From unicellular fatgraphs to trees

Thomas J. X. Li, Christian M. Reidys

*Corresponding author.

Email addresses: thomasli@vt.edu (Thomas J. X. Li), duck@santafe.edu (Christian M. Reidys)

Biocomplexity Institute of Virginia Tech
1015 Life Sciences Circle, Blacksburg, VA 24061, USA

Department of Mathematics of Virginia Tech
225 Stanger Street, Blacksburg, VA 24061, USA

Abstract

In this paper we study the minimum number of reversals needed to transform a unicellular fatgraph into a tree. We consider reversals acting on boundary components, having the natural interpretation as gluing, slicing or half-flipping of vertices. Our main result is an expression for the minimum number of reversals needed to transform a unicellular fatgraph to a plane tree. The expression involves the Euler genus of the fatgraph and an additional parameter, which counts the number of certain orientable blocks in the decomposition of the fatgraph. In the process we derive a constructive proof of how to decompose non-orientable, irreducible, unicellular fatgraphs into smaller fatgraphs of the same type or trivial fatgraphs, consisting of a single ribbon. We furthermore provide a detailed analysis how reversals affect the component-structure of the underlying fatgraphs. Our results generalize the Hannenhalli-Pevzner formula for the reversal distance of signed permutations.

Keywords: Fatgraph, Genus, Reversal, Tree, Irreducible

1. Introduction

A fatgraph is an embedding of a connected finite graph into a 2-dimensional, compact surface, such that each connected component of its complement is
homeomorphic to a disc. Fatgraphs are also referred to as maps or ribbon graphs, and are a special type of dessin d’enfant as defined by Grothendieck [9]. In this paper we consider fatgraphs of orientable as well as non-orientable surfaces. A fatgraph is unicellular if it has only one boundary component.

A well-studied example are plane trees, i.e. unicellular fatgraphs embedded into the sphere. Unicellular fatgraphs appear frequently in combinatorics [16, 15, 5, 6] and play a central role in geometry: moduli spaces of curves [11, 18], in knot theory: Bollobás-Riordan-Tutte polynomial [4, 7], in mathematical physics: quantum field theory [2, 14], and in bioinformatics: modeling of macromolecules [19] including proteins [20] as well as RNA structures [17, 1].

Recently, Huang and Reidys [13] give a topological interpretation of the Hannenhalli-Pevzner theory for the reversal distance of signed permutations. The reversal distance problem originates from genome rearrangements in molecular evolution and was analyzed in a seminal paper by Hannenhalli and Pevzner [10] employing breakpoint graphs. The topological framework of [13] constructs a bijection between signed permutations and $\pi$-maps, an equivalence class of particular unicellular fatgraphs, see Section 2 for details. Reversals on signed permutations correspond to three operations on $\pi$-maps, namely gluing, slicing and half-flipping. Fig. 1 illustrates reversals acting on an arbitrary unicellular fatgraph. In the following we will consider gluing, slicing and half-flipping on arbitrary fatgraphs and refer to these simply as reversals. The reversal distance of a signed permutation is connected to the topological genus of the associated $\pi$-map and can be formulated as the solution of a combinatorial optimization problem, namely as the minimum number of reversal actions needed to reduce a $\pi$-map to a tree.

In this paper we study the $r$-distance of a unicellular fatgraph, that is, the minimum number of reversals needed to transform it to a plane tree.

The main result of this paper is an explicit formula for computing the $r$-distance for arbitrary fatgraphs. The formula generalizes the Hannenhalli-Pevzner expression for the reversal distance of signed permutations and can be computed in polynomial time:

$$d(G) = \begin{cases} 
g + h + 1 & \text{if } h \neq 1, h \text{ is odd and all } h \text{ E-blocks are super-blocks}, 
g + h & \text{otherwise}, 
\end{cases}$$

(1.1)

where $g$ is the Euler genus of $G$ and $h$ is the number of E-blocks in $G$, see Section 9.
To prove eq. (1.1), we decompose a unicellular fatgraph into components, i.e. equivalence classes of associated ribbons. Given a unicellular fatgraph, two ribbons $r_1, r_2$ are associated if there exists a sequence of ribbons $(r_1 = w_1, w_2, \ldots, w_{k-1}, w_k = r_2)$, such that each two consecutive ribbons are traversed alternately on the boundary component, see Section 3. A component is trivial if it consists of a single bi-directional ribbon\(^1\). This decomposition allows us to introduce the component tree, $T$, see Section 3.

By deleting trivial components, $G$ disects into blocks. This gives rise to the block tree of $G$, see Section 3. A block containing only orientable components is called orientable and exposed (E-block) if it is orientable and not contained in a path in the block tree, joining two other orientable blocks.

Eq. (1.1) follows in three steps: first we show that any non-orientable component can be reduced into a tree via $g$ reversals. Here we can identify upfront the set of mono-directional ribbons, whose slicing produces a fatgraph consisting of exclusively non-orientable as well as trivial components. Compared with [13] and [10], our proof provides new insight and is constructive. We can explicitly characterize the relevant ribbons as opposed to showing the mere their existence via a proof by contradiction.

Calling a block, containing at least one non-orientable component, non-orientable we secondly show that any non-orientable block can be transformed into a tree via exactly $g$ slicings. The idea will be to successively

---
\(^1\)corresponding to the cut edges, considered in graph theory
merge two non-orientable and orientable components into a single, non-orientable component via slicings. By successively eliminating orientable components, we thereby reduce the problem to non-orientable components.

Thirdly, it remains to consider blocks containing exclusively orientable components. Their analysis involves certain paths within the block tree and blocks, not contained in these paths, i.e. exposed blocks (E-blocks). The E-blocks are of central importance for deriving eq. (1.1). Certain blocks contained in the above mentioned paths are merged into a single non-orientable block by means of gluings and half-flippings. This reduces the problem to non-orientable blocks.

Fatgraphs appear in the context of RNA pseudoknots, RNA structures that exhibit cross-serial interactions. Pseudoknot RNA is of relevance in the context of aging (RNA telomerase) [3] and disease (non-coding RNA) [8]. The topological classification of RNA pseudoknots, employing orientable fatgraphs, has been studied in [17, 1] and led to folding algorithms that construct a partition function of such configurations in $O(n^6)$ time and $O(n^4)$ space complexity [21]. [12] presents a context-free grammar for uniformly generating RNA structures of fixed topological genus. The grammar is facilitated by a bijection reducing the genus of orientable, unicellular fatgraphs by successively tri-slicing vertices [5].

The framework presented here deals with arbitrary fatgraphs and employs reversals, i.e. bi-slicing vertices, by which we transform arbitrary, unicellular fatgraphs into trees. We present a simple formula for the minimum number of reversals needed to obtain such a tree. This constitutes as a first step towards a bijection between unicellular fatgraphs and unicellular fatgraphs of lower genera along the lines of [5]. Such a "constructive" genus reduction has the potential to allow to recursively construct more complex RNA structures or proteins, extending the results of [12].

The paper is organized as follows: in Section 2, we introduce fatgraphs. In Section 3, we discuss irreducibility and components and introduce the notion of crossing ribbons and show that any unicellular fatgraph can be decomposed into components. In Section 4 we study reversals on fatgraphs and in Section 5 we analyze the effect of reversals on the direction and crossing of ribbons. We prove in Section 6 that for any non-orientable component its $r$-distance equals its genus $g$. Section 7 investigates the effect of reversals on the component tree $T$ and the block tree $B$. In Section 8, we show that any fatgraph with only non-orientable blocks has its $r$-distance $g$. Finally we analyze orientable blocks and prove an analogue of the Hannenhalli-Pevzner
formula for the $r$-distance of arbitrary fatgraph in Section 9.

2. Fatgraphs

Following [13], a fatgraph can be expressed using sectors, i.e. oriented wedges, represented as pairs $(i, \omega(i))$, where $i$ is a label and $\omega(i) \in \{-1, +1\}$ is the orientation. We denote counterclockwise and clockwise orientations by $\omega(i) = +1$ and $\omega(i) = -1$, respectively. By abuse of notation, we shall identify a sector with its label and only consider a sector as a pair, when its orientation is of particular relevance.

**Definition 1.** A rooted fatgraph with $n$ ribbons is a quadruple $G = ([2n + 1], \sigma, \gamma, \omega)$, where $[2n + 1] = \{1, \ldots, 2n + 1\}$ is a set of labeled sectors, $\sigma$ and $\gamma$ are permutations representing vertices and boundary components with $\gamma(2n + 1) = \sigma(2n + 1) = 1$, the function $\omega: [2n + 1] \rightarrow \{-1, +1\}$ represents the orientation of sectors with $\omega(1) = \omega(2n + 1)$, and finally for each pair $(x, \sigma(x))$ (where $x \neq 2n + 1$), there exists a unique pair $(y, \sigma(y))$ such that

- **Case 1:**
  
  \[
  \begin{align*}
  x & \xrightarrow{\sigma} \sigma(x) \\
  \downarrow & \quad \downarrow \\
  \gamma \quad \gamma & \\
  \sigma(y) & \xleftarrow{\sigma} y
  \end{align*}
  \]
  
  with $\omega(x) = \omega(\sigma(y)), \omega(\sigma(x)) = \omega(y)$,

- **Case 2:**
  
  \[
  \begin{align*}
  x & \xrightarrow{\sigma} \sigma(x) \\
  \downarrow & \quad \downarrow \\
  \gamma & \\
  \sigma(y) & \xleftarrow{\sigma} y
  \end{align*}
  \]
  
  with $\omega(x) = -\omega(y), \omega(\sigma(x)) = -\omega(\sigma(y))$.

The directions of the $\gamma$-verticals are implied by the orientations of pairs of sectors $(x, \sigma(x))$ and $(y, \sigma(y))$ and we shall refer to the above diagrams as untwisted and twisted ribbons, respectively.

The above definition is equivalent to that in [13], since a rooted fatgraph can be obtained from a fatgraph via bisecting the sector, 1, into two sectors 1 and $2n + 1$, see Fig. 2 (A) and (B). In the presentation of a fatgraph, we will use an arrow or a sign to denote orientation.

A ribbon is given by two pairs of sectors, $((x, \sigma(x)), (y, \sigma(y)))$. Ribbons with mono- and bi-directional verticals are called $m$- and $b$-ribbons, respectively. For $m$-ribbons we have $\omega(x) = -\omega(\sigma(x))$ and $\omega(y) = -\omega(\sigma(y))$, while
Fig. 2: Fatgraphs. (A) A non-orientable fatgraph as in [13]. (B) The same fatgraph, using our notation: $\sigma = (1, 5, 3, 7)(2, 4, 6)$, $\gamma = (1, 2, 3, 4, 7)(5, 6)$ and orientations $\omega(1) = \omega(2) = \omega(5) = \omega(6) = \omega(7) = +1$, $\omega(3) = \omega(4) = -1$. The ribbon $((1, 5), (6, 2))$ is untwisted and bi-directional, while the ribbon $((5, 3), (4, 6))$ is untwisted and mono-directional. (C) Flipping: vertex $(2, 4, 6)$ in (B) is flipped into $(2, 6, 4)$ in (C).

for $b$-ribbons $\omega(x) = \omega(\sigma(x))$ and $\omega(y) = \omega(\sigma(y))$ hold. An untwisted ribbon (also for twisted ribbon) can be either a $m$- or $b$-ribbon, see Fig. 2 (B).

A fatgraph $G$ represents a cell-complex of a surface $F(G)$: the topological quotient space $F(G)$ is obtained by identifying the sides of ribbons consistent with their orientations, see [20]. We call a fatgraph orientable if it induces an orientable surface $F(G)$, and non-orientable, otherwise. Since we work with orientable as well as non-orientable fatgraphs in this paper, we define the genus $g$ of $G$ as the Euler genus $g^*$ (or modified genus) of $F(G)$, i.e.,

$$2 - g^* = v - e + b,$$

where $v, e, b$ are the numbers of vertices, ribbons and boundary components.

A flip of a vertex $v$ reverses the cyclic ordering of its sectors and changes their respective orientations, see Fig. 2 (B) and (C). Two fatgraphs $G_1$ and $G_2$ are isomorphic if $G_2$ can be obtained from $G_1$ by relabeling of sectors and flipping of vertices. By abuse of notion we shall refer to an equivalence class of isomorphic fatgraphs simply as a fatgraph.

A unicellular fatgraph is a fatgraph in which $\gamma$ is a unique cycle. Without loss of generality we can assume that $\gamma = (1, \ldots, 2n + 1)$. In particular, a unicellular fatgraph of genus zero is a plane tree. A $\pi$-map [13] is a unicellular fatgraph having a distinguished vertex (called the center) and any other vertices (called external vertices), such that any ribbon is incident to the center and an external vertex. We generalize a criterion for orientability of $\pi$-maps [13] to arbitrary unicellular fatgraphs.
Proposition 1 (Orientability). A unicellular fatgraph $G$ is non-orientable if and only if $G$ contains a $m$-ribbon.

3. Some basic facts

In this section, we introduce irreducibility and decompose unicellular fatgraphs into their components, extending the results of [13].

Let $G = ([2n + 1], \sigma, \gamma, \omega)$ be a unicellular fatgraph. Given a ribbon $r = ((x, \sigma(x)), (y, \sigma(y)))$, the origin $r^L$ and the terminus $r^R$ of $r$ are $r^L = \min_{\gamma} \{x, \sigma(x), y, \sigma(y)\}$ and $r^R = \max_{\gamma} \{x, \sigma(x), y, \sigma(y)\}$, respectively. In the traversal of the boundary component, $r^L$ is the first $r$-sector, whereas $r^R$ is the last.

Two ribbons $r_1$ and $r_2$ are $G$-crossing or simply crossing if

$$r^L_1 <_\gamma r^L_2 <_\gamma r^R_1 <_\gamma r^R_2 \quad \text{or} \quad r^L_2 <_\gamma r^L_1 <_\gamma r^R_2 <_\gamma r^R_1.$$ 

Otherwise, $r_1$ and $r_2$ are called $G$-non-crossing. Two ribbons are $G$-associated or associated if there exists a sequence of ribbons $(r_1 = w_1, w_2, \ldots, w_{k-1}, w_k = r_2)$ such that $w_i, w_{i+1}$ are $G$-crossing.

Definition 2. (Irreducibility) A unicellular fatgraph $G$ is irreducible if any two $G$-ribbons $r_1$ and $r_2$ are associated.

Clearly, ribbon association is an equivalence relation and the set of ribbons partitions into equivalence classes of associated ribbons, see Fig. 3 (A).

Definition 3. (Component) A $G$-component is an equivalence class of associated ribbons in $G$. A $G$-component is trivial if it consists of only one $b$-ribbon.

The following proposition is straightforward, see the Supplementary Material (SM).

Proposition 2. For any unicellular fatgraph $G$ the following assertions hold:

1. $G$ uniquely decomposes into a set of components $C$,
2. each component $C$ induces an irreducible unicellular fatgraph $G_C$,
3. $g_G = \sum_C g_{G^C}$.

Fig. 3 (A) depicts a decomposition of a fatgraph into components.

For any two sectors $i$ and $j$, we call $[i, j]_\gamma = \{s : i \leq_{\gamma} s \leq_{\gamma} j\}$ the interval of $\gamma$ and $(i, j)_\gamma$ and $(i, j)]_{\gamma}$ are defined analogously. The $C$-trace $I_C$ is the set of sectors of $C$-ribbons. Clearly, we have $I_C = \bigcup_{i} [a_i, c_i]_{\gamma}$, where $a_1 <_{\gamma} \cdots <_{\gamma} a_l$ and $[c_i, a_{i+1}]_{\gamma}$ is a $C$-gap, where $1 \leq i \leq l$. 

7
Proposition 3. Any two traces of components are either subsequent or nested, i.e. one is contained in a gap of the other.

The proof is straightforward and given in the SM. Let $G$ be a fatgraph having the components $C_1,\ldots,C_k$ with traces $I_{C_h} = \bigsqcup [c_{i_{h}}, a_{i_{h}+1}]_\gamma$, where $[c_{i_{h}}, a_{i_{h}+1}]_\gamma$ are $C_h$-gaps, $1 \leq h \leq k$. Clearly, the set of gaps is partially ordered by set inclusion.

Definition 4. (Component Tree) The component tree $T$ of $G$ is a rooted, bicolored tree having $V_1 = \{C_1,\ldots,C_k\}$ (black) and $V_2 = \{(c_{i_{h}},a_{i_{h}+1})_\gamma \mid 1 \leq h \leq k, i\} \cup \{[1,2n+1]_\gamma\}$ (white) as vertices. $T$ has the following adjacency relations: a white vertex is a child of a black vertex $C_h$ if it is a $C_h$-gap, a black vertex is a child of the white root if its trace is not contained in any gap, and a black vertex, $C_j$ is a child of a white vertex, $[c_{i_{h}}, a_{i_{h}+1}]_\gamma$, if $[c_{i_{h}}, a_{i_{h}+1}]_\gamma$ is the minimum gap containing the $C_j$-trace.

Remark: Proposition 3 guarantees that $T$ has no cycles and thus is well-defined.

Fig. 3 (A) and (B) present a unicellular fatgraph and its component tree. Each trivial component corresponds to a cut vertex or a leave in the component tree $T$. By deleting all vertices corresponding to trivial components, we obtain a set of subtrees $B_h$ of $T$. A block $B_h$ of $G$ is the fatgraph induced by the components corresponding to $B_h$, see Fig. 3 (C).
The set of sectors of $B_h$-ribbons is called the $B_h$-trace, $I_{B_h}$. The trace
$I_{B_h} = \bigcup_{k=1}^{k_h} \left[ a^{(h)}_j, c^{(h)}_j \right]_{\gamma}$, where $[c^{(h)}_j, a^{(h)}_{j+1}]_{\gamma}$ are the $B_h$-gaps for $1 \leq j < k_h$. As in Proposition 3, any two blocks are either subsequent around the boundary component or one is contained in a gap of the other. This property allows us to define a tree on blocks.

**Definition 5. (Block Tree)** The block tree, $B$, is the rooted bicolored tree having the set of trivial components and blocks, $V_1$ (black) and the set of all gaps of trivial components and blocks, as well as the root $[1, 2n + 1]_{\gamma}$, $V_2$ (white) as vertex set. The adjacency relations in $T$ are as follows: a white vertex $v_2$ is a child of a black vertex $v_1$ if $v_2$ is a $v_1$-gap. A black vertex is a child of the white root if its trace is not contained in any gap. A black vertex $v_1$ is a child of a white vertex $v_2$ if $v_2$ is the minimal gap containing the $v_1$-trace.

**Remark:** A gap is called a boundary gap if it is contained in some $B_h$ and adjacent to some trivial component. Given a component tree $T$, by definition, its corresponding block tree $B$ is constructed by first contracting each subtree $B_h$ excluding boundary gaps into a black vertex $v_i$ and then each boundary gap is connected with its corresponding black vertex $v_i$ and trivial components, see Fig. 3 (B) and (D).

4. Reversals

In this section, we generalize the reversal operations on $\pi$-maps [13] to unicellular fatgraphs. Given two sectors $i, j$ with $i < j$, we consider

**Gluing:** suppose $i$ and $j$ are located at two distinct vertices $v_1 = (i, i_1, \ldots, i_p)$ and $v_2 = (j, j_1, \ldots, j_q)$. Without loss of generality (flipping), we can always assume that $i$ and $j$ have different orientations, i.e. $\omega(i) = -\omega(j)$. Then gluing $v_1$ and $v_2$ along sectors $i$ and $j$, produces the new vertex

$$v = (i, j_1 \ldots, j_q, j, i_1, \ldots, i_p),$$

and the fatgraph $\bar{G}$, in which $v_1$ and $v_2$ merge into $v$, see Fig. 4.

**Slicing:** suppose $i$ and $j$ are located at $v = (i, j_1 \ldots, j_q, j, i_1, \ldots, i_p)$ and have different orientations, i.e. $\omega(i) = -\omega(j)$. Then by slicing $v$ along sectors $i$ and $j$, we derive the vertices $v_1$ and $v_2$

$$v_1 = (i, i_1, \ldots, i_p), \quad v_2 = (j, j_1 \ldots, j_q),$$
and the fatgraph $\tilde{G}$, in which $v$ is replaced by $v_1$ and $v_2$, see Fig. 5.

**Half-flipping:** suppose $i$ and $j$ are located at $v = (i, j_1, \ldots, j_q, j, i_1, \ldots, i_p)$, have the same orientation and $\omega(i) = \omega(j) = +1$. Then half-flipping is obtained as follows: we first split the vertex $v$ along the sectors $i$ and $j$ into $v_1$ and $v_2$, secondly flip $v_2$ containing the sector $j$ and thirdly glue $v_1$ and $v_2$ along sectors $i$ and $j$ into $\tilde{v}$, see Fig. 6. This sequence of operations is tantamount to flipping all ribbons attaching to $v$ from sector $i$ to sector $j$ counterclockwise (clockwise if $\omega(i) = \omega(j) = -1$) and produces the new fatgraph $\tilde{G}$, in which $v$ is replaced by $\tilde{v} = (i, j_q, \ldots, j_1, j, i_1, \ldots, i_p)$, see Fig. 6.

**Lemma 1.** Let $G$ be a unicellular fatgraph with the boundary component

$$\gamma = (1, \ldots, i-1, i, i+1, \ldots, j-1, j, j+1, \ldots, 2n+1).$$
Then \( \tilde{G} \) is unicellular and has the boundary component

\[
\tilde{\gamma} = (1, \ldots, i - 1, i, j - 1, \ldots, i + 1, j, j + 1, \ldots, 2n + 1).
\]  

(4.1)

In case of gluing or slicing, we have

\[
\tilde{\omega}(k) = \begin{cases} 
\omega(k) & \text{if } k \leq \gamma_i \text{ or } k \geq \gamma_j \\
-\omega(k) & \text{if } i + 1 \leq k \leq \gamma_j - 1.
\end{cases}
\]

(4.2)

Thus gluing, slicing, or half-flipping manifest as a reversal of the boundary component and we shall refer to them as reversals or \( i, j \)-reversals. In Fig. 7, we illustrate the three types of reversals.

The \( r \)-distance, \( d(G) \), of a unicellular fatgraph \( G \) is the minimum number of reversals needed to transform it into a unicellular fatgraph of genus zero, i.e. a tree.

**Corollary 1.** Any unicellular fatgraph \( G \) of genus \( g \) has \( d(G) \geq g \).

By construction, each reversal reduces the number of vertices by at most one. The corollary follows directly from Euler’s characteristic formula.

5. **Reversals and ribbons**

In this section, we investigate how reversals affect ribbons. Let \( r \) denote a \( G \)-ribbon and \( r^{i,j} \) denote the \( \tilde{G} \)-ribbon that is produced from \( r \) by the \( i, j \)-reversal. By construction, there is a natural bijection between ribbons of \( G \) and those of \( \tilde{G} \).

A ribbon \( r \) intersects an interval \( [i, j]_\gamma \) if \( r^L < \gamma i < r^R \leq \gamma j \) or \( i \leq \gamma r^L < \gamma j < \gamma r^R \). In some sense, this generalizes the notion of crossing: two ribbons \( r_1 \) and \( r_2 \) cross in \( G \) if and only if \( r_1 \) intersects the interval \( [r_2^L, r_2^R]_\gamma \). The directional status of a ribbon \( r \) indicates it is mono- or bi-directional.
Lemma 2. The directional status of \( r \) changes after the \( i,j \)-reversal if and only if \( r \) intersects the interval \([i,j]\). 

The lemma follows via direct verification of the orientation of the two pairs \( \{r^L, \gamma(r^L)\} \) and \( \{\gamma^{-1}(r^R), r^R\} \), utilizing eq. (4.2), see the SM.

Let \( r_1, r_2 \) be two ribbons and let \( \text{cr}(r_1, r_2) = 1, 0 \) depending on whether \( r_1, r_2 \) are crossing or not. We call \( \text{cr}(r_1, r_2) \) the crossing status.

Lemma 3. An \( i,j \)-reversal changes \( \text{cr}(r_1, r_2) \) if and only if \( r_1, r_2 \) intersect \([i,j]\).

Proof. We first establish a criterion for maintaining the crossing status of two ribbons.

Criterion: if we have the equality of sets \( (r^L_1, r^R_1) = ((r^{ij}_1)^L, (r^{ij}_1)^R) \), then \( r_1 \) and \( r_2 \) cross in \( G \) if and only if \( r^{ij}_1 \) and \( r^{ij}_2 \) cross in \( \tilde{G} \).

In view of \( (r^L_1, r^R_1) = ((r^{ij}_1)^L, (r^{ij}_1)^R) \), the set of intermediate sectors between origin and terminus of \( r_1 \) and \( r^{ij}_1 \) remains unchanged. We verify that \( r_2 \) intersects \( [r^L_1, r^R_1] \) if and only if \( r^{ij}_2 \) intersects \( [(r^{ij}_1)^L, (r^{ij}_1)^R] \).

Claim 1: if \( r_1 \) does not intersect the interval \([i,j]\), then \( r_1 \) and \( r_2 \) cross in \( G \) if and only if \( r^{ij}_1 \) and \( r^{ij}_2 \) cross in \( \tilde{G} \).

We distinguish the four cases: \( r^L_1 <_\gamma i <_\gamma j <_\gamma r^R_1, \ i <_\gamma r^L_1 <_\gamma r^R_1 <_\gamma j, \ r^L_1 <_\gamma i <_\gamma j, \ i <_\gamma j <_\gamma r^L_1 <_\gamma r^R_1 \). Since all cases can be argued analogously, it suffices to prove the first: in view of eq. (4.1), the ribbon \( r^{ij}_1 \) has the same origin and terminus as \( r_1 \), i.e. \( (r^{ij}_1)^L = r^L_1 \) and \( (r^{ij}_1)^R = r^R_1 \). Moreover, \( (r^{ij}_1)^L <_\tilde{\gamma} i <_\tilde{\gamma} j <_\tilde{\gamma} (r^{ij}_1)^R \). The boundary component \( \tilde{\gamma} \) reverses only the sequence \( i + 1, \ldots, j - 1 \), contained in \([r^L_1, r^R_1]\) and \([((r^{ij}_1)^L, (r^{ij}_1)^R)]\). Thus \( [r^L_1, r^R_1] = [((r^{ij}_1)^L, (r^{ij}_1)^R)] \) and the Criterion implies that \( r_1, r_2 \) cross in \( G \) if and only if \( r^{ij}_1, r^{ij}_2 \) cross in \( \tilde{G} \).

It remains to show

Claim 2: if \( r_1 \) and \( r_2 \) intersect the interval \([i,j]\), then they change their crossing status.

Without loss of generality we may assume that \( r^L_1 <_\gamma r^L_2 \) and \( r_1, r_2 \) are \( G \)-crossing. We distinguish three cases: \( i <_\gamma r^L_1 <_\gamma r^L_2 <_\gamma j <_\gamma r^R_1 <_\gamma r^R_2, \ r^L_1 <_\gamma i <_\gamma r^L_2 <_\gamma r^R_1 <_\gamma j <_\gamma r^R_2 \) and \( i <_\gamma r^L_1 <_\gamma i <_\gamma r^R_1 <_\gamma r^R_2 <_\gamma j \). It suffices to check the first case, as the other two follow analogously. Then we have

\[
i <_\gamma r^L_1 <_\gamma \gamma(r^L_1) <_\gamma r^L_2 <_\gamma \gamma(r^L_2) <_\gamma j \leq \gamma^{-1}(r^R_1) <_\gamma r^R_1 <_\gamma \gamma^{-1}(r^R_2) <_\gamma r^R_2.
\]
Since the boundary component \( \tilde{\gamma} \) differs only by reversing the part \( i + 1, \ldots, j - 1 \), we obtain

\[
i \leq \tilde{\gamma}(r^L_2) < \tilde{\gamma}(r^L_1) < \gamma^{-1}(r^R_1) < \gamma^{-1}(r^R_2) < \gamma(r^R_j) \leq \gamma(r^L_1) < \gamma(r^L_2) < \gamma^{-1}(r^R_1) < \gamma^{-1}(r^R_2) < \gamma(r^R_j),
\]

and \( (r^L_{1,j})^L = \gamma(r^L_1), \quad (r^L_{1,j})^R = r^R_1, \quad (r^L_{2,j})^L = \gamma(r^L_2), \quad (r^L_{2,j})^R = r^R_2 \). Thus

\[
i \leq \tilde{\gamma}(r^L_{2,j}) < \tilde{\gamma}(r^L_{1,j}) < \gamma(r^R_{1,j}) < \gamma(r^R_{2,j}) < \gamma^{-1}(r^R_{1,j}) < \gamma^{-1}(r^R_{2,j}),
\]

i.e. \( r^L_{1,j} \) and \( r^L_{2,j} \) do not cross in \( \tilde{G} \).

**Remark:** In the following we oftentimes write \( r \) instead of \( r^i,j \).

### 6. Non-orientable irreducible fatgraphs

In this section we show that any non-orientable component \( C_r \) of \( G \) and thus any non-orientable, irreducible fatgraph, can be transformed into a tree using \( g_i \) reversals.

We begin by considering a particular class of slicings, namely that of \( m \)-ribbons, \( r \), at \( \gamma(r^L) \) and \( r^R \). Any \( m \)-ribbon \( r \) consists of a set of four sectors, \( \{r^L, \gamma(r^L), \gamma^{-1}(r^R), r^R\} \), where \( \gamma(r^L) \) and \( r^R \) (\( r^L \) and \( \gamma^{-1}(r^R) \)) are located at the same vertex and have different orientations, see Fig. 8. Thus, slicing a \( m \)-ribbon is well-defined and results in a new, unicellular fatgraph \( \tilde{G} \). Slicing \( r \) produces the set \( \{r^L, \gamma(r^L), \gamma^{-1}(r^R)\} \), i.e. \( r \) becomes a trivial component of \( \tilde{G} \), see Fig. 8.

Since \( r_1 \) and \( r \) are \( G \)-crossing if and only if \( r_1 \) intersects \([\gamma(r^L), r^R]_\gamma \), we derive the following corollary of Lemma 3.

**Corollary 2.** Suppose that \( r_1, r_2 \neq r \). Then slicing \( r \), changes the crossing status of \( r_1 \) and \( r_2 \) if and only if \( r_1 \) and \( r_2 \) cross \( r \).
Lemma 4. Any irreducible, non-orientable unicellular fatgraph, $G$, can be sliced via a $m$-ribbon into non-orientable and trivial components.

This is the analogue of Theorem 4 of [10] on signed permutations and Lemma 10 of [13] on $\pi$-maps. The proof given here is new: we identify upfront a distinguished $m$-ribbon, whose slicing facilitates the lemma.

Proof. For a fixed $m$-ribbon $r$, any other ribbon can be bi- (b) or mono-directional (m) and either cross $r$ (+) or not (−). This partitions the set of ribbons into $E^+_b(r), E^-_b(r), E^+_m(r)$ and $E^-_m(r)$. Then $z = |E^+_b(r)| + |E^-_m(r)|$ is a function of $r$. Let $e$ be a $m$-ribbon that maximizes $z$ and $G^e$ be the fatgraph obtained by slicing $e$.

To prove the lemma, it suffices to show that any ribbon, $r_1$, is contained in a non-orientable $G^e$-component or a trivial $G^e$-component. To this end we inspect how slicing affects the directional- and crossing status of $r_1$.

- Case 1: $r_1 \in E^+_b(e) \cup E^-_m(e)$. Lemma 2 shows that $r_1$ is mono-directional in $G^e$, whence $r_1$ is contained in a non-orientable component.

- Case 2: $r_1 \in E^+_m(e)$. We distinguish two sub-cases:
  - Case 2(a): there exists a ribbon $r_2$, such that $r_2 \in E^+_b(e) \cup E^-_m(e)$ and $r_2 \not\subseteq E^+_b(r_1) \cup E^-_m(r_1)$.
    If $r_2 \in E^+_b(e)$, then $r_2$ is a $b$-ribbon of $G$ and $r_2 \in E^-_b(r_1)$. Since $r_2$ and $r_1$ cross $e$, Corollary 2 implies that their crossing status changes from $G$-non-crossing to $G^e$-crossing. Case 1 shows that $r_2$ is a $m$-ribbon in $G^e$ and thus $r_1$ is contained in a non-orientable component.
    If $r_2 \in E^-_m(e)$, then $r_2$ is a $m$-ribbon of $G$ and $r_2 \in E^+_m(r_1)$. Since $r_2$ does not cross $e$, Corollary 2 guarantees that $r_2$ does not change its crossing status after slicing $e$. Since $r_2$ and $r_1$ are $G$-crossing, they are $G^e$-crossing, which implies that $r_1$ crosses $m$-ribbon $r_2$ in $G^e$ and thus is in a non-orientable component.
  - Case 2(b): for $r_2 \in E^+_b(e) \cup E^-_m(e)$, we have $r_2 \not\subseteq E^+_b(r_1) \cup E^-_m(r_1)$. This implies $E^+_b(e) \subseteq E^+_b(r_1)$ and $E^-_m(e) \subseteq E^-_m(r_1)$. Since $|E^+_b(e)| + |E^-_m(e)|$ is maximal, we have $E^+_b(e) = E^+_b(r_1)$ and $E^-_m(e) = E^-_m(r_1)$. Moreover $E^-_b(e) = E^-_b(r_1)$ and $E^-_m(e) \backslash \{r_1\} = E^+_m(r_1) \{e\}$. This means that $e$ and $r'$ are $G$-crossing if and only if $r_1$ and $r'$ are $G$-crossing for any other ribbon $r'$. By Corollary 2,
e and \( r' \) are \( G^e \)-crossing if and only if \( r_1 \) and \( r' \) are \( G^e \)-crossing. Therefore both \( e \) and \( r_1 \) are trivial components of \( G^e \).

- Case 3: \( r_1 \in E^{-}_b(e) \). Since \( G \) is irreducible, there exists a sequence of ribbons \( r_1, \ldots, r_k \) such that \( r_i \) and \( r_{i+1} \) are \( G \)-crossing and \( r_i \in E^{-}_b(e) \) for \( 1 \leq i \leq k-1 \), but \( r_k \not\in E^{-}_b(e) \). We shall show that \( r_k \) does not belong to Case 2(b): since \( r_i \) does not cross \( e \), Corollary 2 guarantees that \( r_i, r_{i+1} \) are \( G^e \)-crossing and \( r_1, r_k \) are \( G^e \)-associated, whence \( r_k \) is not a trivial component. As a result \( r_k \) belongs to Case 1 and Case 2(a) and is consequently associated with some \( m \)-ribbon in \( G^e \), and thus \( r_1, r_k \) are in a non-orientable component.

\[ \square \]

**Theorem 1.** Any irreducible, non-orientable fatgraph \( G \) has \( d(G) = g \).

Fig. 9 depicts a sequence of slicings, that transforms an irreducible, non-orientable unicellular fatgraph into a tree.

**7. Reversals on components and blocks**

In this section we study \( i,j \)-reversals, mapping \( G \) into \( \tilde{G} \). Given the component tree \( T \), we define a partial order on \( T \)-vertices by setting \( v_i <_T v_j \) if \( v_i \) is a descendant of \( v_j \) in \( T \). Suppose \( [p,q]_\gamma \) is a \( T \)-gap and the set of \( [p,q]_\gamma \)-children is \( \{C_h\} \) such that \( I_{C_h} = [i^{(h-1)}, c_i^{(h)}]_\gamma \cup \cdots \cup [a_i^{(h)}, i^{(h)}]_\gamma \), where \( p = i^{(0)} <_\gamma i^{(1)} <_\gamma \cdots <_\gamma i^{(k)} = q \).

**Definition 6.** A sector \( s \) is attached to a component \( C \) if \( s \in \bigcup_i [a_i, c_i]_\gamma \), where \( I_C = \bigcup_i [a_i, c_i]_\gamma \). A sector \( s \) is attached to \( [p,q]_\gamma \) if \( s \in \{i^{(0)}, i^{(1)}, \ldots, i^{(k-1)}, i^{(k)}\} \).
Remark: by construction, each sector is attached to a unique vertex in the component tree.

Since there is a natural bijection between ribbons of $G$ and those of $\tilde{G}$, we may, by abuse the notion, describe a $\tilde{G}$-component or a $\tilde{G}$-block by specifying its $G$-ribbons. For example, we may say several $G$-components form one $\tilde{G}$-component, meaning that a $\tilde{G}$-component consists of the ribbons of these $G$-components.

Suppose $i, j$ are attached to $v_i$ and $v_j$ in the component tree, $T$. Let $P_{i,j}$ denote the $T$-path joining $v_i$ and $v_j$.

Lemma 5. (Reversal Lemma) Each $G$-component not on the path $P_{i,j}$ is a $\tilde{G}$-component with its orientability unchanged. Furthermore, if the path $P_{i,j}$ contains at least two $G$-components, then all components on $P_{i,j}$ form the $\tilde{G}$-component $C^*$.

We depict the effect of gluing, slicing and half-flipping for component trees in Fig. 10, Fig. 11 and Fig. 12, respectively.

Proof. We first prove two claims.

Claim 1: given a component $C$ with $I_C = \cup_i [a_i, c_i]_\gamma$, the following statements are equivalent
Fig. 12: The effect of half-flipping on the component tree and the block tree.

1. \( a_1 <_\gamma i <_\gamma c_l \),
2. there exists a \( C \)-ribbon \( r \) such that \( r^L <_\gamma i <_\gamma r^R \),
3. \( v_i \leq_T C \), i.e. vertex \( v_i \) is either \( C \) or a descendant of \( C \) in \( T \).

(1)\( \Rightarrow \) (2): suppose to the contrary that for each \( C \)-ribbon \( r \), we have either \( r^L \geq_i r^R \leq_i \). Note that any ribbon \( r_1 \) satisfying \( r^L_1 \geq_i r^R_1 \) does not cross with any ribbon \( r_2 \) with \( r^R_2 \leq_i r^L_2 \). Since any two \( C \)-ribbons are associated, we can assume that \( r^L \geq_i r^R \) holds for all \( C \)-ribbons. Thus \( a_1 \geq_i \), a contradiction. (2)\( \Rightarrow \) (1): is straightforward. (1)\( \Leftrightarrow \) (3): by construction, if a gap \( [s_1, s_2]_\gamma \) is a child of \( C \), then \( [s_1, s_2]_\gamma \subset [a_1, c_l]_\gamma \). Similarly, if \( C \) is a child of a gap \( [s_1, s_2]_\gamma \), then \( [a_1, c_l]_\gamma \subset [s_1, s_2]_\gamma \). Therefore, we have that any \( i \in (a_1, c_l)_\gamma \), if and only if \( i \) is attached to \( C \) or its descendant in \( T \).

Claim 2: a component \( C \) contains a ribbon \( r \) intersecting the interval \( [i, j]_\gamma \) if and only if \( C \) is contained in \( P^{i,j} \) (\( P^{i,j} \)-component) in \( T \).

\( \Leftarrow \): Suppose \( C \) is a \( P^{i,j} \)-component, we have three cases

1. \( v_i \leq_T C \) but \( v_j \not\leq_T C \). By Claim 1, \( a_1 <_\gamma i <_\gamma c_l \leq j \) and there exists a \( C \)-ribbon \( r \) such that \( r^L <_\gamma i <_\gamma r^R \leq_j \), i.e. \( r \) intersects the interval \( [i, j]_\gamma \).
2. \( v_j \leq_T C \) but \( v_i \not\leq_T C \): this is analogous to the previous case.
3. \( v_i \leq_T C \) and \( v_j \leq_T C \). Suppose that any \( C \)-ribbon \( r \) does not intersect the interval \( [i, j]_\gamma \), that is, \( r^R \leq_i r^L \leq_j \) or \( r^L \leq_i r^R \leq_j \). Since the set of \( C \)-sectors consists of the sectors of its ribbons, \( [i, j]_\gamma \) is contained in a \( C \)-gap. I.e. \( i, j \) are contained in the same \( C \)-gap. As an \( T \)-ancestor of this gap, \( C \) is not a \( P^{i,j} \)-component, a contradiction. Therefore there exists a \( C \)-ribbon \( r \) intersecting the interval \( [i, j]_\gamma \).

\( \Rightarrow \): \( r \) intersects the interval \( [i, j]_\gamma \) and, without loss of generality, we can assume that \( r^L <_\gamma i <_\gamma r^R \leq_j \). Claim 1 implies that \( C \) is either \( v_i \) or its ancestor in \( T \). If \( C \) is not a \( P^{i,j} \)-component, then \( C \) is an ancestor of
the minimum common ancestor of $v_i$ and $v_j$ in $T$. By construction, $i,j$ are contained in the same $C$-gap, whence $[i,j]_\gamma$ is not intersected by any $C$-ribbon, a contradiction to the assumption that a $C$-ribbon $r$ intersects the interval $[i,j]_\gamma$.

By Claim 2, each ribbon of a component not contained in $P_{i,j}$ does not intersect the interval $[i,j]_\gamma$. Lemma 3 guarantees that the reversal does not change its crossing status with any other ribbon. Therefore, each component not contained in $P_{i,j}$, is a $\tilde{G}$-component with its orientability unchanged.

For each $P_{i,j}$-component, $C$, the set of ribbons $E_C$ splits into $E_C^+$ and $E_C^-$, representing the ribbons intersecting and not intersecting $[i,j]_\gamma$, respectively. By Claim 2, we know that $E_C^+ \neq \emptyset$. First suppose that $r_1, r_2$ are two ribbons intersecting the interval $[i,j]_\gamma$. Claim 2 guarantees that they belong to $P_{i,j}$-components. If $r_1, r_2$ belong to two different $P_{i,j}$-components, then $r_1, r_2$ are $\tilde{G}$-crossing. This follows from Lemma 3 since both of them intersect the interval $[i,j]_\gamma$ and change their crossing status from $G$-non-crossing to $\tilde{G}$-crossing. If $r_1, r_2$ belong to the same $P_{i,j}$-component, $C$, we use that fact that $P_{i,j}$ contains at least two distinct components. In view of this fact, we can choose $r \in E_{C_1}^+$, where $C_1$ is a $P_{i,j}$-component other than $C$. Then $r_i$ and $r$ are $\tilde{G}$-crossing for $i = 1, 2$, which implies that $r_1$ and $r_2$ are $\tilde{G}$-associated. Secondly, for each ribbon $w_1$ in $E_C^-$, there exists a sequence of ribbons $(w_1, w_2, \ldots, w_{k-1}, w_k)$ such that $w_i, w_{i+1}$ are $G$-crossing, $w_k \in E_C^+$ and $w_i \in E_C^-$ for $1 \leq i \leq k - 1$. Lemma 3 guarantees that the reversal does not change the crossing status of a $E_C^-$-ribbon with any other ribbon, whence $w_i$ and $w_{i+1}$ are $\tilde{G}$-crossing. Thus each ribbon $w_1 \in E_C^-$ is still associated to $w_k \in E_C^+$ in $\tilde{G}$. Accordingly, any two ribbons of $P_{i,j}$-components become $\tilde{G}$-associated, that is, all components on the path $P_{i,j}$ form one component. $\square$

**Remark:** extending the proof of Lemma 5, we can construct the component tree $\tilde{T}$ of $\tilde{G}$ utilizing Lemma 1, see the SM.

We call two components adjacent if their corresponding vertices in the component tree have distance two, i.e. either one is the child of a gap of the other component, or both are children of the same gap. Two components, that are adjacent and contain some $m$-ribbon incident to their common vertex, are called $m$-adjacent.

**Proposition 4.** If $C$ is a $G$-component not contained in $P_{i,j}$ that is adjacent to a $P_{i,j}$-gap or a $P_{i,j}$-component, then $C$ is adjacent to $C_*$ in $\tilde{G}$.
Proposition 5. Any two adjacent $G$-components not contained in $P^{i,j}$ are either again adjacent or both are adjacent to $C_*$ in $\tilde{G}$. In particular, any two adjacent $G$-components not contained in $P^{i,j}$ are in the same block of $\tilde{G}$.

We next analyze how reversals affect blocks. A block, $B$, is covered by a path, $Q$, in the block tree if $B$ is either contained in $Q$ or $B$ is adjacent to a gap contained in $Q$.

Suppose $v_i$ and $v_j$ belong to blocks $B_i$ and $B_j$, respectively. Let $Q^{i,j}$ denote the path joining $B_i$ and $B_j$ in the block tree $B$ of $G$.

Lemma 6. Each $G$-block not covered by $Q^{i,j}$ is a $\tilde{G}$-block of the same orientability. Furthermore, if the path $P^{i,j}$ contains at least two $G$-components, then all blocks covered by $Q^{i,j}$ together with the trivial components contained in $Q^{i,j}$ form the $\tilde{G}$-block $B_*$.

Proof. In the block tree, $B$, each $G$-block not covered by $Q^{i,j}$ is adjacent to some trivial components not contained in $Q^{i,j}$. Proposition 5 implies that this block is adjacent to the same set of trivial components and thus consists of the same components in $\tilde{G}$ and this $\tilde{G}$-block has the same orientability. Suppose that $B$ is a $G$-block covered by $Q^{i,j}$. By definition, there exists a $B$-component, $C_B$, that is either contained in $Q^{i,j}$ or adjacent to a gap contained in $Q^{i,j}$. If $C_B$ is contained in $Q^{i,j}$, Lemma 5 shows that $C_*$ contains all $C_B$-ribbons. If $C_B$ is adjacent to a $Q^{i,j}$-gap, Proposition 4 implies that $C_B$ is adjacent to $C_*$ in $\tilde{G}$. Proposition 5 guarantees that any $B$-component is contained in the $\tilde{G}$-block containing $C_*$, i.e. contained in $B_*$. Therefore, $B_*$ consists of ribbons of trivial $Q^{i,j}$-components and blocks covered by $Q^{i,j}$. □

8. The Pac-Man game

A block that contains at least one non-orientable component is called non-orientable and orientable, otherwise. A fatgraph whose blocks are all non-orientable is called block-non-orientable.

In this section, we shall show that block-non-orientable fatgraphs of genus $g$ have $r$-distance $g$.

Proposition 6. (Pac-Man Proposition) Suppose we have an orientable $G$-component, $C_1$, and a non-orientable $G$-component, $C_2$, that are $m$-adjacent. Then there exists an $i_1,i_2$-slicing that generates a fatgraph $\tilde{G}$ such that

1. $C_1$ and $C_2$ merge into a non-orientable $\tilde{G}$-component,
2. any other $G$-component is a $\tilde{G}$-component of the same orientability.
3. if $G$ is block-non-orientable, then $\tilde{G}$ is.
Proof. We shall apply Lemma 5 and to that end identify two sectors $i_1, i_2$ such that $P^{i_1,i_2}$ contains $C_1$ and $C_2$ but no other components. Since $C_1$ and $C_2$ are adjacent, there exists a gap $[g_1,g_2]_\gamma$ adjacent to both $C_1$ and $C_2$ in $T$. Let $v$ be the common vertex of $C_1$ and $C_2$. Since neither $C_1$ nor $C_2$ are trivial components, each contains at least two ribbons incident to $v$. Consequently, there exist at least three $C_1$-sectors as well as three $C_2$-sectors at $v$. Let $i_1$ be a sector of $I_{C_1}\setminus\{g_1,g_2\}$ located at $v$. Since $C_1$ and $C_2$ are $m$-adjacent and all $C_1$-ribbons are $b$-ribbons, $C_2$ contains a $m$-ribbon incident to $v$. That is, there exist two $C_2$-sectors at $v$ with different orientations. Thus we can choose $i_2$ to be a sector of $I_{C_2}\setminus\{g_1,g_2\}$ at $v$ with an orientation, different from $i_1$. Since $i_k \in I_{C_k}\setminus\{g_1,g_2\}$, the sector $i_k$ is attached to either the component $C_k$ or a gap incident to $C_k$, where $k = 1, 2$. Furthermore this gap cannot coincide with $[g_1,g_2]_\gamma$. This implies that the only two $P^{i_1,i_2}$-components are $C_1$ and $C_2$.

In order to prove (1) and (2), Lemma 5 reduces the work to showing that $C_\ast$ is non-orientable. Claim 2 of Lemma 5 guarantees that $C_1$ contains a ribbon, $r$, intersecting the interval $[i,j]_\gamma$. By Lemma 2, $r$ changes from a $b$-ribbon in $\mathcal{G}$ to a $m$-ribbon of $\mathcal{G}$, implying that $C_\ast$ is non-orientable. To establish (3), we let $B_i$ be the block containing $C_1$ and $C_2$. By construction, the path $Q^{i,j}$ covers only the block $B_i$. Lemma 6 implies, that each $\mathcal{G}$-block other than $B_i$ remains a non-orientable $\tilde{\mathcal{G}}$-block and $B_i$ becomes the $\tilde{\mathcal{G}}$-block $B_\ast$ containing $C_\ast$. Since $C_\ast$ is non-orientable, $B_\ast$ is non-orientable, whence $\tilde{\mathcal{G}}$ is block-non-orientable.

Let us next consider the case of two adjacent components, that are not $m$-adjacent.

**Proposition 7.** Suppose $\mathcal{G}$ is block-non-orientable and $C$ is a non-orientable component such that all orientable components adjacent to $C$ are not $m$-adjacent. Then there exists a $m$-ribbon $e$ in $C$, such that $\tilde{\mathcal{G}}$ obtained by slicing $e$ is block-non-orientable.

*Proof.* Let $B_C$ be the $\mathcal{G}$-block containing $C$. By Lemma 4, $C$ can sliced via some $m$-ribbon, $e$, into non-orientable components $C_1, \ldots, C_l$ and trivial components $T_1, \ldots, T_k$, where $T_1 = \{e_1\}, \ldots, T_k = \{e_k\}$.

First we show that each $\mathcal{G}$-block other than $B_C$, is non-orientable in $\tilde{\mathcal{G}}$. Since slicing $m$-ribbon $e$ is defined as slicing with respect to the sectors $\gamma(e^L)$ and $e^R$, the path $Q^{\gamma(e^L),e^R}$ covers only $B_C$. Lemma 6 guarantees that any
\( G \)-block other than \( B_C \) is a \( \tilde{G} \)-block of the same orientability, i.e. it is non-orientable in \( \tilde{G} \).

Since \( B_C \) could be sliced into several \( G \)-blocks, \( \tilde{B}_1, \ldots, \tilde{B}_q \) and trivial components \( T_1, \ldots, T_k \). It remains to show that each block \( \tilde{B}_i \) is non-orientable.

By construction, in each block \( \tilde{B}_i \), there exists a component \( \tilde{C}_i \) adjacent to some trivial component \( T_j \). We distinguish two cases:

1. \( \tilde{C}_i \) is one of the \( C_1, \ldots, C_l \), created by the slicing: as the \( C_1, \ldots, C_l \) are all non-orientable, we are done.

2. \( \tilde{C}_i \) corresponds to a \( G \)-component in \( B_C \). Since \( \tilde{C}_i \) is adjacent to a trivial component \( T_j \) in \( \tilde{G} \), \( \tilde{C}_i \) is adjacent to \( C \) in \( G \). Thus the ribbon \( e_j \) changes its crossing status and becomes the trivial component \( T_j \) after slicing, which implies that \( e_j \) intersects the interval \([\gamma(e^L), e^R]_\gamma \) in \( G \). Consequently, \( e_j \) is a \( m \)-ribbon in \( C \) and incident to the common vertex of \( \tilde{C}_i \) and \( C \). As a result \( \tilde{C}_i \) and \( C \) are \( m \)-adjacent. By assumption, all the orientable components adjacent to \( C \) are not \( m \)-adjacent. Therefore \( \tilde{C}_i \) and \( \tilde{B}_i \) are non-orientable.

\[ \Box \]

**Remark:** Suppose that \( C \) is non-orientable and \( C_1 \) is orientable, adjacent but not \( m \)-adjacent to \( C \). Let \( r \) denote a \( C \)-ribbon incident to their common vertex and as such, \( r \) is a \( b \)-ribbon. Successively applying Lemma 4 to \( C \), i.e. successively slicing \( m \)-ribbons, renders \( r \) as a trivial component at some point. This means that, along this process, \( r \) changes its crossing status and Lemma 2 guarantees that then \( r \) also changes its direction, becoming a \( m \)-ribbon.

Proposition 7 thus shows that after successive \( C \)-slicings, \( C_1 \) becomes in fact \( m \)-adjacent to a non-orientable component derived from successive slicings of \( C \).

Fig. 13 shows how to transform an orientable component into non-orientable components by employing Propositions 6 and 7.

**Theorem 2.** For each block-non-orientable unicellular fatgraph, \( G \) of genus \( g \), we have \( d(G) = g \).

**Proof.** We prove this by induction on \( g \). Clearly, the assertion holds for \( g = 1 \). For \( G \) having genus \( g \), we distinguish two cases:

Case 1: there exist one orientable component and one non-orientable component that are \( m \)-adjacent. Then Proposition 6 allows us to merge the two
Fig. 13: The Pac-Man game: (A): via successive slicings, two components not $m$-adjacent first become $m$-adjacent and then merge into a non-orientable component. (B): the component trees: we represent a non-orientable component by a Pac-Man, an orientable component by a blue dot, and a trivial component by a no-entry sign.

components by slicing and to obtain a block-non-orientable, unicellular fatgraph $\tilde{G}$ of genus $g - 1$.

Case 2: for any non-orientable component, all the orientable components are not $m$-adjacent. By Proposition 7, slicing an appropriate $m$-ribbon, $e$, of a non-orientable component we obtain a block-non-orientable, unicellular fatgraph $\tilde{G}$ of genus $g - 1$.

9. A Hannenhalli-Pevzner formula for $r$-distance

First we study how half-flipping and gluing transforms orientable blocks into non-orientable blocks. Let $G$ be a unicellular fatgraph having an orientable block $B_1$. Let $C$ be an orientable component in $B_1$ and $r$ be a $C$-ribbon. By $r^L, r^R$-half-flipping, we obtain a unicellular fatgraph $\tilde{G}$.

**Proposition 8.** The following assertions hold:
(1) each $G$-block other than $B_1$ is a $\tilde{G}$-block of the same orientability and $B_1$ becomes a non-orientable $\tilde{G}$-block;
(2) the block tree $\tilde{B}$ of $\tilde{G}$ is the same as the block tree $B$ of $G$.

**Proof.** Since $C$ is orientable, $r$ is a $b$-ribbon and sectors $r^L$ and $r^R$ are located at the same vertex and have the same orientation. Thus the $r^L, r^R$-half-flipping is well-defined. Note that the path $P_{r^L, r^R}$ contains only one
component $C$ and thus the path $Q^{r_L,r_R}$ covers only $B_1$. In view of Lemma 6, it remains to show that $B_1$ is non-orientable in $\tilde{G}$. Lemma 3 shows that $C$ is a $\tilde{G}$-component. By construction, $r$ becomes a $m$-ribbon in $\tilde{G}$, whence both $C$ and $B_1$ are non-orientable in $\tilde{G}$. By definition of block trees, the adjacency between any gap and any block or trivial component remains unchanged, proving assertion (2).

Let $B_1$ and $B_2$ be two orientable $G$-blocks and let $Q$ denote the path connecting $B_1$ and $B_2$ in the block tree $B$.

**Proposition 9.** There exists an $i,j$-gluing such that $\tilde{G}$ satisfies

1. each $G$-block not covered by $Q$ is a $\tilde{G}$-block having the same orientability and all blocks covered by $Q$, together with all trivial components contained in $Q$ merge into the non-orientable $\tilde{G}$-block $\tilde{B}_*$.
2. the block tree $\tilde{B}$ of $\tilde{G}$ is obtained from $B$ by contracting all blocks covered by $Q$ and trivial components contained in $Q$ into one block vertex.

**Proof.** Let $C_k$ be an orientable component of $B_k$ where $k = 1, 2$. Let $P$ denote the path joining $C_1$ and $C_2$ in the component tree $T$. Let $[s_k, t_k]$ be the gap adjacent to $C_k$ on $P$ for $k = 1, 2$. We select $i \in I_{C_1} \setminus \{s_1, t_1\}$ and $j \in I_{C_2} \setminus \{s_2, t_2\}$. By construction, the path $P^{i,j}$ contains all the components on $P$. By construction, the path $Q^{i,j}$ joins $B_1$ and $B_2$, i.e. $Q^{i,j} = Q$. In order to prove (1), in view of Lemma 6, it suffices to show that $\tilde{B}_*$ is non-orientable. According to Lemma 5, all components contained in $P$ merge into the $\tilde{G}$-component $C_*$ in $\tilde{B}_*$. Since at least one $b$-ribbon of $C_1$ intersects the interval $[i, j]_{\gamma}$ and changes to $m$-ribbon in $\tilde{G}$, both $C_*$ and $\tilde{B}_*$ are non-orientable. To prove (2), we verify, employing Lemma 1, that the adjacencies between gaps and blocks or trivial components in $\tilde{B}$ are induced from $B$: if a block not covered by $Q$ and a gap are adjacent in $B$, then they remain adjacent in $\tilde{B}$. If a gap is adjacent to a block covered by $Q$ then it is adjacent to $\tilde{B}_*$ in $\tilde{B}$. Therefore the block tree $\tilde{B}$ of $\tilde{G}$ is obtained from $B$ by contracting all blocks covered by $Q$ and trivial components on $Q$ into one block.

In difference to non-orientable blocks, that could be resolved individually, orientable blocks have to be considered as an ensemble. This is due to Proposition 9: the action of reversals affects the entire set of components or blocks.

**Definition 7.** (E-block) A block, $B_i$, is exposed (E-block) if it is orientable and not covered by any path joining any two other orientable blocks in the block tree $B$. Let $h$ denote the number of E-blocks.
E-blocks are closely related to hurdles of signed permutations and π-maps [10, 13].

Let $\bar{B}$ denote the minimum $B$-subtree, containing all orientable blocks. By construction, each leaf of $\bar{B}$ is an orientable block. A block, $B_i$, is an E-block if and only if $B_i$ is a $\bar{B}$-leave and the gap adjacent to $B_i$ has degree two in $\bar{B}$.

As a result, suppose an E-block, $B_i$, is replaced by a non-orientable block, then, any other exposed block remains exposed and furthermore, at most one additional, orientable block becomes exposed.

**Definition 8. (S-block)** An E-block, $B_i$, is called a super-block (S-block) if changing $B_i$ into a non-orientable block does not affect the number of E-blocks in the block tree $B$.

**Proposition 10.** Let $\tilde{G}$ be the fatgraph obtained from $G$ via an $i,j$-reversal. Then

$$g(G) + h(G) - (g(\tilde{G}) + h(\tilde{G})) \leq 1.$$  

(9.1)

That is, a reversal reduces $(g + h)$ by at most one.

**Proof.** We distinguish three cases

1. $i,j$-reversal is a slicing, i.e. $i$ and $j$ are located at the same vertex with different orientations. By construction, they are attached to components or gaps in the non-orientable block, $B_0$. Lemma 6 shows that the $i,j$-reversal affects only $B_0$ and that any E-block persists in $\tilde{G}$. Thus the number of E-blocks of $\tilde{G}$ satisfies $h(\tilde{G}) \geq h(G)$. Since $g(\tilde{G}) = g(G) - 1$, eq. (9.1) holds.

2. $i,j$-reversal is half-flipping, i.e. $i$ and $j$ are located at the same vertex with the same orientation. By construction, they are attached to components or gaps in the block, $B_1$. Lemma 6 shows that the $i,j$-reversal affects only $B_1$. Proposition 8 shows that $\tilde{G}$ has the same block tree as $G$ and all $G$-blocks except $B_1$ have the same orientability. Furthermore, any E-block other than $B_1$ persists in $\tilde{G}$. Thus the number of E-blocks is reduced by at most one, i.e. $h(\tilde{G}) \geq h(G) - 1$ and in view of $g(\tilde{G}) = g(G)$, eq. (9.1) follows.

3. $i,j$-reversal is a gluing, i.e. $i$ and $j$ are located at two distinct vertices. We have two cases

   - $i$ and $j$ are located at two distinct vertices of the same block, $B_2$. Then Lemma 6 shows that all $G$-blocks except $B_2$ keep their
orientability in $\tilde{G}$ and in particular, any E-block other than $B_2$ persists to be an E-block in $\tilde{G}$. Consequently, the number of E-blocks of $\tilde{G}$ is reduced by at most one, i.e. $h(\tilde{G}) \geq h(G) - 1$, whence eq. (9.1).

- $i$ and $j$ are located at two distinct vertices in the two blocks, $B_3$ and $B_4$. Proposition 9 shows that the block tree, $\tilde{B}$, of $\tilde{G}$ is obtained from $B$ by contracting all blocks covered by the path connecting $B_3$ and $B_4$. Thus any E-block other than $B_3$ and $B_4$ remains to be an E-block in $\tilde{G}$. As a result, the number of E-blocks of $\tilde{G}$ is reduced by at most two, i.e. $h(\tilde{G}) \geq h(G) - 2$. In view of $g(\tilde{G}) = g(G) + 1$, eq. (9.1) follows.

\[ \square \]

**Corollary 3.** For any unicellular fatgraph $G$, we have $d(G) \geq g + h$.

**Proposition 11.** If the block tree $B$ has $2k$ E-blocks, then there exists a collection of $k$ $B$-paths joining all E-blocks such that each orientable block is covered by at least one path.

**Theorem 3.** Let $G$ be a unicellular fatgraph of genus $g$ with $h$ E-blocks. Then

\[
d(G) = \begin{cases} 
g + h + 1 & \text{if } h \neq 1, \text{ } h \text{ is odd and all } h \text{ E-blocks are super-blocks,} \\
g + h & \text{otherwise.}
\end{cases}
\]

(9.2)

**Proof.** Suppose first that $h \neq 1$, $h$ is odd and all $h$ E-blocks are S-blocks. Note that, the number of S-blocks cannot be reduced by half-flipping. Since gluing reduces the number of S-blocks by at most two, we need at least $(h + 1)/2$ number of gluings to eliminate all S-blocks. Since each gluing increases the genus by one, we obtain the lower bound $g + (h+1)/2 + (h+1)/2 = g + h + 1$ on $r$-distance $d(G)$.

In view of Corollary 3, it is sufficient to identify a sequence of reversals that realizes the above lower bound. We distinguish the following cases

1. $h = 1$. Then $G$ has only one orientable block and Proposition 8 shows that half-flipping one ribbon in this block transforms $G$ into a block-non-orientable fatgraph $G_1$ of genus $g$. Theorem 2 shows that $d(G_1) = g$, whence $d(G) = g + 1$. 

25
Fig. 14: Eliminating all the orientable blocks. The fatgraph $G$ (Left) has genus 12 and contains 3 E-blocks, 2 of which are S-blocks. All orientable $G$-blocks are transformed into non-orientable blocks in $G'$ (Right), via one half-flipping and one slicing. Theorem 2 shows that $d(G') = g(G') = 13$. Therefore we have $d(G) = g(G) + h(G) = 15$.

2. $h = 2k$. By Proposition 11, there exists a collection of $k$ paths joining all E-blocks and covering all orientable blocks. Proposition 9 implies, that one can always glue blocks covered by each path and thus obtain a block-non-orientable fatgraph, $G_2$, via $k$ gluings. Therefore $d(G_2) = g(G_2) = g + k$ and $d(G) = d(G_2) + k = g + h$.

3. $h = 2k + 1$ and there exists at least one E-block, $B_1$, which is not a S-block. Proposition 8 shows that half-flipping one $B_1$-ribbon transforms $G$ into a fatgraph $G_3$ having genus $g$ and $2k$ E-blocks. As in the case of $h = 2k$, we have $d(G_3) = g + 2k$ and thus $d(G) = d(G_3) + 1 = g + h$.

4. $h = 2k + 1$ and all $h$ E-blocks are S-blocks. Let $B_1, \ldots, B_h$ be all $h$ S-blocks. Set $B'$ to be the minimum subtree of $B$ containing all $B_2, \ldots, B_h$, and let $P'$ be the path connecting $B_1$ and $B'$. By Proposition 9, we obtain the fatgraph $G_4$ by gluing all blocks covered by $P'$. By construction, the number of E-blocks in $G_4$ is reduced to $2k$. The case $h = 2k$ shows that $d(G_4) = g + 1 + 2k$, whence $d(G) = d(G_4) + 1 = g + h + 1$.

Fig. 14 depicts a sequence of reversals eliminating E- and S-blocks.

Acknowledgments

The second author is a Thermo Fisher Scientific Fellow in Advanced Systems for Information Biology and acknowledges their support of this work.
[1] J. E. Andersen, R. C. Penner, C. M. Reidys, M. S. Waterman, Topological classification and enumeration of RNA structures by genus, J. Math. Biol. 67 (5) (2013) 1261–1278.

[2] D. Bessis, C. Itzykson, J. B. Zuber, Quantum field theory techniques in graphical enumeration, Adv. in Appl. Math. 1 (2) (1980) 109–157.

[3] M. A. Blasco, Telomeres and human disease: ageing, cancer and beyond, Nature Reviews Genetics 6 (8) (2005) 611–622.

[4] B. Bollobás, O. Riordan, A Polynomial Invariant of Graphs On Orientable Surfaces, Proc. Lond. Math. Soc. 83 (3) (2001) 513–531.

[5] G. Chapuy, A new combinatorial identity for unicellular maps, via a direct bijective approach, Adv. in Appl. Math. 47 (4) (2011) 874–893.

[6] G. Chapuy, V. Féray, E. Fusy, A simple model of trees for unicellular maps, J. Combin. Theory Ser. A 120 (8) (2013) 2064–2092.

[7] J. A. Ellis-Monaghan, I. Moffatt, Graphs on Surfaces: Dualities, Polynomials, and Knots, SpringerBriefs in Math., Springer-Verlag, New York, 2013.

[8] M. Esteller, Non-coding RNAs in human disease, Nature Reviews Genetics 12 (12) (2011) 861–874.

[9] A. Grothendieck, Esquisse d’un programme, in: L. Schneps, P. Lochak (Eds.), Geometric Galois Actions 1, Vol. 242 of London Math. Soc. Lecture Note Ser., Cambridge University Press, 1997, pp. 5–48.

[10] S. Hannenhalli, P. A. Pevzner, Transforming Cabbage into Turnip: Polynomial Algorithm for Sorting Signed Permutations by Reversals, J. ACM 46 (1) (1999) 1–27.

[11] J. Harer, D. Zagier, The Euler characteristic of the moduli space of curves, Invent. Math. 85 (1986) 457–485.

[12] F. W. D. Huang, C. M. Reidys, Topological language for RNA, Mathematical Biosciences 282 (2016) 109–120.

[13] F. W. D. Huang, C. M. Reidys, A topological framework for signed permutations, Discrete Math. 340 (9) (2017) 2161–2182.
[14] M. Kontsevich, Intersection theory on the moduli space of curves and the matrix airy function, Comm. Math. Phys. 147 (1) (1992) 1–23.

[15] S. K. Lando, A. K. Zvonkin, Graphs on Surfaces and Their Applications, Vol. 141 of Encyclopaedia Math. Sci., Springer-Verlag, Berlin, 2004, with an appendix by Don B. Zagier.

[16] B. Mohar, C. Thomassen, Graphs on Surfaces, Johns Hopkins Stud. Math. Sci., Johns Hopkins University Press, Baltimore, MD, 2001.

[17] H. Orland, A. Zee, RNA folding and large $n$ matrix theory, Nucl. Phys. B 620 (2002) 456–476.

[18] R. C. Penner, Perturbative series and the moduli space of Riemann surfaces, J. Differential Geom. 27 (1) (1988) 35–53.

[19] R. C. Penner, Moduli spaces and macromolecules, Bull. Amer. Math. Soc. 53 (2) (2016) 217–268.

[20] R. C. Penner, M. Knudsen, C. Wiuf, J. E. Andersen, Fatgraph models of proteins, Comm. Pure Appl. Math. 63 (10) (2010) 1249–1297.

[21] C. M. Reidys, F. W. D. Huang, J. E. Andersen, R. C. Penner, P. F. Stadler, M. E. Nebel, Topology and prediction of RNA pseudoknots. Bioinformatics 27 (8) (2011) 1076–1085.