arXiv:1704.07902v4 [math.CO] 17 Jan 2018

A bijection between the set of nesting-similarity classes and L&P matchings

Megan A. Martinez [1]   Manda Riehl[2]

[1] *Department of Mathematics, Ithaca College, Ithaca, NY*
[2] *Department of Mathematics, Rose-Hulman Institute of Technology*

Matchings are frequently used to model RNA secondary structures; however, not all matchings can be realized as RNA motifs. One class of matchings, called the L & P matchings, is the most restrictive model for RNA secondary structures in the Largest Hairpin Family (LHF). The L & P matchings were enumerated in 2015 by Jefferson, and they are equinumerous with the set of nesting-similarity classes of matchings, enumerated by Klazar. We provide a bijection between these two sets. This bijection preserves noncrossing matchings, and preserves the sequence obtained reading left to right of whether an edge begins or ends.

Keywords: perfect matchings, nesting-similarity, bijective combinatorics

Ribonucleic acid (RNA) is an essential molecule found in the cells of all living things. Usually, RNA is formed by a string of nucleotides which folds over on itself, creating secondary bonds. The structure of these secondary bonds is a topic of great interest and has been studied from a biological and mathematical perspective [2, 3]. Mathematically, RNA secondary structures can be modeled by considering each nucleotide as a vertex and bonds between nucleotides that are not part of the RNA backbone as edges. Each vertex is incident to at most one edge and thus the graph obtained is a matching.

We represent these matchings as $2n$ points drawn along a horizontal line (the backbone) and arcs drawn between pairs of points represent the nucleotide bonds. For simplicity, we assume all matchings are *complete* (i.e., all vertices are incident to an edge), and therefore contain n edges. RNA often has nucleotides that are not bonded, but these structures can be reconstructed by adding any number of isolated vertices to a complete matching. We notate the set of complete matchings with n edges by $\mathcal{M}(n)$.

For a matching, M, we denote the set of edges $E(M)$ and label these edges with $[n]$ in increasing order from left to right by their left endpoints. For the edge labeled i, we write $i = (i_1, i_2)$ where i represents the label of the edge and i_1, i_2 represent the position of the left and right endpoints of the edge, respectively. A pair of edges $i = (i_1, i_2)$ and $j = (j_1, j_2)$ are said to be *nested* if $i_1 < j_1 < j_2 < i_2$ and *crossing* if $i_1 < j_1 < i_2 < j_2$. For a matching $M \in \mathcal{M}(n)$, let $ne(M)$ and $cr(M)$ denote the number of pairs of nested edges and crossing edges in M, respectively. Additionally, the edges $i = (i_1, i_2)$ and $j = (j_1, j_2)$ are said to form a *hairpin* if $j_1 = i_1 + 1$ and $j_2 = i_2 + 1$ (Figure 1a). A nested sequence of edges that can be drawn above the backbone is called a *ladder* (Figure 1b).
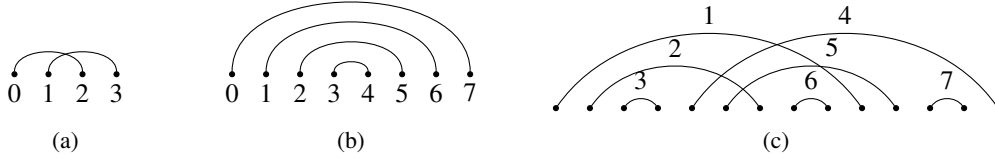
Fig. 1: The left matching is a hairpin, the middle matching is a ladder of four edges, and the right matching is an example of an L & P matching with edges labeled by left endpoint.

Matchings that contain no crossing pairs of edges are called *noncrossing matchings*. The number of noncrossing matchings with n edges is well-known to be counted by the nth Catalan number, $C_n = \frac{1}{n+1}\binom{2n}{n}$, and has been studied in several contexts, including pattern avoidance [1].

For each matching, we can examine the vertices from left to right and list whether vertices are left or right endpoints of an edge. For example, for the matching in Figure 2a, this sequence is LLRLLRRRLR; we call this the *LR-sequence* of a matching. LR-sequences are in bijection with Dyck paths, and each non-crossing matching has a distinct LR-sequence. For any $M \in \mathcal{M}(n)$, define $nc(M)$ to be the noncrossing matching with the same LR-sequence as M.

In order to model RNA structures that contain crossings, different families of matchings have been studied. In this paper, we focus on the family of L & P matchings. The L & P matchings were first rigorously defined by Condon et al. [2] and this definition was later refined by Jefferson [3]. Each matching in the L & P family can be constructed inductively by starting from either a hairpin or a single edge, and either a) inflating an edge by a ladder, or b) inserting a noncrossing matching into an L & P matching [3]. For an example of an L & P matching, see Figure 1c. The counting sequence for the number of L & P matchings with n edges begins 1, 3, 12, 51, 218, 926, 16323, 67866, 280746 and is given by the formula $2 \cdot 4^{n-1} - \frac{3n-1}{2n+2}\binom{2n}{n}$ [3].

It was noted by Jefferson that L & P matchings with n edges, which we denote $\mathcal{LP}_n$, are equinumerous to the equivalence classes on matchings with n edges determined by the nesting-similarity equivalence, $\sim_{ne}$, defined by Klazar [4]. We say that two matchings $M, N \in \mathcal{M}(n)$ are *nesting-similar*, and write $M \sim_{ne} N$, if and only if $ne(M) = ne(N)$ and M, N have the same LR-sequence. An example of this equivalence is shown in Figure 2. Klazar showed that there are $2 \cdot 4^{n-1} - \frac{3n-1}{2n+2}\binom{2n}{n}$ nesting-similarity equivalence classes for matchings in $\mathcal{M}(n)$; however, no explicit bijection was discovered connecting L & P matchings to nesting-similarity classes. Our main result is to construct a bijective correspondence between these two structures.

Klazar [4] utilized a different, but equivalent, definition for nesting-similarity in his work. He examined the *tree of matchings* where the vertex set is the infinite set of matchings, $\bigcup_{n=0}^{\infty} \mathcal{M}(n)$. Matchings are
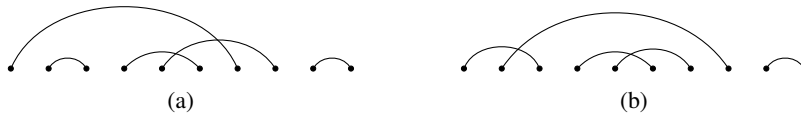

Fig. 2: Two nesting-similar matchings. Each matching has LR-sequence LLRLLRRRLR and two pairs of nested edges.

connected if one can be obtained from the other by inserting a new edge whose left endpoint occurs earliest. With this construction, two matchings are said to be nesting-similar if we can record the number of nestings for the children of each matching, which have one added edge, and obtain the same multiset. It is straightforward to show that this definition is equivalent to the definition above.

In his paper, Klazar shows a correspondence between his nesting-similarity matchings and tunnel pairs in Dyck paths. Our bijection between L & P matchings and the set of nesting-similarity classes will be a composition of two bijections. The composed mapping will pass through the set of nestings in noncrossing matchings, which correspond to tunnel pairs in Dyck paths. Define $\mathcal{NCN}_n = \{(M, a, b) \mid M \in \mathcal{M}(n)$ noncrossing; $a = b = 0$ or edges $a < b$ are nested in $M\}$; so, $\mathcal{NCN}_n$ is the set of noncrossing matchings with a chosen nested pair of edges ($a = b = 0$ indicates no nested pair has been indicated).

In Section 1, we define a bijection between L & P matchings and $\mathcal{NCN}_n$, and in Section 2, we define a bijection between nesting-similarity classes and $\mathcal{NCN}_n$. Thus We show that both L & P matchings and nesting-similarity classes are equinumerous to $\mathcal{NCN}_n$.

1 L & P matchings and noncrossing matchings

The process of constructing an L & P matching implies that such a matching contains a crossing exactly if the matching can be built inductively from a hairpin. As a result, any L & P matching that contains a crossing will have all crossings occur in an inflated hairpin. Given a matching, M, we will label edges by left endpoint, as in Figure 1c. In this Figure, edges 1, 2 crossing edges 4, 5 comprise an inflated hairpin. Below we provide a precise definition of this structure.

Definition 1. *A* maximal inflated hairpin *in an L & P matching is two sets of edges $A = \{a_1, a_2, \ldots, a_k\}$ and $B = \{b_1, b_2, \ldots, b_\ell\}$ such that*

1. *every pair of edges in A and every pair of edges in B is nested,*

2. *for every $a_i \in A$, a_i crosses every edge in B, and*

3. *every crossing in M occurs between edges in A and B.*

We let A be the set of edges with smaller labels (the left side of the inflated hairpin), and we say that M contains the inflated hairpin (A, B).

Any L & P matching consists of a maximal inflated hairpin with noncrossing matchings inserted below the hairpin as in Figure 3. It is possible for the inflated hairpin to be empty, yielding a noncrossing matching. For example, the matching in Figure 2a is L & P: the first and third edges crossing the fourth edge form an inflated hairpin. However, the matching in Figure 2b is not L & P, since the first four edges are all involved in crossings, but these four edges do not form an inflated hairpin.

Notice that every edge not in the inflated hairpin of an L & P matching must begin and end between two vertices that are adjacent in the inflated hairpin, otherwise it would be part of an inflated hairpin itself. This fact will be used in our bijection.

Lemma 1. *Let $M \in \mathcal{LP}_n$ such that M contains an inflated hairpin (A, B). Then the right endpoints of the inflated hairpin appear in the order $(a_k, a_k - 1, \ldots, a_1, b_\ell, b_\ell - 1, \ldots, b_1)$ and can be rearranged to $(b_\ell, b_\ell - 1, \ldots, b_1, a_k, a_k - 1, \ldots, a_1)$ to construct $nc(M)$. Additionally, every pair of edges in $A \cup B$ are nested in $nc(M)$.*
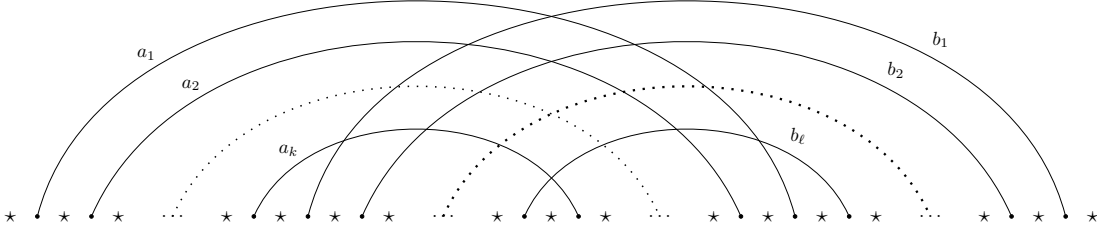
Fig. 3: The structure of an L & P matching; Each $\star$ indicates a position where a noncrossing matching may be inserted. The edges $a_1, a_2, \ldots, a_k$ cross the edges $b_1, b_2, \ldots, b_\ell$ to form the inflated hairpin.

Proof: Since we defined A to contain edges with smaller labels than B, the result that the right endpoints are in the order $(a_k, a_k - 1, \ldots, a_1, b_\ell, b_\ell - 1, \ldots, b_1)$ follows immediately. Now consider rearranging the right endpoints such that they appear in the order $(b_\ell, b_\ell - 1, \ldots, b_1, a_k, a_k - 1, \ldots, a_1)$. It is easy to see that every pair of edges in $A \cup B$ is now nested and that no additional crossings can be created by this rearrangement. Therefore the result is a noncrossing matching. Notice that, since noncrossing matchings are in direct bijection with LR-sequences, the resulting matching is in fact $nc(M)$. $\square$

For an example of this conversion of an L & P matching with an inflated hairpin into a noncrossing matching, see Figure 4. The results of Lemma 1 will be useful as we define our bijection from $\mathcal{LP}_n$ to $\mathcal{NCN}_n$.

Definition 2. *Define* $\phi : \mathcal{LP}_n \to \mathcal{NCN}_n$ *where,*

$$\phi(M) = \begin{cases} (M, 0, 0), & \text{if } M \text{ noncrossing,} \\ (nc(M), \max(A), \max(B)), & \text{otherwise.} \end{cases}$$

where (A, B) *is the inflated hairpin of* M.

The results of Lemma 1 imply that $\max(A), \max(B)$ are a nested pair of edges in $nc(M)$, and therefore ϕ is well-defined. All that remains is to show that ϕ is in fact a bijection.

Theorem 1. *The mapping* $\phi : \mathcal{LP}_n \to \mathcal{NCN}_n$ *is a bijection.*

Proof: We will construct the inverse of ϕ. First note that $\phi^{-1}(M, 0, 0) = M$.

Now consider some noncrossing matching M with nested pair of edges (a, b). Let $A = \{a_i \in E(M) \mid (a_i, a) \text{ are nested and } a_i \leq a \text{ such that } a_1 < a_2 < \cdots < a_k = a\}$ and $B = \{b_i \in E(M) \mid (b_i, b) \text{ are nested and } a < b_i \leq b \text{ such that } b_1 < b_2 < \cdots < b_\ell = b\}$. Then, since edges are labeled by left endpoints and M is noncrossing, the right endpoints of $A \cup B$ appear in the order $(b_\ell, b_\ell - 1, \ldots, b_1, a_k, a_k - 1, \ldots, a_1)$.

Let M' be the resulting matching when the right endpoints of $A \cup B$ are reordered to appear as $(a_k, a_k - 1, \ldots, a_1, b_\ell, b_\ell - 1, \ldots, b_1)$. It is straightforward to show that M' is an L & P matching with inflated hairpin (A, B) where $(\max(A), \max(B)) = (a, b)$. It follows that $\phi^{-1}(M, a, b) = M'$. Therefore ϕ is a bijection, as desired. $\square$
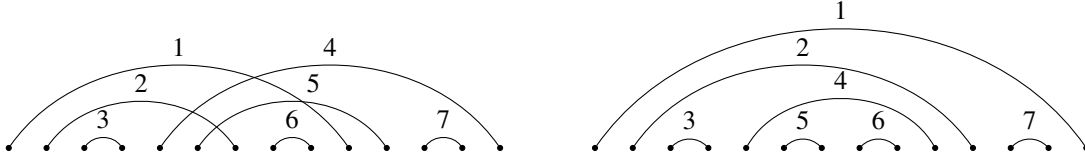
Fig. 4: On the left is an L & P matching with labeled edges. In this matching, the inflated hairpin contains edges 1, 2 crossing edges 4, 5. On the right is the corresponding noncrossing matching obtained by swapping the right endpoints of edges in the inflated hairpin.

2 Nesting-Similarity Classes and Noncrossing Matchings

Recall that $M \sim_{ne} N$ if and only if $ne(M) = ne(N)$ and M, N have the same LR-sequence. Klazar showed that the nesting-similarity classes are equinumerous with tunnel pairs in Dyck Paths through the use of transpositions that swap the endpoints of nestings with minimal width. Using this map iteratively, for any LR-sequence, if M is the corresponding noncrossing matching, then for every i where $0 \leq i \leq ne(M)$, Klazar proved that there exists some matching with the same LR-sequence and i pairs of nested edges. Notice that the noncrossing matchings contain the maximum possible number of nestings for a particular LR-sequence, so Klazar's result encompasses all the nesting-similarity classes.

We define a bijection between nesting-similarity classes and the set $\mathcal{NCN}_n$ which explicitly defines a representative for each equivalence class. Let $nep(M)$ denote the list of nested pairs of edges in M ordered lexicographically by prioritizing the second element. For example,

$$nep(M) = \{(1,2), (1,3), (2,3), (1,4), (2,4), (1,5), (2,5), (4,5), (1,6), (2,6), (4,6), (1,7)\}$$

for the noncrossing matching of Figure 4. Let $rperm(M)$ be the order in which the right endpoints of edges appear. For example, $rperm(M) = 3564271$. We define $rperm$ for all matchings, not necessarily noncrossing. However, when we have noncrossing matchings, rperm is useful for identifying nestings.

Lemma 2. *In a noncrossing matching M, edges a, b are nested, with $a < b$, if and only if b appears before a in $rperm(M)$.*

This follows quickly from the method of labeling edges and the definition of nestings.

Consider a noncrossing matching M with k pairs of nested edges. Given some i with $0 \leq i \leq k$, we define a process by which we rearrange the vertices in a noncrossing matching to obtain a matching with the same LR-sequence and i pairs of nested edges. These matchings will be the representatives of the nesting-similarity classes.

In the definition below, given edges a, b where $1 \leq a, b \leq n$ ($a \neq b$), let $(a, b).M$ denote the matching that results by swapping the left endpoints of edges a and b.

Definition 3. *Let M be a noncrossing matching with k pairs of nested edges, and $nep(M) = \{(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k)\}$. Define the sequence of matchings $M_0, M_1, M_2, \ldots, M_k$ by $M_0 = M$ and $M_i = (a_i, b_i).M_{i-1}$ for all $i \in [k]$. Additionally, let $lperm(M_i)$ denote the order in which the left endpoints of matching M_i appear.*

For an example of this definition, see Figure 5. It is clear that each M_i has the same LR-sequence as M. Our goal is to additionally show that $ne(M_i) = k - i$. This would imply that the matchings in

$\{M_0, M_1, \ldots, M_k \mid M_0 \text{ noncrossing}\}$ form a set of representatives of the $k+1$ distinct nesting-similarity classes for matchings with the same LR-sequence as M.

$$M_0 = M$$
$$lperm(M_0) = 1234$$
$$ne(M_0) = 4$$

$$M_1 = (1,2).M_0$$
$$lperm(M_1) = 2134$$
$$ne(M_1) = 3$$

$$M_2 = (1,3).M_1$$
$$lperm(M_2) = 2314$$
$$ne(M_2) = 2$$

$$M_3 = (1,4).M_2$$
$$lperm(M_3) = 2341$$
$$ne(M_3) = 1$$
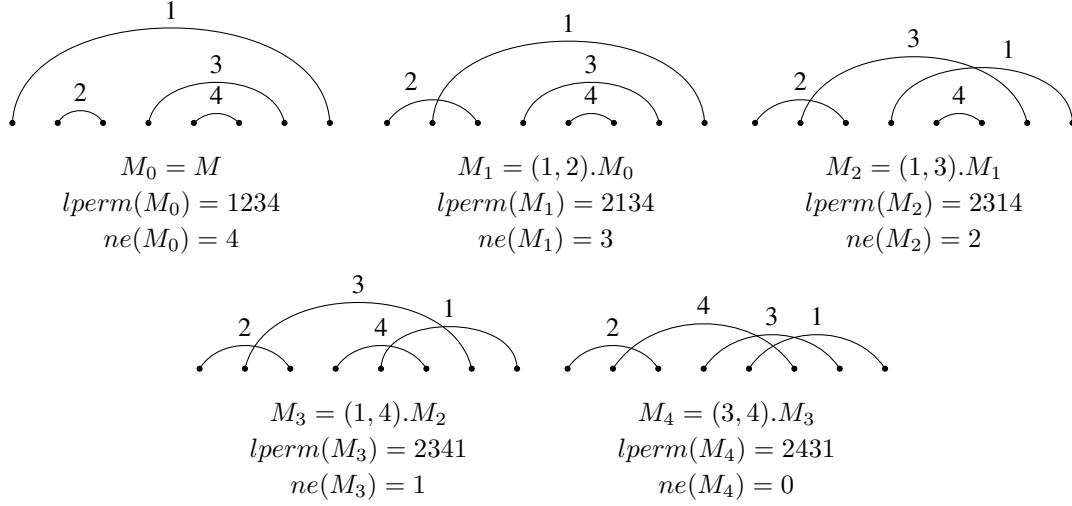
$$M_4 = (3,4).M_3$$
$$lperm(M_4) = 2431$$
$$ne(M_4) = 0$$

Fig. 5: An example of the matching obtained by swapping left endpoints in a noncrossing matching, as in Definition 3. Note the $nep(M) = \{(1,2),(1,3),(1,4),(3,4)\}$, which defines the order in which left endpoints are swapped.

Lemma 3. *Let M be a noncrossing matching, let $M_0, M_1, \ldots, M_k$ be the sequence of matchings as defined in Definition 3, and let $nep(M) = \{(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k)\}$. Then for every $i \in [k]$, a_{i+1}, b_{i+1} appear in order and are adjacent in $lperm(M_i)$.*

Proof: We use proof by induction. The result clearly holds for M_0. Assume that a_{j+1}, b_{j+1} are adjacent and in order in $lperm(M_j)$ for $0 \leq j < i$. First notice that this assumption immediately implies that a_{i+1}, b_{i+1} appear in order in $lperm(M_i)$.

To obtain a contradiction, assume that a_{i+1}, b_{i+1} are not adjacent in $lperm(M_i)$. So, there exists some c such that c appears between a_{i+1}, b_{i+1}. Notice that nestings in M_0 that are lexicographically smaller than (a_{i+1}, b_{i+1}) have had their left endpoints swapped to obtain M_i. This fact will allow us to obtain a contradiction. We will consider three cases based on the size of the label c.

First assume that $a_{i+1} < b_{i+1} < c$. This implies that c and b_{i+1} appear out of order in $lperm(M_i)$, and must have been swapped. However, (b_{i+1}, c) is lexicographically larger than (a_{i+1}, b_{i+1}), which gives a contradiction.

Next assume that $c < a_{i+1} < b_{i+1}$. In this case, c and a_{i+1} appear out of order in $lperm(M_i)$. The inductive assumption implies that at some step, (c, a_{i+1}) must have been swapped and were a nesting in M_0. However, this would also imply that (c, b_{i+1}) is a nesting in M_0 that is lexicographically smaller than (a_{i+1}, b_{i+1}); applying this swap would reverse the order of c, b_{i+1} as well. So, c could not appear between a_{i+1} and b_{i+1} and we have a contradiction.

Finally assume that $a_{i+1} < c < b_{i+1}$. We need to consider two further subcases dependent on $rperm(M_0)$. If c appears before b_{i+1} in $rperm(M_0)$, then c and b_{i+1} are not nested edges by Lemma 2. However, a_{i+1} nests both c and b_{i+1} in this case. It follows that (a_{i+1}, c) is a lexicographically smaller nesting in $nep(M_0)$ and c, a_{i+1} must have already been swapped. As a result, they cannot appear in order in M_i, giving a contradiction.

If instead b_{i+1} appears before c in $rperm(M_0)$, the condition on noncrossing edges in M_0 and the fact that a_{i+1} and b_{i+1} are nested implies that a_{i+1} and c are also nested. Again, this is a lexicographically smaller nesting, which implies that a_{i+1} and c must appear out of order in M_i, giving a contradiction. $\quad\square$

Lemma 4. *Let M be a noncrossing matching, let $M_0, M_1, \ldots, M_k$ be the sequence of matchings as defined in Definition 3, and let $nep(M) = \{(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k)\}$. Then $nep(M_i) = \{(a_{i+1}, b_{i+1}), (a_{i+2}, b_{i+2}), \ldots, (a_k, b_k)\}$; in particular, $ne(M_i) = k - i$.*

Proof: We use proof by induction on i. It is straightforward to show that $nep(M_1) = \{(a_2, b_2), \ldots, (a_k, b_k)\}$.

Now assume that $nep(M_{i-1}) = \{(a_i, b_i), \ldots, (a_k, b_k)\}$. By definition, we know that (a_i, b_i) is no longer a nesting in $M_i = (a_i, b_i).M_{i-1}$. Additionally, since (a_i, b_i) were adjacent in $lperm(M_{i-1})$ by Lemma 3, we know that $(a_{i+1}, b_{i+1}), (a_{i+2}, b_{i+2}), \ldots, (a_k, b_k)$ are all still nestings in M_i and that no additional nestings are created by swapping (a_i, b_i). Therefore $nep(M_i) = \{(a_{i+1}, b_{i+1}), (a_{i+2}, b_{i+2}), \ldots, (a_k, b_k)\}$. $\quad\square$

It follows that the matchings we generate by the swaps in Definition 3 form a set of representatives of the $k + 1$ distinct nesting-similarity classes for matchings with the same LR-sequence as M. Let $\mathcal{NS}_n$ be the set of representatives of the nesting-similarity classes; that is $\mathcal{NS}_n = \{N \in \mathcal{M}(n) \mid N = M_i$ for some noncrossing matching M with at least i nestings$\}$. We can now define the other half of our bijection.

Definition 4. *Define $\tau : \mathcal{NCN}_n \to \mathcal{NS}_n$ where, if M is a noncrossing matching with $nep(M) = \{(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k)\}$, then $\tau(M, a_i, b_i) = M_i$ and $\tau(M, 0, 0) = M$.*

So, τ will take the noncrossing matching M, which has the maximum number of nestings of any matching with the same LR-sequence, and perform a sequence of left vertex swaps. Each swap converts precisely one nesting pair of edges into a crossing pair of edges; these swaps continue until the associated pair of edges is no longer nested. For an example of the mapping τ, see Figure 6.

Theorem 2. *The mapping $\tau : \mathcal{NCN}_n \to \mathcal{NS}_n$ is a bijection. Additionally, if $M \in \mathcal{M}(n)$ is noncrossing, $\tau(M, 0, 0) = M$.*

Proof: Given some $N \in \mathcal{NS}_n$, consider $M = nc(N)$. If $N = M$ (implying that N is noncrossing), then $\tau^{-1}(N) = (N, 0, 0)$. Otherwise, by the definition of $\mathcal{NS}_n$, there exists some i such that $N = M_i$. If $nep(M) = \{(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k)\}$, set $\tau^{-1}(N) = (M, a_i, b_i)$. $\quad\square$

Now, combining Theorems 1 and 2, we immediately obtain our desired result.

Theorem 3. *The map $\sigma = \tau \circ \phi : \mathcal{LP}_n \to \mathcal{NS}_n$ is a bijection between L & P matchings and nesting-similarity classes.*
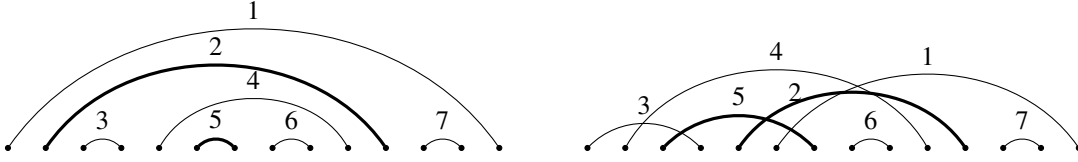
Fig. 6: An example of the bijection τ, mapping from the noncrossing matching on the left, with chosen nesting pair $(2, 5)$ to the representative of the nesting-similarity class with 5 nestings and LR-sequence LLLRLLRLRRRLRR on the right.
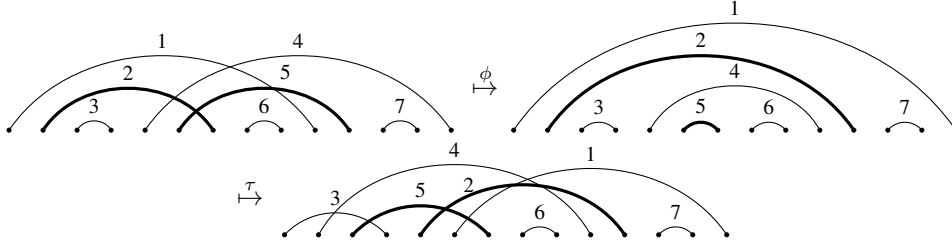

Fig. 7: Above is the result of first applying ϕ to an L & P matching to obtain a noncrossing matching with an indicated nesting pair. Then, we see the result of applying τ to the noncrossing matching with indicated nesting pair to obtain a representative of a nesting equivalence class. Composed, this is the mapping σ.

An example of this composition can be seen in Figure 7.

The images under the map σ also form a set of representatives for the nesting similarity classes $\mathcal{NS}_n$. Although Klazar proved that the classes exist, representatives of those classes were not explicitly provided, and now we have done so.

Corollary 1. *Since the bijection σ has an intermediate step at the noncrossing matching associated to a matching M, σ has the following properties:*

- *If M is a noncrossing matching, then $\sigma(M) = M$.*

- *The LR-sequence of M is the LR-sequence of $\sigma(M)$.*

Other properties that are sometimes preserved in bijections between matchings, such as number of nestings or number of crossings, are not preserved by σ. However these statistics all fail to be equidistributed between the two sets LP_n and NS_n, so no other bijection exists that preserves them.

In the larger context, we note that none of the other Largest Hairpin Family matchings (LHF, D&P, R&G, C&C) have closed forms for their enumeration sequences. However given that our map σ relates an L & P matching to a noncrossing matching and a nesting edge pair, and the fact that all of these families are constructed inductively, we believe it may be possible to find a similar mapping involving a noncrossing matching and some other matching property. The authors hope to explore this possibility in future work.

The authors would like to thank the organizers of the Dagstuhl Seminar 16071: Pattern Avoidance and Genome Sorting, which is where the problem was first presented, and where the authors had generous

References

[1] Jonathan Bloom and Sergi Elizalde. Pattern avoidance in matchings and partitions. *Electron. J. Combin.*, 20(2):#P5, 2013.

[2] Anne Condon, Beth Davy, Baharak Rastegari, Shelly Zhao, and Finbarr Tarrant. Classifying RNA pseudoknotted structures. *Theor. Comput. Sci.*, 320(1):35–50, 2004.

[3] Aziza F. Jefferson. *The substitution decomposition of matchings and RNA secondary structures*. PhD thesis, University of Florida, 2015.

[4] Martin Klazar. On identities concerning the numbers of crossings and nestings of two edges in matchings. *SIAM J. Discrete Math.*, 20(4):960–976, 2006.