Edge-outer graph embedding and the complexity of the DNA reporter strand problem

M. N. Ellingham[a,1,], Joanna A. Ellis-Monaghan[b,2]

[a]*Department of Mathematics, 1326 Stevenson Center, Vanderbilt University, Nashville, Tennessee 37240*
[b]*Department of Mathematics, Saint Michael's College, One Winooski Park, Colchester, Vermont 05458*

Abstract

In 2009, Jonoska, Seeman and Wu showed that every graph admits a route for a DNA reporter strand, that is, a closed walk covering every edge either once or twice, in opposite directions if twice, and passing through each vertex in a particular way. This corresponds to showing that every graph has an *edge-outer embedding*, that is, an orientable embedding with some face that is incident with every edge. In the motivating application, the objective is such a closed walk of minimum length. Here we give a short algorithmic proof of the original existence result, and also prove that finding a shortest length solution is NP-hard, even for 3-connected cubic (3-regular) planar graphs. Independent of the motivating application, this problem opens a new direction in the study of graph embeddings, and we suggest new problems emerging from it.

Keywords: DNA origami, reporter strand, orientable graph embedding, one-face embedding, edge-outer embedding
2000 MSC: 05C10, 05C90, 68Q17, 92D10

---

1. Introduction

DNA self-assembly, and self-assembly in general, is a rapidly advancing field, with [13, 16] providing good overviews. In 2006, Rothemund introduced 'DNA origami', a new self-assembly method that increased the scale of DNA constructs and is one of the major developments in DNA nanotechnology this century [15]. It originally involved combining an M13 single-stranded cyclic viral molecule, called the *scaffolding strand*, with 200-250 short staple strands to produce a 90×90 nm tile (in 2D), but now these strands can also produce

Email addresses: `mark.ellingham@vanderbilt.edu` (M. N. Ellingham), `jellis-monaghan@smcvt.edu` (Joanna A. Ellis-Monaghan)
URL: `https://math.vanderbilt.edu/ellingmn/` (M. N. Ellingham), `http://www.smcvt.edu/pages/get-to-know-us/faculty/ellis-monaghan-jo.aspx` (Joanna A. Ellis-Monaghan)
[1]Supported by Simons Foundation award 429625
[2]Supported by NSF grant DMS-1332411

3D constructs with the structure of graphs or graph fragments [1]. At its most basic level, the design objective for DNA origami assembly of a graph-like structure is a strategy with the scaffolding strand following a single walk that traverses every edge at least once, with any edges that are traversed more than once visited exactly twice, in opposite directions (because DNA strands in a double helix are oppositely directed), and without separating or crossing through at a vertex. See [3, 5] for further work on routing scaffolding strands.

The problem of finding a similarly prescribed walk arises in the context of determining an efficient route for a *reporter strand*, that is, a strand that is recovered and read at the end of an experiment to report on the result of the assembly. In designing the DNA self-assembly of a molecule with the structure of a graph G, the boundary components of a 'thickened' version of G identify the circular DNA strands that assemble (hybridize) into the graph G. For details, see [9], where the objective was to show that every graph has an associated orientable thickened graph, with a boundary component visiting every edge at least once, thus corresponding to the desired route for the reporter strand.

While motivated by a particular application, this problem of finding suitable walks is of independent intrinsic interest in topological graph theory. Thickened graphs are also known as *ribbon graphs*, and are equivalent to embeddings of graphs in compact surfaces. Each face of an embedding corresponds to a boundary component of a thickened graph, which corresponds to a circular strand of DNA. Thus, showing the existence of a suitable walk for a scaffolding or reporter strand is equivalent to proving that every graph admits an orientable embedding where every edge lies on a single face. The facial walk of this face gives the corresponding desired route for the DNA strand. Prompted by this application, we define a *reporter strand walk* in a graph G to be a walk that uses every edge of G at least once and occurs as a facial boundary walk in some orientable embedding of G. See Figure 1 for two examples of embeddings of K_4 on the torus with facial walks that are reporter strand walks. Notice that the walk shown on the right is shorter (has fewer edges) than the one on the left.

Having a face that includes every edge is intermediate between two well-known properties of graph embeddings. A *one-face embedding* is an embedding in which there is only one face, so every edge occurs exactly twice on the boundary of this single face; a one-face embedding is necessarily a maximum genus embedding. An *outer embedding* in a given surface is an embedding in which all vertices appear on a single face (outerplanar graphs are particularly well-studied). It therefore seems appropriate to call an embedding in which there is a face (the 'outer' face) that includes every edge an *edge-outer embedding*. For connected graphs, a one-face embedding is edge-outer, and an edge-outer embedding is outer.

Here we give a short proof of the result from [9] that reporter strand walks always exist, and hence every graph has an edge-outer embedding. Our result provides a polynomial-time algorithm that finds both the embedding and the reporter strand walk. Furthermore, we show that the problem of finding a shortest length reporter strand walk, or equivalently, an embedding with smallest degree outer face, is NP-hard, even for 3-connected cubic planar graphs. We begin with a weaker NP-hardness result in Section 3 and strengthen the result in Section 4.

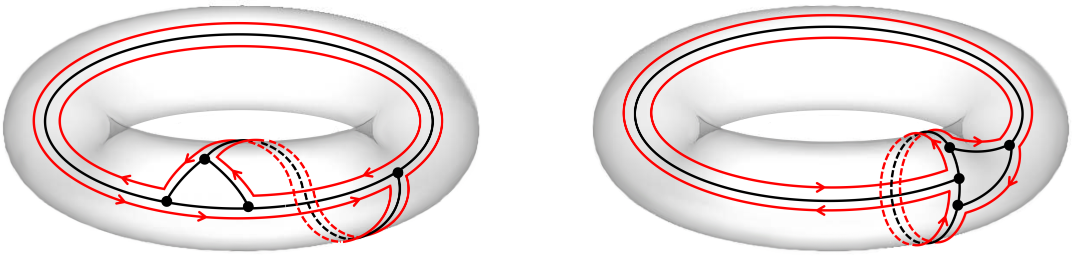

Figure 1: Facial walks corresponding to reporter strands.

2. A short proof that reporter strand walks exist

In this section we provide a short proof of the existence of reporter strand walks in all graphs. In this paper, graphs may have loops and multiple edges. A graph with neither loops nor multiple edges is *simple*. Sometimes we think of an edge as consisting of two distinct *edge-ends* or just *ends*; $D_G(v)$ denotes the set of ends incident with v in G.

All embeddings in this paper will be *cellular*, in which every face is homeomorphic to an open disk. We assume the reader is familiar with combinatorial descriptions of orientable cellular embeddings of graphs, using *rotation schemes* or *rotation systems*, as described in [4, 7, 11]. A rotation scheme assigns to each vertex a *rotation*, which is a cyclic ordering of the edge-ends incident with that vertex, corresponding to their order in the globally consistent clockwise direction in the surface. Every orientable embedding is determined up to homeomorphism by its rotation scheme.

Suppose we have an embedding of a graph G described by a rotation scheme, with the rotation at a given vertex v being $(d_0, d_1, d_2, \ldots, d_{k-1})$, where $k = \deg(v)$ and $D_G(v) = \{d_0, d_1, d_2, \ldots, d_{k-1}\}$. Further, suppose v is incident with a face f and $d_i \in D_G(v)$ is not on f for some i. Then we say we are *flipping* d_i *into* f if we change this rotation at v, and thus the overall rotation scheme, by moving d_i to between d_j and d_{j+1}, where f is between d_j and d_{j+1} for some j (interpreting subscripts modulo k). If the face f visits v several times, there may be more than one such j, but this does not matter for our purposes.

In our terminology, the main result of [9] is that every connected graph has a reporter strand walk. The following algorithm provides a short proof.

Algorithm 2.1. Given a connected graph G (with loops and multiple edges allowed):

Take an arbitrary orientable embedding of G.
Choose an arbitrary face f.
While some edge is not in f {
    Choose an edge e not in f, but incident with a vertex v of f.
    Modify the embedding by flipping an end of e incident with v into f.
    There is a new face using e twice; let f be this face.
}

Assertion. This algorithm runs in polynomial time. It terminates with an orientable embedding of G in which f is a face using every edge of G. Thus, the boundary walk of f is a reporter strand walk in G.

Proof that the algorithm works. The initial embedding exists because we may just give each vertex an arbitrary rotation. The edge e always exists because G is connected. When we flip the end of e into f, we create a new face f' that includes all edges of the old face f, and uses e twice. (If e belonged to two distinct old faces then f' also uses all other edges from those faces. Otherwise, e belonged to a single old face g, and the two occurrences of e split the boundary of g into two pieces; f' includes the edges of one of those pieces. Also, the length of the face f increases both when e has only one end incident with a vertex of f and when it has two, as the example below illustrates.) Since f' becomes the new f, the edge set of f strictly increases at each iteration, until it contains all edges.

The operations in the algorithm are easily implemented in polynomial time using the rotation scheme representation of an embedding. □

An example is shown in Figure 2, where we represent orientable embeddings of a graph as plane drawings with possible edge crossings. With this representation, the rotation scheme just corresponds to the clockwise ordering at each vertex in the drawing, and we can trace faces in the usual way for a plane graph, except that we ignore edge crossings. We show a complete run of the algorithm, which requires two iterations. The tracing of the face f is shown initially and after each iteration (in red, if color appears) and the edges e_1, e_2 used in the two iterations are labeled.

Remark. Often, an important consideration in determining reporter or scaffolding strand walks is assuring that the result is not knotted. The authors of [5] observed that knotted walks can result from A-trails (non-crossing Eulerian circuits) in toroidal meshes, while [12] characterizes knotted and unknotted A-trails in toroidal meshes and [10] gives an an approximation algorithm for unknotted walks in surface triangulations. We note here that Algorithm 2.1 starts with an abstract graph and outputs a walk that is a facial walk of the graph embedded in an orientable surface. This walk thus bounds a disk and hence is unknotted when viewed as a curve in space. This means that every graph has some embedding in 3-dimensional space, in fact an embedding in some orientable surface, with an unknotted reporter strand walk.

3. NP-completeness of short reporter strands

3.1. Two decision problems

Given that reporter (or scaffolding) strand walks exist, for experimental efficiency it is natural to seek a shortest such walk, i.e., one with as few edges as possible, and hence a minimum number of duplicated edges. However, in this section we show that the following decision version of finding a shortest reporter strand walk is NP-complete.

Short Reporter Strand Walk (SRSW). Given a graph G and a nonnegative integer k, does G have a reporter strand walk of length at most k?

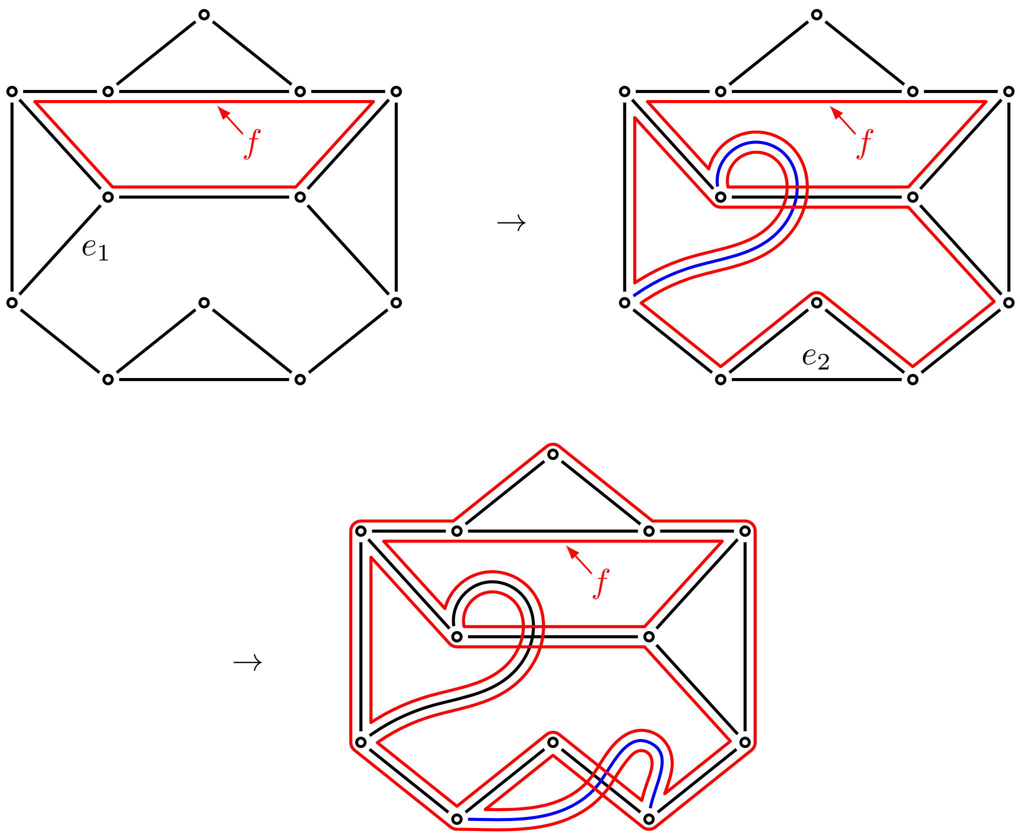

Figure 2: Example of Algorithm 2.1.

In other words, SRSW asks whether G has an orientable embedding with a facial walk that uses all edges and has length at most k. A 'yes' instance can be certified by giving a suitable embedding of G, so SRSW is in NP. The construction used here to prove that SRSW is NP-complete is relatively straightforward; it forms the first step towards a stronger NP-completeness result given in Section 4.

All walks from this point onwards (including paths and cycles) are directed walks. Let W^{-1} denote the reverse of a walk W. For two walks W_1 and W_2, $W_1 \cdot W_2$ denotes their concatenation, which is only defined if the last vertex of W_1 is the first vertex of W_2. A walk is *edge-spanning* if it uses every edge at least once, and *edge-2-bounded* if it uses every edge at most twice. In any walk an edge used exactly once is a *solo* edge, and an edge used exactly twice is a *double* edge.

For much of this section and the next we will be working with simple graphs. So, we can uniquely identify an edge between u and v using the notation uv. We can also describe walks just using sequences of vertices: then uv also means a one-edge walk from u to v, in that direction. Using special notation to distinguish between the (undirected) edge uv and the (directed) walk uv would be unwieldy; we rely instead on context or explicit textual explanation.

We are interested in walks that can occur as face boundaries in an orientable graph embedding. For such a walk W, we should be able to glue a facial disk to the graph, identifying its boundary with W, without preventing the neighborhood of any vertex from being an open disk in an embedding, and without introducing nonorientability. This yields two properties. Given a walk W in G and $v \in V(G)$, let $\mathrm{Rot}_G(W, v)$ be the graph with vertex set $D_G(v)$, where we add one edge between edge-ends d, d' for each time W enters v on d and immediately leaves on d', or vice versa. In an embedding, for each v the union of $\mathrm{Rot}_G(W, v)$ over all facial walks W is a cycle on $D_G(v)$ describing the (undirected) rotation at v, so each $\mathrm{Rot}_G(W, v)$ is a subgraph of such a cycle. Therefore, we say a closed walk W is *rotation-compatible in* G if for every $v \in V(G)$, $\mathrm{Rot}_G(W, v)$ is either a cycle with vertex set $D_G(v)$, or a union of vertex-disjoint paths. Also, a walk is *orientable* if it uses each edge at most once in each direction. A walk that is rotation-compatible or orientable is edge-2-bounded.

A result of Širáň and Škoviera [17, Prop. 1] implies that a closed walk in G occurs as a face boundary in some orientable embedding of G if and only if it is orientable and rotation-compatible in G. Loosely, if W is orientable and rotation-compatible then each $\mathrm{Rot}_G(W, v)$ describes a partial rotation at v that can be arbitrarily completed to a full rotation, giving a rotation scheme. Thus, a reporter strand walk in G is precisely a closed walk that is edge-spanning, orientable, and rotation-compatible in G.

A natural lower bound on the length of a reporter strand walk is the length of a *Chinese postman walk*, an edge-spanning closed walk of minimum length. A Chinese postman walk is edge-2-bounded, but need not be rotation-compatible or orientable. Such walks were first considered by Guan [8]. Thus, we have a more specific decision problem.

Chinese Postman Reporter Strand Walk (CPRSW). Given a graph G, does it have a reporter strand walk that is also a Chinese postman walk? (Such a walk is a *Chinese postman reporter strand walk* or *CPRS walk*.)

Since a Chinese postman walk, and its length, can be found in polynomial time [2], CPRSW is in NP. Moreover, every instance of CPRSW can be transformed in polynomial time to an instance of SRSW, involving the same graph. Therefore, if we show that CPRSW is NP-complete for a class of graphs, SRSW is also NP-complete for that class.

We will prove that CPRSW is NP-complete even when restricted to 2-connected cubic planar graphs, and, in the next section, to 3-connected cubic planar graphs. We do this by reducing the hamilton cycle problem for 3-connected cubic planar graphs, which is known to be NP-complete [6], to CPRSW.

In this section and the following section, we work with cubic graphs and their subgraphs.

3.2. Special properties of cubic graphs

Both Chinese postman and reporter strand walks have special structures in 2-connected cubic graphs.

First we consider Chinese postman walks. Suppose G is a 2-connected cubic graph. Any edge-spanning walk in G uses all three edges at each vertex, so it must use each vertex at least twice, and hence it must have length at least $2|V(G)|$. Thus, its length is $2|V(G)|$

if and only if it uses each vertex exactly twice, if and only if it contains exactly two solo edges and one double edge incident with each vertex. Now, by a well-known result [14] of Petersen, G has a perfect matching M. If we replace each edge of M in G by two parallel edges to obtain G', then G' is eulerian, and an euler tour in G' gives an edge-spanning walk in G of length $2|V(G)|$. Therefore, a Chinese postman walk has length $2|V(G)|$, and hence uses two solo edges and one double edge at each vertex. We summarize this as follows.

Lemma 3.1. *A closed walk in a* 2*-connected cubic graph* G *is a Chinese postman walk if and only if it is edge-spanning, edge-*2*-bounded, and its double edges form a perfect matching of* G*.*

Now we consider reporter strand walks. In cubic graphs, rotation-compatibility can be replaced by a simpler property. A *retraction* in a walk consists of an edge followed immediately by the same edge in the opposite direction. A walk with no retractions is *retraction-free*. If a graph has no vertices of degree 1, every rotation-compatible closed walk is retraction-free. If a graph has no vertices of degree 4 or more, every retraction-free edge-2-bounded closed walk is rotation-compatible. Therefore, in a cubic graph a closed walk is rotation-compatible if and only if it is edge-2-bounded and retraction-free, giving the following.

Lemma 3.2. *A closed walk in a cubic graph* G *is a reporter strand walk if and only if it is edge-spanning, orientable, and retraction-free.*

Corollary 3.3. *A closed walk in a* 2*-connected cubic graph is a Chinese postman reporter strand (CPRS) walk if and only if it is edge-spanning, orientable, retraction-free, and the double edges form a perfect matching of* G*.*

The two walks in K_4 shown in Figure 1 illustrate these characterizations: both satisfy Lemma 3.2 and are reporter strand walks, and the one on the right satisfies Corollary 3.3 and is a CPRS walk.

Suppose W is a CPRS walk in 2-connected cubic G, and $v \in V(G)$. Since W is orientable, the double edge at v, call it $\delta_W(v)$, is used in both directions by W, so one of the solo edges at v, call it $\sigma_W^-(v)$, must be used by W to enter v, and the other, $\sigma_W^+(v)$, must be used by W to leave v. Since W is retraction-free, it must use the edge sequences $\sigma_W^-(v)\delta_W(v)$ and $\delta_W(v)\sigma_W^+(v)$ to pass through v.

Therefore, we can reconstruct W from the choice of double edges (which form a matching) and of orientations for the remaining solo edges (one entering, one leaving each vertex). To consider possible CPRS walks we make such choices and try to trace W by following the edge sequences $\sigma_W^-(v)\delta_W(v)$ and $\delta_W(v)\sigma_W^+(v)$ at v. In general this *tracing procedure* may fail by finding a closed walk that is not edge-spanning. If this does not happen we obtain a CPRS walk.

A connected graph H with two vertices v_1, v_2 of degree 2 and all other vertices of degree 3 is called an *edge gadget*. If G is a graph disjoint from H and $u_1u_2 \in E(G)$, then we say $J = (G - u_1u_2) \cup H \cup \{u_1v_1, u_2v_2\}$ is obtained by *bisecting* u_1u_2 *in* G *with* H. The *cubic completion* of H is $H^+ = H \cup v_1v_2$. We omit the proof of the following standard result.

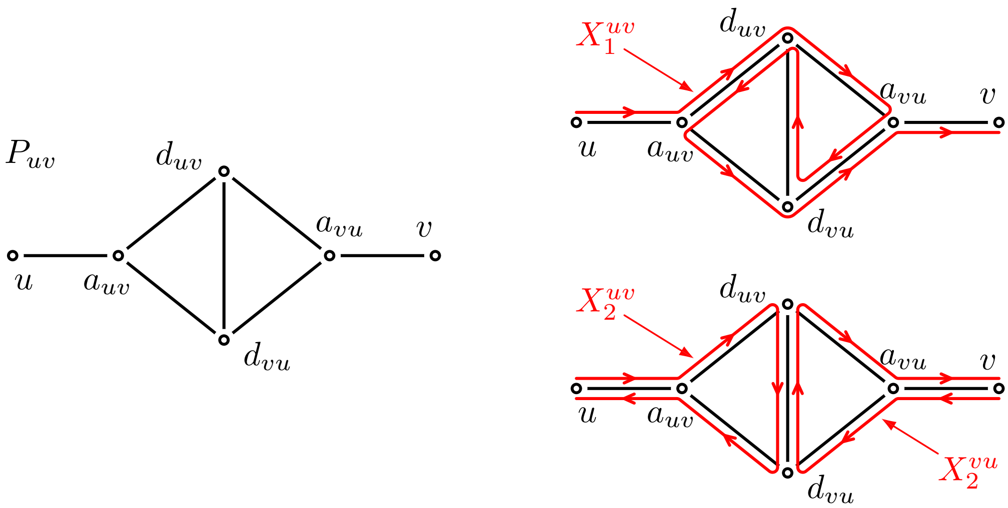

Figure 3: Construction of P, and how a CPRS walk passes through P_{uv}.

Lemma 3.4. *Let G be a cubic graph, and H an edge gadget. Construct J by bisecting an edge of G with H. Then J is cubic. If G and the cubic completion H^+ are both 2-connected, planar, and simple, then J is 2-connected, planar, and simple.*

3.3. The NP-completeness result

Construction 3.5. Given a 3-connected cubic planar simple graph N, construct a new graph P by replacing each edge uv by a subgraph P_{uv} ($= P_{vu}$) consisting of a 4-cycle $(a_{uv}d_{uv}a_{vu}d_{vu})$ on four new vertices and three additional edges ua_{uv}, va_{vu} and $d_{uv}d_{vu}$. Note that order matters for subscripts in new vertex names. We use π_{uv} ($= \pi_{vu}$) to refer to the automorphism of P_{uv} that swaps d_{uv} and d_{vu} and fixes the other vertices. See Figure 3.

Claim. The graph P is a 2-connected cubic planar simple graph.

Proof of claim. The graph $P'_{uv} = P_{uv} - \{u, v\}$ is an edge gadget, and $(P'_{uv})^+ \cong K_4$ is 2-connected, planar, and simple. Replacing uv by P_{uv} is equivalent to bisecting uv with P'_{uv}, so the claim follows by repeated application of Lemma 3.4. □

Lemma 3.6. *Suppose we construct P as in Construction 3.5. Let $X_1^{uv} = ua_{uv}d_{uv}a_{vu}d_{vu}d_{uv}$-$a_{uv}d_{vu}a_{vu}v$ and $X_2^{uv} = ua_{uv}d_{uv}d_{vu}a_{uv}u$. Then a CPRS walk W in P must pass through each subgraph P_{uv} in one of two ways,*

(a) as a single walk X_1^{uv}, $\pi_{uv}(X_1^{uv})$, $(X_1^{uv})^{-1}$ or $(\pi_{uv}(X_1^{uv}))^{-1}$; or

(b) as two walks X_2^{uv} and X_2^{vu}, or $\pi_{uv}(X_2^{uv}) = (X_2^{uv})^{-1}$ and $\pi_{uv}(X_2^{vu}) = (X_2^{vu})^{-1}$.

Proof. If $d_{uv}d_{vu}$ is a double edge of W, then ua_{uv} and va_{vu} are also double edges. The solo edges in P_{uv} form a single cycle, which must be oriented consistently, as either $(a_{uv}d_{uv}a_{vu}d_{vu})$ or its reverse. Applying the tracing procedure described above, (b) holds.

If $d_{uv}d_{vu}$ is not a double edge, the set of double edges in P_{uv} is either $\{a_{uv}d_{uv}, d_{vu}a_{vu}\}$ or $\{a_{uv}d_{vu}, d_{uv}a_{vu}\}$. By symmetry (from π_{uv}) we may assume the former. The solo edges form a single path $ua_{uv}d_{vu}d_{uv}a_{uv}v$ which must be oriented in this direction or its reverse. Applying the tracing procedure, (a) holds. □

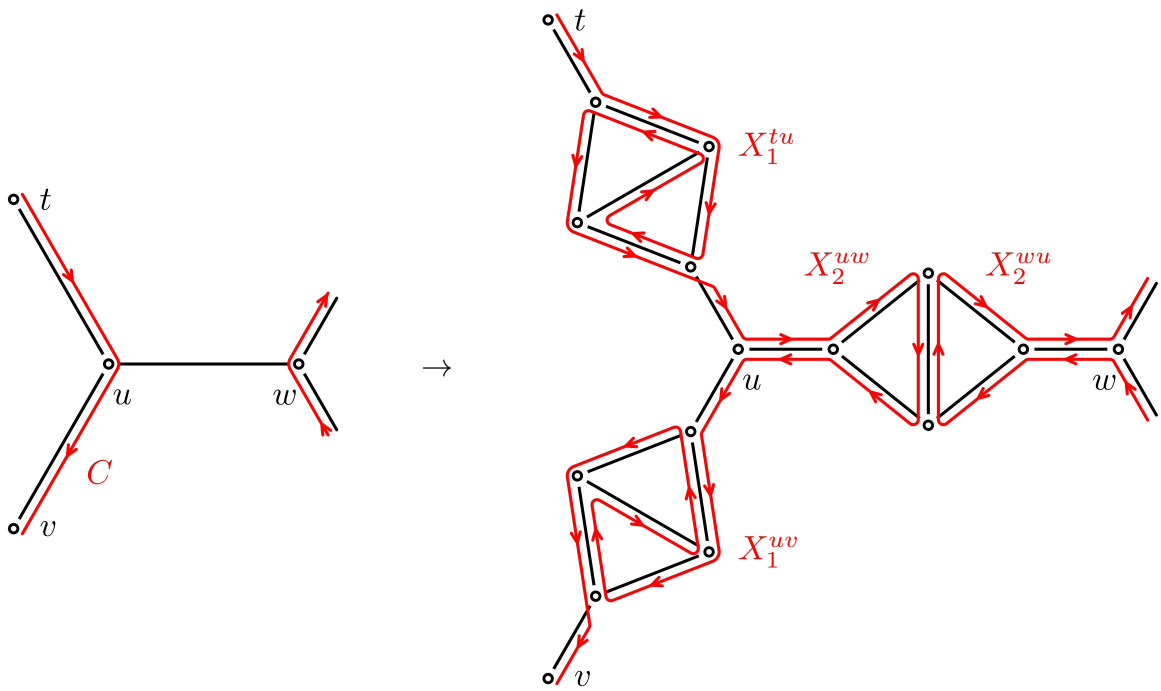

Figure 4: Constructing a CPRS walk in P from a hamilton cycle in N.

Thus, Lemma 3.6 says that up to symmetry or reversal a CPRS walk must pass through P_{uv} in one of the ways shown on the right in Figure 3.

Proposition 3.7. *For N and P as in Construction 3.5, the following are equivalent.*
(a) N has a hamilton cycle.
(b) P has a Chinese postman reporter strand walk.

Proof. Suppose N has a hamilton cycle, C. First, replace each (directed) edge uv of C by the walk X_1^{uv}. This gives a walk in P that uses every vertex of N once. Now for each edge $uw \in E(N) - E(C)$ splice X_2^{uw} into this walk at u, and splice X_2^{wu} into this walk at w. The result is a CPRS walk in P. See Figure 4.

Conversely, suppose P has a CPRS walk W. By Lemma 3.6, at each $u \in V(N)$, $\sigma_W^-(u)$ belongs to some subwalk $W_1^{tu} = X_1^{tu}$ or $\pi_{tu}(X_1^{tu})$ of W, $\sigma_W^+(u)$ belongs to some $W_1^{uv} = X_1^{uv}$ or $\pi_{uv}(X_1^{uv})$, and both occurrences of $\delta_W(u)$ belong to some $W_2^{uw} = X_2^{uw}$ or $(X_2^{uw})^{-1}$, where t, v, w are the neighbors of u in N. Thus, deleting all subwalks W_2^{uw} and replacing each subwalk W_1^{uv} by the edge uv of N gives a hamilton cycle in N. □

Construction 3.5 therefore gives a polynomial time transformation from the hamilton cycle problem for 3-connected cubic planar simple graphs, which is NP-complete [6], to CPRSW for 2-connected cubic planar simple graphs. This yields the following theorem.

Theorem 3.8. *The problems* Short Reporter Strand Walk *and* Chinese Postman Reporter Strand Walk *are NP-complete for 2-connected cubic planar simple graphs.*

4. A stronger NP-completeness result

4.1. Achieving 3-connectedness

While Section 3 provides a simple proof of NP-completeness for the problems SRSW and CPRSW, the class of graphs that it uses does not have a stable 3-dimensional structure, so they are not likely to occur in situations where we design a DNA molecule to have a specified geometric embedding in space. In particular, the graphs P produced by Construction 3.5 have connectivity 2, while the graph formed by the edges of any polyhedron in 3-dimensional space is 3-connected. Theorem 3.8 leaves open the possibility that SRSW and CPRSW can be solved easily for 3-connected graphs, or even that all 3-connected graphs have a CPRS walk. Here we show that for 3-connected graphs (in fact, 3-connected cubic planar graphs) SRSW and CPRSW are NP-complete, and hence unlikely to have polynomial-time solutions.

First we modify the graph P from Construction 3.5 to obtain a new graph Q with improved connectivity, in Construction 4.1. However, CPRS walks in Q do not necessarily correspond to CPRS walks in P, so later we further modify Q into a graph R where we can control the CPRS walks so that they do correspond to CPRS walks in P, and hence to hamilton cycles in N.

Given a graph G with a plane embedding, let $\mathrm{cwn}_G(u, v)$ denote the neighbor of u that is immediately clockwise from v in the rotation around u.

Construction 4.1. Suppose we have N and P as in Construction 3.5. To construct Q, take a plane embedding of N, and a corresponding plane embedding of P in which each 4-cycle $(a_{uv}d_{uv}a_{vu}d_{vu})$ is clockwise. Replace each edge $a_{uv}d_{uv}$ of P by a path $a_{uv}b_{uv}c_{uv}d_{uv}$ involving two new vertices b_{uv}, c_{uv}. Then incident to each vertex c_{uv} add a *bracing edge* $c_{uv}b_{vw}$ where $w = \mathrm{cwn}_N(v, u)$. See Figure 5.

Given a graph G, define a relation E_3^G, or just E_3, on $V(G)$ by uE_3v when there are three edge-disjoint uv-paths in G; equivalently (by the edge version of Menger's Theorem) when no set of fewer than 3 edges separates u and v.

Lemma 4.2. *E_3 is an equivalence relation.*

Proof. E_3 is reflexive (take three copies of the trivial walk at a vertex) and clearly symmetric; we must show it is transitive. Suppose that uE_3v and vE_3w. If we do not have uE_3w then some set of fewer than 3 edges separates u and w. But then this set either separates u and v, contradicting uE_3v, or v and w, contradicting vE_3w. Hence, uE_3w. □

A graph G is 3-edge-connected precisely when all vertices of G are E_3-equivalent.

Lemma 4.3. *The graph Q is 3-connected, planar, and simple.*

Proof. Clearly Q is planar and simple (see Figure 5). For cubic graphs such as N and Q, 3-connectedness is equivalent to 3-edge-connectedness, which is equivalent to showing that all vertices are E_3-equivalent.

Vertices of Q are either *original* vertices, namely vertices of N, or *new* vertices, added by Constructions 3.5 and 4.1. If u and v are original vertices then there are three edge-disjoint

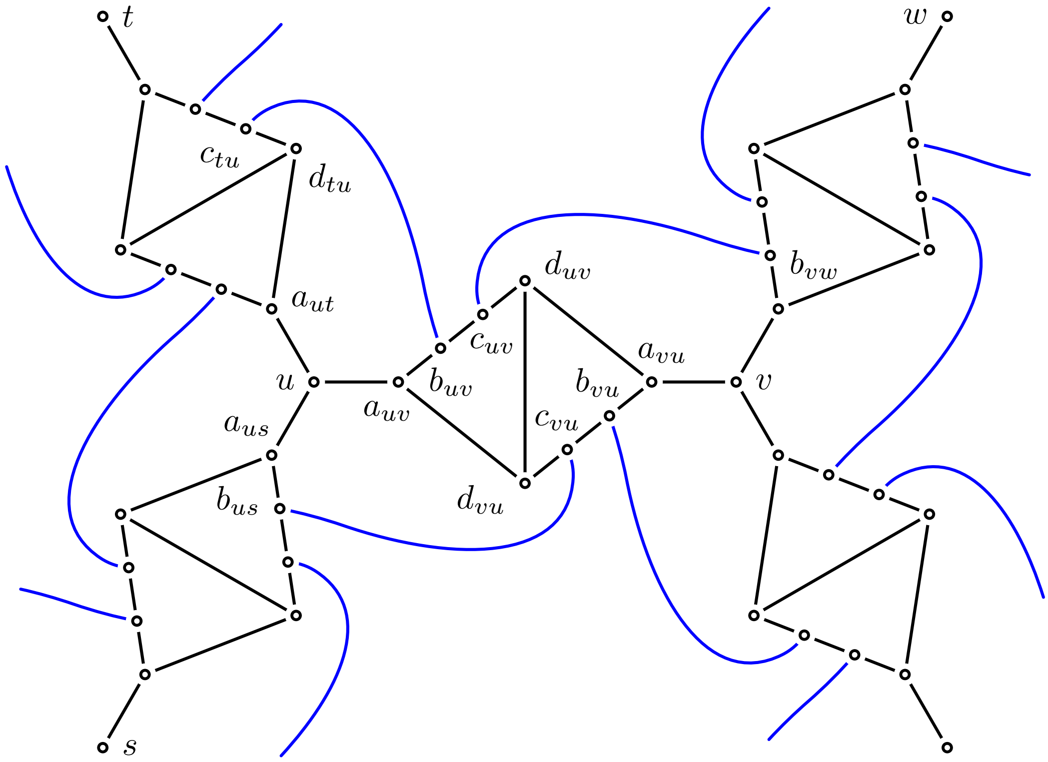

Figure 5: Construction of Q.

uv-paths in N, which easily provide three edge-disjoint uv-paths in Q. Hence all original vertices are E_3^Q-equivalent. So it suffices to show that each new vertex is E_3^Q-equivalent to some original vertex.

Suppose that in the plane embedding of N, the neighbors of u are s, t, v in clockwise order. The following paths from new vertices of Q to the original vertex u (see Figure 5) show that a_{uv}, b_{uv} and d_{uv} are E_3^Q-equivalent to u:

$a_{uv}u$-paths: $a_{uv}u$, $a_{uv}b_{uv}c_{tu}d_{tu}a_{ut}u$, $a_{uv}d_{vu}c_{vu}b_{us}a_{us}u$.

$b_{uv}u$-paths: $b_{uv}a_{uv}u$, $b_{uv}c_{uv}d_{uv}d_{vu}c_{vu}b_{us}a_{us}u$, $b_{uv}c_{tu}d_{tu}a_{ut}u$.

$d_{uv}u$-paths: $d_{uv}d_{vu}a_{uv}u$, $d_{uv}c_{uv}b_{uv}c_{tu}d_{tu}a_{ut}u$, $d_{uv}a_{vu}b_{vu}c_{vu}b_{us}a_{us}u$.

Rather than c_{uv} it is more convenient to show that c_{vu} is E_3^Q-equivalent to u:

$c_{vu}u$-paths: $c_{vu}d_{vu}a_{uv}u$, $c_{vu}b_{us}a_{us}u$, $c_{vu}b_{vu}a_{vu}d_{uv}c_{uv}b_{uv}c_{tu}d_{tu}a_{ut}u$.

Since every new vertex is a_{uv}, b_{uv}, d_{uv} or c_{vu} for some choice of u and v, every new vertex is E_3^Q-equivalent to an original vertex, as required. □

4.2. Controlling walks

A connected graph H with three vertices v_1, v_2, v_3 of degree 2 and all other vertices of degree 3 is called a *vertex gadget*. If G is a graph disjoint from H and $u \in V(G)$ has degree 3 with neighbors u_1, u_2, u_3, then we say the graph $J = (G - u) \cup H \cup \{u_1v_1, u_2v_2, u_3v_3\}$ is obtained by *replacing u in G by H*. The *cubic completion* of H is $H^+ = H \cup \{vv_1, vv_2, vv_3\}$ where v is a new vertex. We omit the proof of the following standard result.

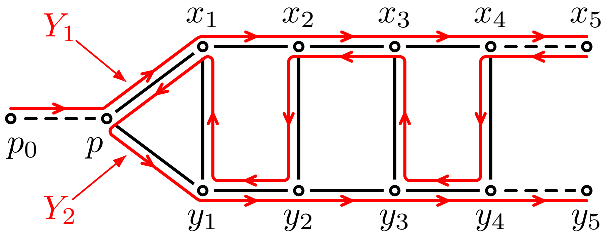

Figure 6: Vertex gadget A, and how a CPRS walk passes through it.

Lemma 4.4. *Let G be a cubic graph, and H a vertex gadget. Construct J by replacing a vertex of G with H. Then J is cubic. If G and the cubic completion H^+ are both 3-connected, planar, and simple, then J is 3-connected, planar, and simple.*

Now we construct subgraphs in which the route taken by a CPRS walk is constrained in various ways.

Let A be the vertex gadget shown (with additional incident edges p_0p, x_4x_5, y_4y_5) in Figure 6. Let α be the automorphism of A that swaps the two paths $x_1x_2x_3x_4$ and $y_1y_2y_3y_4$ while fixing p. Note that the cubic completion A^+ is 3-connected (to see this, observe that for every $v \in V(A^+)$, $A^+ - v$ has a hamilton cycle and is therefore 2-connected). Also, A^+ is planar and simple.

Lemma 4.5. *Suppose the vertex gadget A described above is an induced subgraph of a 2-connected cubic graph G. Let $Y_1 = px_1x_2x_3x_4$ and $Y_2 = x_4y_4y_3x_3x_2y_2y_1x_1py_1y_2y_3y_4$. If W is a CPRS walk in G then W passes through A and its incident edges as two walks, either $p_0p{\cdot}Y_1{\cdot}x_4x_5$ and $x_5x_4{\cdot}Y_2{\cdot}y_4y_5$, or $p_0p{\cdot}\alpha(Y_1){\cdot}y_4y_5$ and $y_5y_4{\cdot}\alpha(Y_2){\cdot}x_4x_5$, or reversing both walks in one of these pairs.*

Proof. Suppose first that p_0p is a double edge in W. If x_1y_1 is not a double edge then x_1x_2 and y_1y_2 are double edges. We have a triangle (px_1y_1) of solo edges; we may assume its edges are oriented in that direction by W. We have another path of solo edges $x_3x_2y_2y_3$. If this is oriented as $y_3y_2x_2x_3$ then then the tracing procedure fails by finding a 4-cycle $(x_1y_1y_2x_2)$. So it is oriented as $x_3x_2y_2y_3$. If x_3y_3 is a double edge then the tracing algorithm fails by finding a 6-cycle $(x_1y_1y_2y_3x_3x_2)$. Thus, x_3y_3 is a solo edge, it must be oriented as y_3x_3, x_3x_4 and y_3y_4 are double edges, and x_4y_4 is a solo edge. If x_4y_4 is oriented as x_4y_4, then the tracing procedure fails by finding a 4-cycle $(y_3x_3x_4y_4)$, and if it is oriented as y_4x_4, then the tracing algorithm fails by finding an 8-cycle $(x_1y_1y_2y_3y_4x_4x_3x_2)$.

If x_1y_1 is a double edge we have a path of solo edges $y_2y_1px_1x_2$ which without loss of generality is oriented in that direction. If x_2y_2 is a double edge, then the tracing procedure fails by finding the 4-cycle $(y_2y_1x_1x_2)$. So x_2y_2 is a solo edge, it must be oriented as x_2y_2, x_2x_3 and y_2y_3 are double edges, and x_3y_3 is a single edge. If x_3y_3 is oriented as x_3y_3 then our tracing algorithm fails by finding a 6-cycle $(x_3y_3y_2y_1x_1y_2)$, and if it is oriented as y_3x_3 then the tracing algorithm fails by finding a 4-cycle $(y_3x_3x_2y_2)$.

Therefore, p_0p is a solo edge; without loss of generality, $p_0p = \sigma_W^-(p)$. By symmetry (from α) we may assume that $\delta_W(p) = px_1$. Then y_1y_2, x_2x_3, y_3y_4 and x_4x_5 must all be

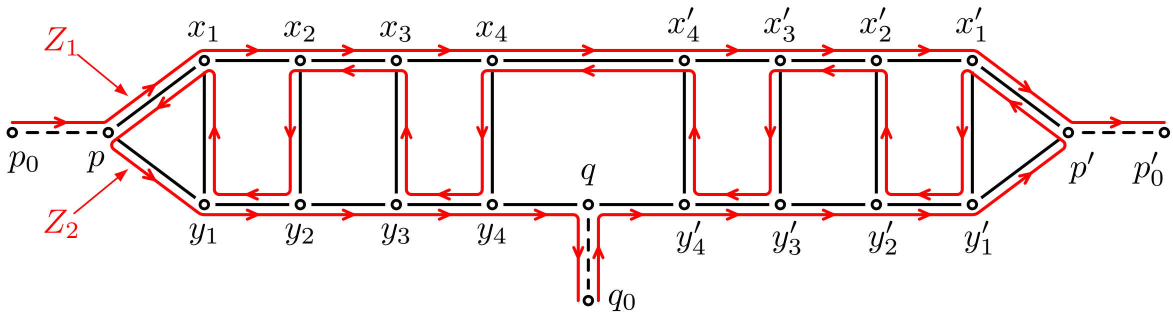

Figure 7: Vertex gadget B, and how a CPRS walk passes through it.

double edges. The solo edges form a single path which is oriented $p_0py_1x_1x_2y_2y_3x_3x_4y_4y_5$. Now the tracing procedure gives $p_0p{\cdot}Y_1{\cdot}x_4x_5$ and $x_5x_4{\cdot}Y_2{\cdot}y_4y_5$. □

Thus, Lemma 4.5 says that up to symmetry or reversal a CPRS walk must pass through A as shown in Figure 6. Loosely, A acts like a vertex, in that a CPRS walk passes through it as two walks of the form (entering solo edge)...(double edge) and (double edge)...(leaving solo edge), but with a restriction: the edge p_0p must be a solo edge.

Now we build a larger vertex gadget. Let A' be a copy of A, with a plane embedding that is the mirror image of the embedding of A in Figure 6. Let p' in A' correspond to p in A, and so on. Let $B = A \cup A' \cup \{x_4x'_4, y_4q, y'_4q\}$ where q is a new vertex. Then B is a vertex gadget. Note that B^+ can be considered as obtained from K_4 by replacing two vertices by copies of A, so by Lemma 4.4 applied twice, B^+ is 3-connected, planar, and simple.

Lemma 4.6. *Suppose the vertex gadget B described above is an induced subgraph of a 2-connected cubic graph G, with incident edges p_0p, qq_0 and $p'p'_0$. Let $Z_1 = Y_1{\cdot}x_4x'_4{\cdot}(Y'_1)^{-1}$ and $Z_2 = qy'_4{\cdot}(Y'_2)^{-1}{\cdot}x'_4x_4{\cdot}Y_2{\cdot}y_4q$. If W is a CPRS walk in G then W passes through B and its incident edges as two walks, either $p_0p{\cdot}Z_1{\cdot}p'p'_0$ and $q_0q{\cdot}Z_2{\cdot}qq_0$, or reversing both of these walks.*

Proof. Applying Lemma 4.5 to both A and A', p_0p and $p'p'_0$ are solo edges. The perfect matching of double edges of W must have an odd number of edges leaving the odd set $V(B)$, so q_0q must be a double edge. Therefore, qy_4 and qy'_4 are solo edges. Now Lemma 4.5, applied to both A and A', gives the result. □

Thus, Lemma 4.6 says that up to reversal (or, equivalently, up to the automorphism of B swapping p and p') a CPRS walk must pass through B as shown in Figure 7.

4.3. NP-completeness for 3-connected cubic planar graphs

Construction 4.7. Suppose we have N, P and Q as in Constructions 3.5 and 4.1. For each vertex b_{uv} take a copy B_{uv} of B, where $p_{uv}, q_{uv}, p'_{uv}, Z_1^{uv}, Z_2^{uv}$ correspond to p, q, p', Z_1, Z_2 in B, respectively. Construct R by replacing each vertex of the form b_{uv} in Q by B_{uv}, so that if b_{uv} is adjacent to a_{uv}, c_{uv}, c_{tu} then the edges incident with B_{uv} are $a_{uv}p_{uv}$, $q_{uv}c_{tu}$ and $p'_{uv}c_{uv}$.

Claim. The graph R is a 3-connected cubic planar simple graph.

Proof of claim. As noted above, B^+ is a 3-connected, planar and simple, and so is Q by Lemma 4.3. The claim follows by repeated application of Lemma 4.4. □

Proposition 4.8. *For N, P, Q and R as in Constructions 3.5, 4.1 and 4.7, the following are equivalent.*

(a) N has a hamilton cycle.
(b) P has a Chinese postman reporter strand walk.
(c) P has a Chinese postman reporter strand walk using every edge of the form $a_{uv}d_{uv}$ as a solo edge.
(d) R has a Chinese postman reporter strand walk.

Proof. By Proposition 3.7, (a) ⇔ (b). Clearly (c) ⇒ (b). Suppose (b) holds and we have a CPRS walk W in P. Suppose some $a_{uv}d_{uv}$ is not a solo edge of W. By Lemma 3.6, W must use X_1^{uv} or its reverse; replacing this by $\pi_{uv}(X_1^{uv})$ or its reverse we still have a CPRS walk, and now $a_{uv}d_{uv}$ (and also $a_{vu}d_{vu}$) is a solo edge. Applying this to all $a_{uv}d_{uv}$ that are not solo edges, we obtain a CPRS walk W' satisfying (c). Thus, (b) ⇒ (c).

So now we show that (c) ⇔ (d). Suppose that (c) holds, with a walk W using each $a_{uv}d_{uv} \in E(P)$ as a solo edge. The bracing edge of Q incident with c_{uv} has the form $c_{uv}b_{vw}$, where w follows u in clockwise order around v in N. Replace each directed edge $a_{uv}d_{uv}$, or its reverse, in W by a walk in R according to the following rules:

W uses $a_{uv}d_{uv}$, $a_{vw}d_{vw}$: $a_{uv}d_{uv} \to T_{uv}^{00} = a_{uv}p_{uv}\cdot Z_1^{uv}\cdot p'_{uv}c_{uv}q_{vw}\cdot Z_2^{vw}\cdot q_{vw}c_{uv}d_{uv}$.
W uses $a_{uv}d_{uv}$, $d_{vw}a_{vw}$: $a_{uv}d_{uv} \to T_{uv}^{01} = a_{uv}p_{uv}\cdot Z_1^{uv}\cdot p'_{uv}c_{uv}q_{vw}\cdot (Z_2^{vw})^{-1}\cdot q_{vw}c_{uv}d_{uv}$.
W uses $d_{uv}a_{uv}$, $a_{vw}d_{vw}$: $d_{uv}a_{uv} \to T_{uv}^{10} = d_{uv}c_{uv}q_{vw}\cdot Z_2^{vw}\cdot q_{vw}c_{uv}p'_{uv}\cdot (Z_1^{uv})^{-1}\cdot p_{uv}a_{uv}$.
W uses $d_{uv}a_{uv}$, $d_{vw}a_{vw}$: $d_{uv}a_{uv} \to T_{uv}^{11} = d_{uv}c_{uv}q_{vw}\cdot (Z_2^{vw})^{-1}\cdot q_{vw}c_{uv}p'_{uv}\cdot (Z_1^{uv})^{-1}\cdot p_{uv}a_{uv}$.

The rules guarantee that in each B_{uv} we use both Z_1^{uv} and Z_2^{uv}, or both $(Z_1^{uv})^{-1}$ and $(Z_2^{uv})^{-1}$. Therefore, the result is a CPRS walk W' in R. Thus, (d) holds.

Conversely, suppose (d) holds, so R has a CPRS walk W. Consider each $a_{uv}d_{uv} \in E(P)$ and the corresponding bracing edge $c_{uv}b_{vw} \in E(Q)$. Applying Lemma 4.6 to B_{uv} and B_{vw}, we see that W must either travel from a_{uv} to d_{uv} along T_{uv}^{00} or T_{uv}^{01} from above, or travel from d_{uv} to a_{uv} along T_{uv}^{10} or T_{uv}^{11}. In the former case, replace this subwalk of W by the edge $a_{uv}d_{uv}$ of P; in the latter case replace it by $d_{uv}a_{uv}$. Making all such replacements gives a CPRS walk W' in P in which each $a_{uv}d_{uv}$ is a solo edge. Thus, (c) holds. □

Constructions 3.5, 4.1 and 4.7 therefore give a polynomial time transformation from the hamilton cycle problem for 3-connected cubic planar simple graphs to CPRSW for the same family of graphs. Applying these constructions to nonhamiltonian 3-connected cubic planar graphs N proves the existence of 3-connected cubic planar simple graphs R with no CPRS walk (or we can construct small examples of such graphs easily using vertex gadgets A and B). Our final theorem also follows immediately.

Theorem 4.9. *The problems* SHORT REPORTER STRAND WALK *and* CHINESE POSTMAN REPORTER STRAND WALK *are NP-complete for 3-connected cubic planar simple graphs.*

5. Conclusion

This application brings to light a new, natural area of investigation in topological graph theory, edge-outer embeddability, which seems quite rich in attractive questions and new directions:

1. Is there a polynomial-time algorithm that will return a reporter strand walk that is within x percent of minimum length? The algorithm in Section 2 is within 100% of optimal (at most twice the length). To what extent can this be improved? A related result appears in [10], where they give a cubic-time $\frac{5}{3}$-approximation algorithm in the special case that the graph is a triangulation of an orientable surface.

2. Are there classes of graphs where it is polynomial-time to find a minimum length reporter strand walk? Eulerian graphs are one such class. We have shown that the problem is NP-hard for 3-connected graphs, but can it be solved in polynomial time for graphs with higher connectivity?

3. What can be said about the genus range of embeddings that yield reporter strand walks, or reporter strand walks of minimum length? Are these ranges intervals?

References

[1] S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf, W. M. Shih, Self-assembly of DNA into nanoscale three-dimensional shapes, Nature, 459 (2009) 414–418.

[2] J. Edmonds, E. L. Johnson, Matching, Euler tours and the Chinese postman, Math. Programming 5 (1973) 88–124.

[3] J. A. Ellis-Monaghan, A. McDowell, I. Moffatt, G. Pangborn, DNA origami and the complexity of Eulerian circuits with turning costs, Nat. Comp. 14 (2014) 1–13.

[4] J. A. Ellis-Monaghan, I. Moffatt, Graphs on surfaces: Dualities, polynomials, and knots, SpringerBriefs in Mathematics, Springer, New York, 2013.

[5] J. A. Ellis-Monaghan, G. Pangborn, N. C. Seeman, S. Blakeley, C. Disher, M. Falcigno, B. Healy, A. Morse, B. Singh, M. Westland, Design tools for reporter strands and DNA origami scaffold strands. Theoret. Comput. Sci. 671 (2017) 69–78.

[6] M. R. Garey, D. S. Johnson, R. Endre Tarjan, The planar Hamiltonian circuit problem is NP-complete, SIAM J. Comput. 5 (1976) 704–714.

[7] J. L. Gross, T. W. Tucker, Topological graph theory, Dover, Mineola, New York, 2001.

[8] Meigu Guan (Guan Meigu), Graphic programming using odd or even points, Acta Mathematica Sinica 10 (1960) 263–266 (in Chinese); translated as Mei-ko Kwan (Kwan Mei-ko), Chinese Mathematics 1 (1962) 273–277.

[9] N. Jonoska, N. Seeman, G. Wu, On existence of reporter strands in DNA-based graph structures, Theoret. Comput. Sci. 410 (2009) 1448–1460.

[10] A. Mohammed, M. Hajij, Unknotted strand routings of triangulated meshes, in: Proceedings of DNA Computing and Molecular Programming: 23rd International Conference, DNA 23 (Austin, TX, USA, September 24-28, 2017), Lecture Notes in Computer Science 10467 (2017) 46-63.

[11] B. Mohar, C. Thomassen, Graphs on surfaces, Johns Hopkins University Press, Baltimore, 2001.

[12] A. Morse, W. Adkisson, J. Greene, D. Perry, B. Smith, G. Pangborn, J. Ellis-Monaghan, DNA origami and unknotted A-trails in torus graphs, preprint. https://arxiv.org/abs/1703.03799

[13] J. Pelesko, Self Assembly: The Science of Things That Put Themselves Together, Chapman and Hall/CRC, 2007.

[14] J. Petersen, Die Theorie der regulären Graphen, Acta Math. 15 (1891) 193–220.
[15] P. W. K. Rothemund, Folding DNA to create nanoscale shapes and patterns, Nature, 440 (2006) 297–302.
[16] N. C. Seeman, Structural DNA Nanotechnology, Cambridge University Press, Cambridge, 2015.
[17] M. Škoviera, J. Širáň, Oriented relative embeddings of graphs, in: Proceedings of the International Conference on Combinatorial Analysis and its Applications (Pokrzywna, 1985), Zastos. Mat. 19 (1987-8) 589–597.