Minimum Path Bases and Relevant Paths

Petra M. Gleiss
Institute for Theoretical Chemistry, University of Vienna, Währingerstrasse 17, A-1090 Vienna, Austria

Josef Leydold
Department for Applied Statistics and Data Processing, University of Economics and Business Administration, Augasse 2-6, A-1090 Wien, Austria

Peter F. Stadler
Bioinformatics Group, Department of Computer Science, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany

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. Although the size distribution of subgraphs is the same in all minimum length bases, the number of cycles and paths may differ. **© 2005 Wiley Periodicals, Inc.** **NETWORKS, Vol. 46(3), 119–123 2005**

Keywords: cycle space; minimum cycle basis; relevant cycles and paths

1. INTRODUCTION

In recent years, the structure of a many diverse large "real-world" graphs—including those of social and technological networks, of the metabolic and regulatory networks of a cell, and chemical reaction networks—has been elucidated experimentally, see, for example, [6]. This has led to a renewed interest in efficient algorithms for computing graph properties that can be used to characterize and classify large network graphs [1].

Although most work concentrates on the path structure (which relates directly to transport phenomena in the network), the cycle structure is equally important for an understanding of robustness and resilience because cycles determine the "detours" when edges or vertices are removed [12]. For example, in [4], the distribution of cycle lengths is used to discriminate the off-equilibrium scale-free networks from equilibrium models. In this context short cycles, that is, short detours, are of particular interest because they convey information on the local ordering in complex networks. Data from real-world networks, for instance, reveal the presence of locally grid-like structures through the frequency of quadrangles [7].

From a mathematical point of view, it is natural to consider minimum cycle bases. They have direct applications in diverse areas of science and engineering, such as the static analysis of rigid framework structure [16], chemical structure storage and retrieval systems [2, 9], and electric circuit theory [8]. Flux Analysis in chemical reaction networks [10, 20] considers stationary fluxes as kernel vectors of the stoichiometric matrix, which in the case of a network of isomerization reactions reduces to the incidence matrix of a directed graph. Stationary "flux modes" then reduce to cycles in this graphs [12].

Often, a network contains a special subset U of vertices that interact with the environment. In such a setting it becomes natural to consider those paths that connect vertices in U with each other in addition to cycles. This becomes particularly apparent for metabolic networks: The nodes in U represent food and waste materials; stationary flux modes then correspond to both cycles and certain paths that begin and end in U. The latter form the so-called biochemical pathways on which our understanding of a cell's metabolism is based. In the case of undirected graphs, Hartvigsen's concept of U-path bases [13] adds exactly these paths to the cycle space.

A disadvantage of minimum length cycle and path bases, however, is the fact that they are in general not unique. Using the union of all minimum cycle bases solves this problem in many cases. In 1997, Vismara [22] published an efficient

Received October 2003; accepted June 2005
Correspondence to: P. F. Stadler; e-mail: studla@bioinf.uni-leipzig.de
Contract grant sponsor: Austrian Fonds zur Förderung der Wissenschaftlichen Forschung; Contract grant number: 14094-MAT
Contract grant sponsor: Bioinformatics Initiative of the Deutsche Forschungs Gemeinschaft; Contract grant number: BIZ–6/1-2
DOI 10.1002/net.20080
Published online in Wiley InterScience (www.interscience.wiley.com).
© 2005 Wiley Periodicals, Inc.

algorithm for computing this set of relevant cycles. In this contribution we extend Vismara's algorithm to computing the union of all minimum bases of the U-path space and discuss some simple properties of this set.

2. PRELIMINARIES

Let $G(V, E)$ be a simple graph with vertex set V and edge set E. A cycle in G is a subgraph of G in which each vertex has even degree. A cycle is *elementary* if it is connected and 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. If a uv-path is not elementary, then removing an elementary uv-path from it leaves only vertices of even degree, and hence, a cycle. Thus, a uv-path is an edge-disjoint union of an elementary uv-path and a collection of elementary cycles.

We reserve the symbol P_{xy} for **a** shortest path between x and y, while $P[x, y]$ denotes the subpath from x to y of a given elementary path or cycle P.

The incidence vector $\bar{H}$ 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. We use the same symbol H for a subgraph of G, its edge set, and the corresponding incidence vector. It is customary to consider H as a vector 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, for example, [5].

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 $\omega(e) > 0$.

3. *U*-PATH SPACES

Let $U \subseteq V$ be a nonempty set of vertices, $|U| \geq 2$, and consider the vector space $\mathfrak{U}^*$ generated by the incidence vectors of the uv-paths, for all $u, v \in U$. 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 [10, 12, 21]. Hartvigsen [13] 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 [15] for minimum length bases of the cycle space $\mathfrak{C}(G)$. The following result shows that these two vector spaces are equal on proper graphs.

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

Proof. Let $u, v \in U$. Because $\mathfrak{C}(G)$ is spanned by the elementary cycles, it is sufficient to show that any elementary cycle C is the sum of some uv-paths. 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 can also split the elementary cycle C into two paths, C_1 and C_2, with x and y as their endnodes, such that $C = C_1 \oplus C_2$. If some of the points u, v, x, and y coincide, then some of these paths are empty, that is, 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 in 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, because 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 sub-path $H[v, h]$ has no edge in common with the subpaths $D_2[h, y]$ 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 $u \in U$, provided that G is biconnected. It is therefore meaningful to extend the definition of $\mathfrak{U}^*$ to the special case $|U| = 1$. Then $\mathfrak{U}^*$ 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.*

Remark. *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 U-path space $\mathfrak{U}(G)$ of a graph $G(V, E)$ can be transformed into the cycle space of a graph $G^*(V^*, E^*)$ that is constructed by adding a vertex z^* and edges $\{z^*, u\}$ of weight zero for each $u \in U$. U-paths then correspond to cycles through z^*: every uv-path can be transformed into a cycle in G^* by adding the path (u, z^*, v); conversely, by deleting z^*

every cycle containing z^* becomes the union of cycles and uv-paths in G where u, v are neighbors of z^*, that is, a U-path. The length of a uv-path is obviously equal to the length of the corresponding cycle. As an immediate consequence we get the dimension of the U-path space. The dimension of the cycle space of a connected graph is the cyclomatic number $\mu(G) = |E|-|V|+1$. Hence, the dimension of the U-space is given by $\mu(G^*) = |E^*|-|V^*|+1 = (|E|+|U|)-(|V|+1)+1$, see also [14, Prop. 6.5].

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

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. In [14], a more general model is considered, where U_i-paths connect endpoints within one or more disjoint subsets $U_i \subset V$. We may deal with this construction in the same way as above by adding a distinct extra vertex for each set U_i.

4. MINIMUM $\mathfrak{U}$-BASES AND RELEVANT $\mathfrak{U}$-ELEMENTS

Definition 3. *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 of $\mathfrak{V}$-relevant subgraphs is the union of all minimum length bases of $\mathfrak{V}$.

Lemma 4 (Vismara [22]). *$A \in \mathfrak{V}$ is relevant if and only if A cannot be written as the $\oplus$-sum of strictly shorter elements of $\mathfrak{V}$.*

Horton's [15] minimum cycle basis algorithm is based on an easy-to-check necessary condition for relevance: A cycle C is *edge-short* if it contains an edge $e = \{x, y\}$ and a vertex z such that $C = C^{xy,z} := \{x, y\} \cup P_{xz} \cup P_{yz}$, where P_{xz} and P_{yz} are shortest paths. Hartvigsen [13] generalized this notion to paths: A uv-path P is *edge-short* if there is an edge $e = \{x, y\}$ in P such that both $P[u, x]$ is a shortest ux-path and $P[y, v]$ is shortest yv-path. Horton and Hartvigsen furthermore showed that it is sufficient to consider the cycles $C^{xy,z}$ and paths $P_{uv}^{x,y} = P_{ux} \cup \{x, y\} \cup P_{yv}$ 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 [13, 15].

A different approach based on the exchange property of matroid is pursued in [3, 18]. In the case of directed graphs it appears oftentimes more appropriate to consider linear independence over $\mathbb{Q}$ instead of GF(2). In this setting, the minimum cycle basis problem can be solved by analogs of both Horton's and Berger's approaches, see [17, 19].

The related problem of computing all relevant cycles or $\mathfrak{U}$-elements cannot, in general, be solved in polynomial time because the number of relevant cycles may grow exponentially with $|V|$ in some graph families; see [22, Fig. 2] for an example. It is possible, however, to define a set of *prototypes* for the relevant cycles that grows only polynomially with $|V|$. Each relevant cycle C can be then 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. Lemma 4 shows that either both C and C^p are relevant or neither cycle is relevant. Furthermore, a minimum cycle basis contains at most one of these two cycles. Vismara [22] describes a construction of prototypes for cycles:

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 w.r.t. this ordering. 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 in C are 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} precedes r in the given ordering. Vismara shows that the members of a cycle family are related by (2), and that the relevant cycle families form a partition of the set of relevant cycles. Any minimum cycle basis contains at most one representative from each cycle family.

This idea can be extended to U-paths. Thus, the notion of *even-balanced* and *odd-balanced* paths is completely analogous, and we define the *U-path family* associated with a prototype P as follows:

Definition 5. *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); see Figure 1.

The relevant $\mathfrak{U}$-elements can be computed using the two-stage approach proposed by Vismara [22]. This can be done

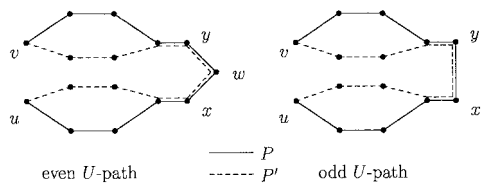


FIG. 1. Two U-paths P and P' belonging to the same family $\mathcal{F}(P)$ differ only by the shortest paths from u to x and/or from v to y that they include.

by solving the equivalent Minimum Cycle Basis Problem for the extended graph $G^*(V^*, E^*)$, but with fixing the additional vertex z^* as the largest element in the arbitrary ordering. We do not need to consider all balanced U-paths. It is sufficient to consider those that are obtained from balanced cycles through z^*. The U-path prototypes are then obtained by deleting z^*. The algorithm below compiles the necessary steps.

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

Input: Connected graph $G(V, E)$, set U.

1: Create extended graph $G^*(V^*, E^*) = (V \cup \{z^*\}, E \cup \bigcup_{u \in U}\{\{z^*, u\}\})$.
2: Fix ordering of vertices with z^* as largest element.
3: Compute shortest path P_{uv} for all $u, v \in V^*$.
4: Compute cycle prototypes C_r^{pq} and C_r^{pxq}, see [22], and store in $\mathcal{P}^*$.
5: Sort $\mathcal{P}^*$ by length and set $\hat{\mathcal{R}} = \emptyset$.
6: 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 $\hat{\mathcal{R}}$.
7: Replace each prototype in $\hat{\mathcal{R}}$ that contains z^* by corresponding U-path.
8: List all relevant U-elements by recursive backtracking from $\hat{\mathcal{R}}$.

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

Vismara [22] showed that the prototypes for the relevant cycles can be computed in $\mathcal{O}(\mu(G^*)|E^*|^3)$ steps. Because $|U| \le |V| \le |E|$, this is the same as $\mathcal{O}((\mu(G) + |U|)|E|^3)$, that is, the inclusion of the U-paths does not substantially increase the complexity.

In the second part the relevant U-elements are extracted by means of a recursive backtracking scheme. For each cycle prototype C_r^{pq} (which corresponds to a U-path iff $r = z^*$) we have to replace the paths $C_r^{pq}[p, r]$ and $C_r^{pq}[q, r]$ by all possible alternative paths with the same length. These can be generated using the recursive function `List_Paths()` from [22], where we have to obey additional constraints that r is the vertex with largest index in the given ordering.

5. EXCHANGEABILITY OF $\mathfrak{U}$-ELEMENTS

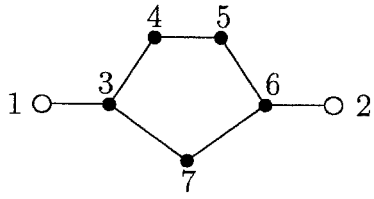
In [11], a partition of the set of relevant cycles is introduced that is coarser than Vismara's cycle families. This construction generalizes directly to the U-space:

Definition 6. *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 minimum linearly dependent set of relevant $\mathfrak{U}$-elements that contains C' and C'' and has no element that is longer than C'.*

Interchangeability is an equivalence relation. The theory developed in [11] 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. Thus, we have the following.

Proposition 7. *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* $\mathrm{knar}(\mathcal{W}) = |\mathcal{W} \cap \mathcal{B}|$ *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 [11]. 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 $P_1 = (1, 3, 7, 6, 2)$, $P_2 = (1, 3, 4, 5, 6, 2)$, and the cycle $C = (3, 4, 5, 6, 7, 3)$ with $|P_1| = 4$ and $|P_2| = |C| = 5$. Furthermore, $C = P_2 \oplus P_1$, that is, the path P_2 and the cycle C belong to the same $\leftrightarrow$-equivalence class.

ACKNOWLEDGMENTS

We thank two anonymous referees for their valuable comments, which substantially streamlined our arguments.

REFERENCES

[1] L.A.N. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley, Classes of small world networks, Proc Natl Acad Sci USA 97 (2000), 11149–11152.

[2] F. Berger, C. Flamm, P.M. Gleiss, J. Leydold, and P.F. Stadler, Counterexamples in chemical ring perception, J Chem Inf Comput Sci 44 (2004), 323–331.

[3] F. Berger, P. Gritzmann, and S. de Vries, Minimum cycle bases for network graphs, Algorithmica 40 (2004), 51–62.

[4] G. Bianconi, Number of cycles in off-equilibrium scale-free networks and in the internet at the autonomous system level, Eur Phys J B 38 (2004), 223–230.

[5] B. Bollobás, Modern graph theory, Springer, New York, 1998.

[6] S. Bornholdt and H.G. Schuster (Editors), Handbook of graphs and networks: From the genome to the internet, Wiley-VCH, Berlin, 2002.

[7] G. Caldarelli, R. Pastor-Satorras, and A. Vespignani, Structure of cycles and local ordering in complex networks, Eur Phys J B 38 (2004), 183–186.

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

[9] 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 (1989), 172–187.

[10] D.A. Fell, Understanding the control of metabolism, Portland Press, London, 1997.

[11] P.M. Gleiss, J. Leydold, and P.F. Stadler, Interchangeability of relevant cycles in graphs, Elec J Comb 7 (2000), R16.

[12] P.M. Gleiss, P.F. Stadler, A. Wagner, and D.A. Fell, Relevant cycles in chemical reaction networks, Adv Complex Syst 4 (2001), 207–226.

[13] D. Hartvigsen, Minimum path bases, J Algorithms 15 (1993), 125–142.

[14] D. Hartvigsen, The planar multiterminal cut problem, Discr Appl Math 85 (1998), 203–222.

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

[16] A. Kaveh, Structural mechanics: Graph and matrix methods, Research Studies Press, Exeter, UK, 1992.

[17] T. Kavitha and K. Mehlhorn, A polynomial time algorithm for minimum cycle basis in directed graphs, "STACS 2005", Lecture Notes in Computer Science, Springer, Heidelberg, (2004) vol. 3404, pp. 654–665.

[18] T. Kavitha, K. Mehlhorn, D. Michail, and K. Paluch, A faster algorithm for minimum cycle basis of graphs, "ICALP 2004," Lecture Notes in Computer Science, Springer, Heidelberg, (2004) vol. 3142, pp. 846–857.

[19] C. Liebchen and R. Rizzi, A greedy approach to compute a minimum cycle bases of a directed graph, Inf Process Let 94 (2005), 107–112.

[20] C.H. Schilling, D. Letscher, and B.O. Palsson, Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from pathway-oriented perspective, J Theor Biol 203 (2000), 229–248.

[21] O.N. Temkin, A.V. Zeigarnik, and D. Bonchev, Chemical reaction networks: A graph-theoretical approach, CRC Press, Boca Raton, FL, 1996.

[22] P. Vismara, Union of all the minimum cycle bases of a graph, Electr J Comb 4 (1997), 73–87.