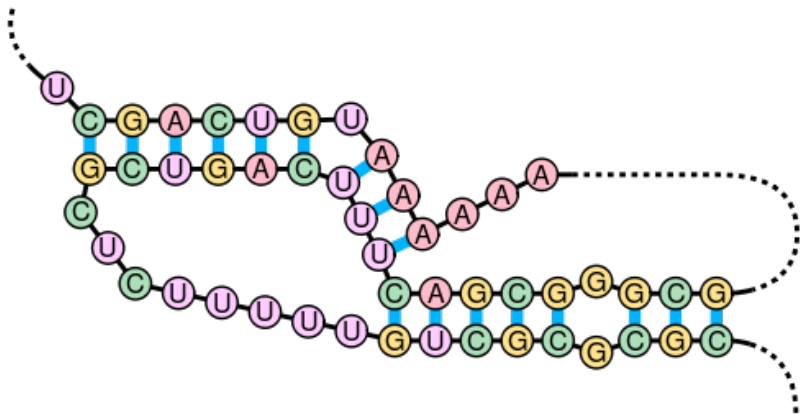


The substitution decomposition of matchings
and
RNA secondary structures

Aziza Jefferson and Vince Vatter
University of Florida

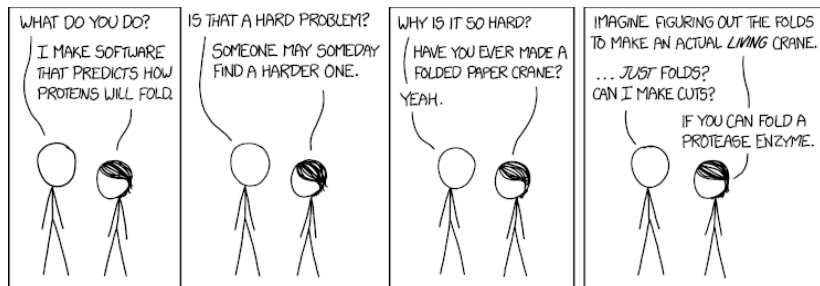
Permutation Patterns 2018
July 13, 2018

PREDICTING HOW RNA FOLDS

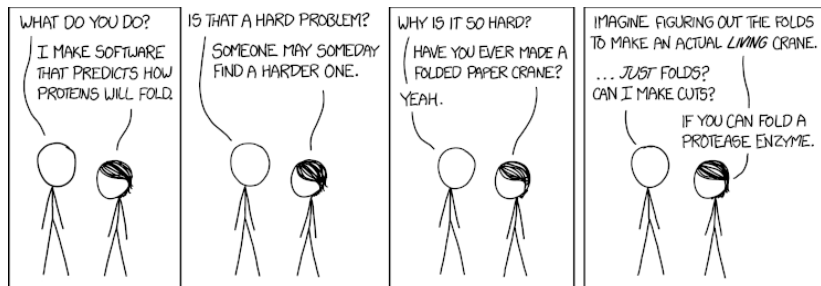


Problem: Given the primary structure, predict the secondary structure.

PREDICTING HOW RNA FOLDS



PREDICTING HOW RNA FOLDS



How about algorithms that only predict certain secondary structures?

There are *oodles* of them.

ORTHODOX STRUCTURES

For a long time biologists thought the edges of the corresponding matchings could not cross. Secondary structures without crossings are called *orthodox structures*.

We call those *non-crossing matchings*.

Counted by the Catalan numbers.

Crossings are called *pseudoknots*.

THE D&P FAMILY

Matchings which can be reduced to empty by consecutively deleting edges according to three rules. The edge (i, j) can be deleted if it is

- ▶ a directly adjacent edge, i.e., $j = i + 1$,
- ▶ a directly nested edge, i.e., there exists an edge (i', j') such that $i' = i + 1$ and $j' = j - 1$, or
- ▶ a "hairpin", i.e., there exists an edge (i', j') where either

$$j' = j + 1 = i' + 2 = i + 3 \text{ or } j = j' + 1 = i + 2 = i' + 3.$$

Dirks and Pierce. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J Comput. Chem.* 24 (2003), 1664–1677.

THE D&P FAMILY

Counted by Saule et al. using the context-free language

$$S \rightarrow dS\bar{d}S \mid P,$$

$$P \rightarrow pSX\bar{p}S \mid \epsilon,$$

$$X \rightarrow xSX\bar{x}S \mid ySY\bar{y}S,$$

$$Y \rightarrow ySY\bar{y}S \mid \epsilon.$$

Saule, Régnier, Steyaert, and Denise. Counting RNA pseudoknotted structures. *J. Comput. Biol.* 18, 10 (2011), 1339–1351.

There must be a better way to...

- ▶ describe families and
- ▶ count them.

There must be a better way to...

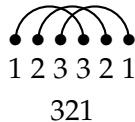
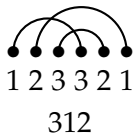
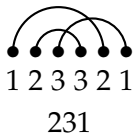
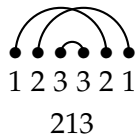
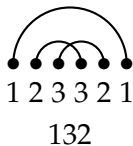
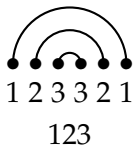
- ▶ describe families and
- ▶ count them.

There is!

PP IS A SPECIAL CASE OF MP

PP IS A SPECIAL CASE OF MP

(LEARN MORE AT MP2019 IN ZÜRICH)



We'll call these *permutational matchings* (following Jelínek for the term but Bloom and Elizalde for the "backwards" convention). The permutational matchings are precisely those matchings that avoid $\wedge\wedge$.

AVOIDING A SHORT PATTERN

- ▶ $\text{Av}(\mathfrak{M})$ — Catalan enumeration, very simple structure, poset is isomorphic to $\text{Av}(231)$.

AVOIDING A SHORT PATTERN

- ▶ $\text{Av}(\mathfrak{M})$ — Catalan enumeration, very simple structure, poset is isomorphic to $\text{Av}(231)$.
- ▶ $\text{Av}(\hat{\wedge})$ — Catalan enumeration, more complicated structure, poset is *not* isomorphic to $\text{Av}(321)$, but behaves quite a bit like it (especially w.r.t. all properties of $\text{Av}(321)$ mentioned in Brignall's talk) — see Albert and V (pre-preprint).

AVOIDING A SHORT PATTERN

- ▶ $\text{Av}(\mathfrak{M})$ — Catalan enumeration, very simple structure, poset is isomorphic to $\text{Av}(231)$.
- ▶ $\text{Av}(\hat{\wedge})$ — Catalan enumeration, more complicated structure, poset is *not* isomorphic to $\text{Av}(321)$, but behaves quite a bit like it (especially w.r.t. all properties of $\text{Av}(321)$ mentioned in Brignall's talk) — see Albert and V (pre-preprint).
- ▶ $\text{Av}(\wedge\wedge)$ — $n!$ enumeration, all of permutation patterns.

AVOIDING A SHORT PATTERN

- ▶ $\text{Av}(\mathfrak{M})$ — Catalan enumeration, very simple structure, poset is isomorphic to $\text{Av}(231)$.
- ▶ $\text{Av}(\curvearrowright)$ — Catalan enumeration, more complicated structure, poset is *not* isomorphic to $\text{Av}(321)$, but behaves quite a bit like it (especially w.r.t. all properties of $\text{Av}(321)$ mentioned in Brignall's talk) — see Albert and V (pre-preprint).
- ▶ $\text{Av}(\wedge\wedge)$ — $n!$ enumeration, all of permutation patterns.

Note: $\text{Av}(\mathfrak{M})$ is isomorphic as a poset to $\text{Av}(\wedge\wedge, \curvearrowright)$ — the smallest-yet example of "unbalanced Wilf-equivalence" (Burstein and Pantone).

THE SUBSTITUTION DECOMPOSITION

First famous use by Gallai in 1967. (Though the idea dates back to at least Fraïssé in 1953.)

Gallai. Transitiv orientierbare Graphen. *Acta Math. Acad. Sci. Hungar* 18 (1967), 25–66.

THE SUBSTITUTION DECOMPOSITION

First famous use by Gallai in 1967. (Though the idea dates back to at least Fraïssé in 1953.)

Gallai. Transitiv orientierbare Graphen. *Acta Math. Acad. Sci. Hungar* 18 (1967), 25–66.

Also called modular decomposition, disjunctive decomposition, and X-join.

THE SUBSTITUTION DECOMPOSITION

First famous use by Gallai in 1967. (Though the idea dates back to at least Fraïssé in 1953.)

Gallai. Transitiv orientierbare Graphen. *Acta Math. Acad. Sci. Hungar* 18 (1967), 25–66.

Also called modular decomposition, disjunctive decomposition, and X-join.

Useful in a number of algorithmic contexts (though not yet explicitly in RNA secondary structure prediction).

Möhring. Algorithmic aspects of the substitution decomposition in optimization over relations, sets systems and Boolean functions. *Ann. Oper. Res.* 4, 1-4 (1985), 195–225.

THE SUBSTITUTION DECOMPOSITION

First famous use by Gallai in 1967. (Though the idea dates back to at least Fraïssé in 1953.)

Gallai. Transitiv orientierbare Graphen. *Acta Math. Acad. Sci. Hungar* 18 (1967), 25–66.

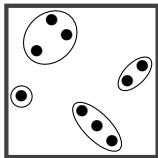
Also called modular decomposition, disjunctive decomposition, and X-join.

Useful in a number of algorithmic contexts (though not yet explicitly in RNA secondary structure prediction).

Möhring. Algorithmic aspects of the substitution decomposition in optimization over relations, sets systems and Boolean functions. *Ann. Oper. Res.* 4, 1-4 (1985), 195–225.

And also in the enumeration of permutation classes...

INTERVALS AND SIMPLE PERMUTATIONS

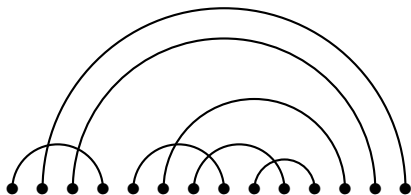


Ovals enclose the *intervals* of this permutation. A permutation is *simple* if all of its intervals are trivial (single entries or the whole permutation).

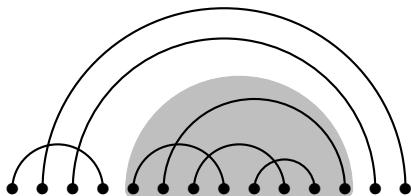
Albert and Atkinson (2005). *If a permutation class has only finitely many simple permutations then it...*

- ▶ *is defined by finitely many minimal forbidden permutations (finite basis),*
- ▶ *does not contain an infinite antichain, and*
- ▶ *has an algebraic generating function.*

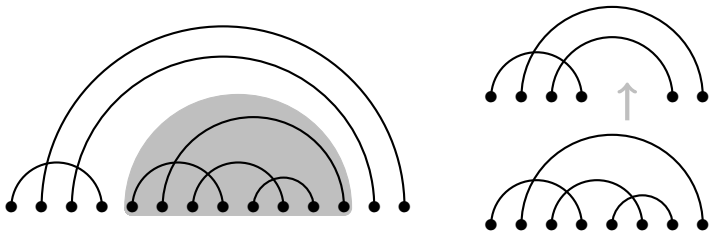
VERTEX MODULES (ANALOGUES OF INTERVALS)



VERTEX MODULES (ANALOGUES OF INTERVALS)

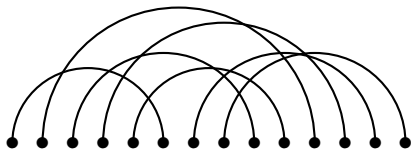


VERTEX MODULES (ANALOGUES OF INTERVALS)

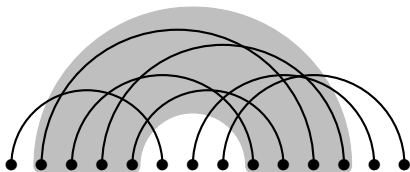


Matchings without vertex modules are *weakly indecomposable*.

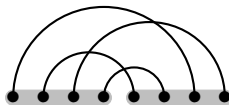
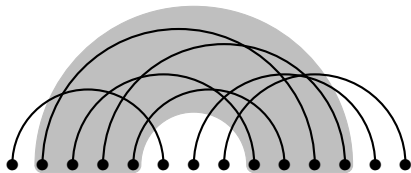
EDGE MODULES (ALSO ANALOGUES OF INTERVALS?)



EDGE MODULES (ALSO ANALOGUES OF INTERVALS?)



EDGE MODULES (ALSO ANALOGUES OF INTERVALS?)



- ▶ Note: we can only inflate an edge by a permutational matching.
- ▶ Matchings without vertex or edge modules are *strongly indecomposable*.

COUNTING INDECOMPOSABLE MATCHINGS

A Variety of People (multiple times) have "counted" the simple permutations. Asymptotically, $1/e^2$ of all permutations are simple.

COUNTING INDECOMPOSABLE MATCHINGS

A Variety of People (multiple times) have "counted" the simple permutations. Asymptotically, $1/e^2$ of all permutations are simple.

Nijenhuis and Wilf (1979) counted the weakly indecomposable matchings ("connected linked diagrams" to them):

$$w_n = (n - 1) \sum_{i=1}^{n-1} w_i w_{n-i}.$$

COUNTING INDECOMPOSABLE MATCHINGS

A Variety of People (multiple times) have "counted" the simple permutations. Asymptotically, $1/e^2$ of all permutations are simple.

Nijenhuis and Wilf (1979) counted the weakly indecomposable matchings ("connected linked diagrams" to them):

$$w_n = (n - 1) \sum_{i=1}^{n-1} w_i w_{n-i}.$$

Jefferson and V. Asymptotically, $1/e$ of all matchings are weakly indecomposable.

COUNTING INDECOMPOSABLE MATCHINGS

A Variety of People (multiple times) have "counted" the simple permutations. Asymptotically, $1/e^2$ of all permutations are simple.

Nijenhuis and Wilf (1979) counted the weakly indecomposable matchings ("connected linked diagrams" to them):

$$w_n = (n - 1) \sum_{i=1}^{n-1} w_i w_{n-i}.$$

Jefferson and V. Asymptotically, $1/e$ of all matchings are weakly indecomposable.

Jefferson and V. Asymptotically, $1/e^2$ of all matchings are strongly indecomposable.

THE SAME PROPORTION...

But there are a lot more matchings than permutations.

THE SAME PROPORTION...

But there are a lot more matchings than permutations.

Simple permutations of length 5:

24153, 25314, 31524, 35142, 41352.

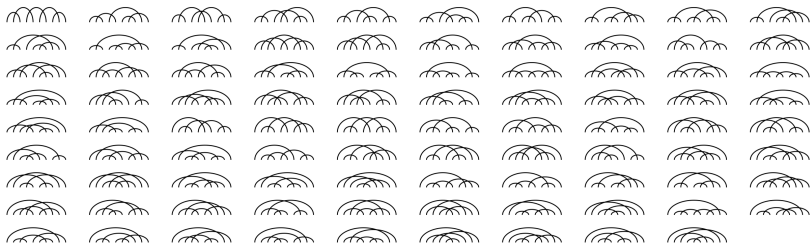
THE SAME PROPORTION...

But there are a lot more matchings than permutations.

Simple permutations of length 5:

24153, 25314, 31524, 35142, 41352.

Strongly indecomposable matchings on 5 edges:



BUILDING/ENUMERATING A FAMILY

1. Identify the strongly indecomposable members of the family.
2. *Inflate* their edges to form the weakly indecomposable members.
3. Perform *insertions* to form all members of the family.

BUILDING / ENUMERATING A FAMILY

1. Identify the strongly indecomposable members of the family.
2. *Inflate* their edges to form the weakly indecomposable members.
3. Perform *insertions* to form all members of the family.

Jefferson and V. *If a downset of matchings has only finitely many strongly indecomposable members then it...*

- ▶ *is defined by finitely many minimal forbidden matchings (finite basis),*
- ▶ *does not contain an infinite antichain, and*
- ▶ *has an algebraic generating function.*

(So the Albert–Atkinson Theorem holds in MP as well as PP.)

THE D&P FAMILY

Matchings which can be reduced to empty by consecutively deleting edges according to three rules. The edge (i, j) can be deleted if it is

- ▶ a directly adjacent edge, i.e., $j = i + 1$,
- ▶ a directly nested edge, i.e., there exists an edge (i', j') such that $i' = i + 1$ and $j' = j - 1$, or
- ▶ a "hairpin", i.e., there exists an edge (i', j') where either

$$j' = j + 1 = i' + 2 = i + 3 \text{ or } j = j' + 1 = i + 2 = i' + 3.$$

THE D&P FAMILY

Matchings which can be reduced to empty by consecutively deleting edges according to three rules. The edge (i, j) can be deleted if it is

- ▶ a directly adjacent edge, i.e., $j = i + 1$,
- ▶ a directly nested edge, i.e., there exists an edge (i', j') such that $i' = i + 1$ and $j' = j - 1$, or
- ▶ a "hairpin", i.e., there exists an edge (i', j') where either

$$j' = j + 1 = i' + 2 = i + 3 \text{ or } j = j' + 1 = i + 2 = i' + 3.$$

Strongly indecomposable matchings: \wedge and \cap .

Allowed edge inflations: Only \cap , $\cap\cap$, $\cap\cap\cap$, \dots

Allowed insertions: All.

THE D&P FAMILY

Strongly indecomposable matchings: \wedge and \mathfrak{M} "hairpin".

Allowed edge inflations: Only \frown , \smile , $\smile\smile$, $\smile\smile\smile$, ... "ladders".

Allowed insertions: All.

Generating function f satisfies

$$x^3f^6 - x^2f^5 + 2xf^3 - xf^2 - f + 1 = 0.$$

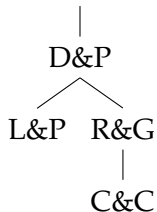
Sequence:

1, 3, 13, 65, 351, 1994, 11747, 71117, 439765, 2765775, ...

Given the description above, this is all completely routine.

HAIRPIN-ONLY FAMILIES

All inflations & insertions allowed



THE L&P FAMILY

Strongly indecomposable matchings: \wedge and \pitchfork "hairpin".

Allowed edge inflations: Only \frown , \smile , $\smile\smile$, $\smile\smile\smile$, ... "ladders".

Allowed insertions: *Only insert non-crossing matchings.*

Generating function f satisfies

$$(16x^3 - 8x^2 + x)f^2 + (-28x^2 + 15x - 2)f + (25x^2 - 14x + 2) = 0.$$

Sequence:

1, 3, 12, 51, 218, 926, 3902, 16323, 67866, 280746, ...

The same sequence arose in work of Klazar (2006), and the enumerative coincidence has been explained bijectively by Martinez and Riehl (2017).

THE L&P FAMILY

Strongly indecomposable matchings: \wedge and \pitchfork "hairpin".

Allowed edge inflations: Only \smile , $\smile\smile$, $\smile\smile\smile$, ... "ladders".

Allowed insertions: *Only insert non-crossing matchings.*

Generating function f satisfies

$$(16x^3 - 8x^2 + x)f^2 + (-28x^2 + 15x - 2)f + (25x^2 - 14x + 2) = 0.$$

Sequence:

$$1, 3, 12, 51, 218, 926, 3902, 16323, 67866, 280746, \dots$$

The same sequence arose in work of Klazar (2006), and the enumerative coincidence has been explained bijectively by Martinez and Riehl (2017).

Thank you.