider a shortest ux-path Q. In this case we have $|Q| < |P[u, x]| < \frac{1}{2}|P|$.

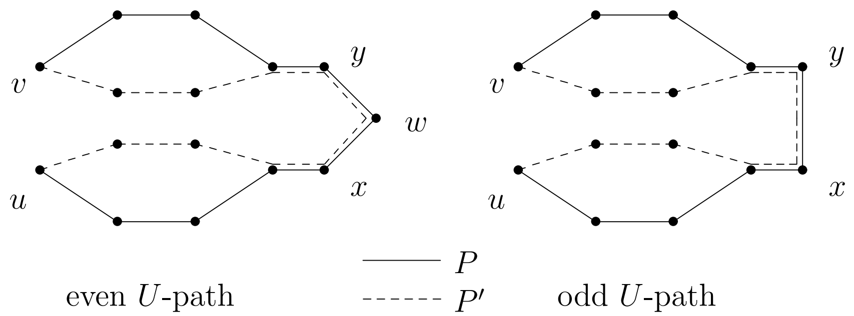
Set $C = Q \oplus P[u, x]$. Clearly C is a cycle or an edge disjoint union of cycles and $|C| \leq |Q| + |P[u, x]| < |P|$. Now consider the path $P' = Q \oplus \{x, y\} \oplus P[y, v]$; it is a short U-path satisfying $|P'| \leq |Q| + |\{x, y\}| + |P[y, v]| < |P[u, x]| + |\{x, y\}| + |P[y, v]| = |P|$. We have

$$P = P[u, x] \oplus \{x, y\} \oplus P[y, v] = P[u, x] \oplus Q \oplus Q \oplus \{x, y\} \oplus P[y, v] = C \oplus P' \quad (4)$$

Hence again P can be written as an $\oplus$-sum of strictly shorter elements of the U-space, and therefore it is not relevant. $\square$

We are now in the position to construct *prototypes* of relevant U-paths in the same manner as Vismara's prototypes of relevant cycles. For any relevant U-path P including the vertices x, y and eventually w, as defined in Theorem 6, we define the *U-path family* associated with P as follows:

Definition 7. *The U-path family $\mathcal{F}(P)$ belonging to the prototype P is the set of all balanced U-paths P' such that $|P'| = |P|$ and P' consists of the vertices u and v, the edge $\{x, y\}$ or the path (x, w, y) and two shortest paths P_{ux} and P_{vy}.*



Hence, two U-paths P and P' belonging to the same family $\mathcal{F}(P)$ differ only by the shortest paths from u to x and/or from v to y that they include. Consequently, $P = P' \oplus S_1 \oplus S_2 \oplus \cdots \oplus S_k$ where the S_j are cycles (or edge-disjoint unions of cycles).

Theorem 8. *Each relevant $\mathfrak{U}$-element belongs to exactly one U-path family or cycle family.*

Proof. By construction, cycles and paths belong to different families. The proof that the cycles families form a partition of the set of relevant cycles is given in [8].
Each relevant path is either even-balanced or odd-balanced and therefore belongs to the U-path family that is characterized by the end-vertices u and v and the "middle part" (x, w, y) or $\{x, y\}$, respectively. $\square$

The relevant $\mathfrak{U}$-elements can be computed using the two-stage approach proposed by Vismara [8]. In the first step a set of prototypes is extracted by means of the greedy procedure from candidate set with a polynomial number of cycles. Algorithm 1 is a straightforward extension of Vismara's approach. We have to add Algorithm 2 in order to include all potential path prototypes; the following greedy step on the collection of all balanced cycles and U-paths remains unchanged. Vismara [8] showed that the relevant cycle families can be computed in $\mathcal{O}(|E|^2|V|)$ steps. There are at most $|U|^2|E|$ families of relevant U-paths, hence the algorithm remains polynomial.

In the second part the relevant U-elements are extracted by means of a recursive backtracking scheme. For each cycle or path prototype, C_r^{pq} or P_{uv}^{xy}, we have to replace the paths $C_r^{pq}[p, r]$ and $C_r^{pq}[q, r]$ or $P_{uv}^{xy}[u, x]$ and $P_{uv}^{xy}[v, y]$ by all possible alternative paths with the same length. These can be generated using the recursive function `List_Paths()` from [8] for the cycles (where we have to obey additional constraints

Algorithm 1 Relevant $\mathfrak{U}$-elements.

Input: Connected graph G.
1: Compute shortest path P_{uv} for all $u, v \in V$.
2: Compute cycle prototypes C_r^{pq} and C_r^{pxq}, see [8], and store in $\mathcal{P}$.
3: Compute path prototypes P_{uv}^{xy}, P_{uv}^{xwy} (Algorithm 2), and store in $\mathcal{P}$.
4: Sort $\mathcal{P}$ by length and set $\hat{\mathcal{R}} = \emptyset$.
5: For each length k, check if $Q \in \mathcal{P}$ with $|Q| = k$ is independent of all shorter elements in $\hat{\mathcal{R}}$. If yes, add Q to $\hat{\mathcal{R}}$.
6: List all relevant U-elements by recursive backtracking from $\hat{\mathcal{R}}$.

In practice one checks linear independence only against a partial minimal basis.

Algorithm 2 Prototypes for Relevant U-paths.

```
for all (u, v) ∈ U do
  /* calculate even prototypes: */
  for all w ∈ V do
    if |P_uw| = |P_vw| then
      for all x ∈ V adjacent to w do
        for all y ∈ V adjacent to w do
          if |P_ux| + |P_xw| = |P_uw| and |P_vy| + |P_yw| = |P_vw| then
            P_uv^xwy = P_ux ⊕ {x, w} ⊕ {w, y} ⊕ P_yv
  /* calculate odd prototypes: */
  for all e = {x, y} ∈ E do
    if |P_ux|, |P_vy| < (|P_ux| + |P_xy| + |P_yv|) then
      P_uv^xy = P_ux ⊕ {x, y} ⊕ P_yv
    if |P_uy|, |P_vx| < (|P_uy| + |P_yx| + |P_xv|) then
      P_uv^xy = P_uy ⊕ {y, x} ⊕ P_xv
```

that r is the vertex with largest index in the given ordering) and an analogous function (without constraint) for the U-paths.

4. Exchangeability of $\mathfrak{U}$-Elements

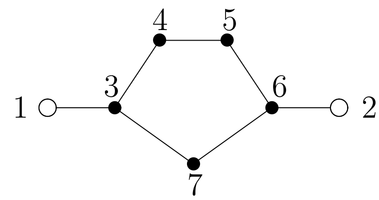
In [3] a partition of the set of relevant cycles is introduced that is coarser than Vismara's cycle families. This construction generalizes directly to the U-space:

Definition 9. *Two relevant $\mathfrak{U}$-elements $C', C'' \in \mathcal{R}_{\mathfrak{U}}$ are* interchangeable*, $C' \leftrightarrow C''$, if (i) $|C'| = |C''|$ and (ii) there exists a minimal linearly dependent set of relevant $\mathfrak{U}$-elements that contains C' and C'' and with each of its elements not longer than C'.*

Interchangeability is an equivalence relation. The theory developed in [3] does not depend on the fact that one considers cycles; indeed it works for all finite vector spaces over $GF(2)$ and hence in particular for U-spaces. Hence we have the following

Proposition 10. *Let $\mathcal{B}$ be a minimum length $\mathfrak{U}$-basis and let $\mathcal{W}$ be a $\leftrightarrow$-equivalence class of relevant $\mathfrak{U}$-elements. Then $|\mathcal{W} \cap \mathcal{B}|$ is independent of the choice of the basis $\mathcal{B}$.*

The quantity $\mathrm{knar}(\mathcal{W}) = |\mathcal{W} \cap \mathcal{B}|$ has been termed the *relative rank* of the equivalence class $\mathcal{W}$ in [3]. It is tempting to speculate that the $\leftrightarrow$-partition might distinguish between cycles and paths. As the example below shows, however, this is not the case:



Here $U = \{1, 2\}$ and the relevant $\mathfrak{U}$-elements are the paths $P_1 = (1, 3, 7, 6, 3)$, $P_2 = (1, 3, 4, 5, 6, 2)$, and the cycle $C = (3, 4, 5, 6, 7, 3)$. with $|P_1| = 4$ and $|P_2| = |C| = 5$. Furthermore $C = P_2 \oplus P_1$, i.e., the path P_2 and the cycle C belong to the same $\leftrightarrow$-equivalence class.

Acknowledgements. This work was supported in part by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung* Proj. 14094-MAT.

References

[1] B. Bollobás. *Modern Graph Theory.* Springer, New York, 1998.

[2] D. A. Fell. *Understanding the Control of Metabolism.* Portland Press, London, 1997.

[3] P. M. Gleiss, J. Leydold, and P. F. Stadler. Interchangeability of relevant cycles in graphs. *Elec. J. Comb.*, 7:R16 [16pages], 2000.

[4] P. M. Gleiss, P. F. Stadler, A. Wagner, and D. A. Fell. Relevant cycles in chemical reaction network. *Adv. Complex Syst.*, 4, 2001. in press.

[5] D. Hartvigsen. Minimum path bases. *J. Algorithms*, 15:125–142, 1993.

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

[7] O. N. Temkin, A. V. Zeigarnik, and D. Bonchev. *Chemical Reaction Networks: A Graph-Theoretical Approach.* CRC Press, Bocal Raton, FL, 1996.

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