# Cherry reduction sequences and linear extensions

**Joan Carles Pons**

Department of Mathematics and Computer Science
University of the Balearic Islands

Universitat de les Illes Balears

From Phylogenetics to Phylogenomics: Mathematical and Statistical Challenges in the Era of Big Data, 22/10/2024

# Table of contents

# Basic definitions and terminology

# Basics on networks

- Rooted binary phylogenetic networks
- Cherries and reticulated cherries. Terminal nodes. Cherry reduction $CR(N, u)$
- Cherry reduction sequence of $N$ is $(N_0, \ldots, N_s)$ where $N_{i+1} = CR(N_i, u_i)$
- Orchard networks (and tree-child networks)
- $CR - Seq(N)$ is the set of cherry reduction sequences of $N$
- $T(N)$ is the set of tree nodes of $N$
- $\mathring{N}$ is the result of removing pendant arcs on $N$

# Basics on posets

- Order $\preceq$ and poset $(X, \preceq)$
- Order-dual $\preceq^{op}$ and order-dual poset $(X, \preceq^{op})$
- Linear extension of $(X, \preceq)$ is a bijection $\pi : \{1, \ldots, |X|\} \to X$ s.t. $x \preceq y \to \pi^{-1}(x) \leq \pi^{-1}(y)$
- $LinExt(X)$ is the set of linear extensions of (poset) $X$
- Cover graph of a poset $(X, \preceq)$ is the graph $(X, \{xy \mid \text{ if } x \prec y \text{ and } \nexists z : x \prec z \prec y\})$

# Posets on networks

- Reachability relation on $V(N)$: $u \rightsquigarrow v$ when $u$ is an ancestor of $v$ in $N$
- Poset $(V(N), \rightsquigarrow)$ or its restriction $(T(\mathring{N}), \rightsquigarrow)$. What are the maximals?
- Linear extensions on $(T(\mathring{N}), \rightsquigarrow)$
- Cover graph of $(T(\mathring{N}), \rightsquigarrow)$

# The question

# What's the question?

From Erdös, P. L., Semple, C., and Steel, M. (2019) in Mathematical Biosciences
"A class of phylogenetic networks reconstructable from ancestral profiles".

Can we count the number of cherry reduction sequences on ~~orchard~~ networks?

Can we count the number of cherry reduction sequences on tree-child networks?

In other words, if $N$ is a tree-child network, what is $|CR-Seq(N)|$?

# Why this question?

1. CR-Seq represents different ways to simplify evolutionary information
2. Yule process on tree generation. The probability $P_Y(T)(n-1)! = |CR - Seq(T)^*|$
3. Cherry reductions are used for different purposes: identification, generation, etc
4. …

Impossible to resist for one interested in phylogenetic combinatorics

# The lemmas

# Lemma 1

Let $N$ be an orchard network, $u \in T(\mathring{N})$ a terminal node, and $N' = CR(N, u)$.

Then $T(\mathring{N}) = T(\mathring{N}') \sqcup \{u\}$.

# Lemma 2

Let $N$ be an orchard network.

The assignment $\sigma : CR - Seq(N) \to \mathfrak{S}_{T(\mathring{N})}$ that maps each cherry reduction sequence $(N_0, \ldots, N_s)$ of $N$ to the permutation of internal tree nodes $(u_0 \ \cdots \ u_{s-1})$ of $N$, where $N_{i+1} = CR(N_i, u_i)$, is well-defined and injective.

Moreover, $\sigma$ can be restricted to linear extensions of $(T(\mathring{N}), \rightsquigarrow^{op})$.

# Lemma 3

Let $N$ be a tree-child network. Let $u \in T(\mathring{N})$.

Then $u$ is terminal in $N$ if, and only if, $u$ is maximal in $(T(\mathring{N}), \rightsquigarrow)$.

*Note*. That is not true for (general) orchard networks.

# The main result

# Theorem

Let N be a tree-child network.

Then, there is a bijection between $CR - Seq(N)$ and $LinExt((T(\mathring{N}), \rightsquigarrow))$.

# The algorithm(s)

# Network parameters

- Tree-width
- Level

Both terms are related: For a level-$k$ network $N$, $tw(N) \leq k + 1$

# Counting linear extensions

The problem of counting linear extensions of a poset is #P-complete in general (Brightwell and Winkler, 1991).

The problem of counting linear extensions of a poset of $p$ elements parameterized by the tree-width of its underlying *cover graph* is not FPT.

There is an algorithm to count the number of linear extensions of a poset with $p$ elements with complexity $O(p^{tw+4})$ where $tw$ is the tree-width of the *cover graph* of the poset (Kangas et al., 2016).

# In our context

Our poset is $(T(\mathring{N}), \rightsquigarrow)$.

Note that $|T(\mathring{N})| = n - 1$ in tree-child networks.

The cover graph of $(T(\mathring{N}), \rightsquigarrow)$ is easy to compute and it is a minor of $N$.

For the tree-width, we have $tw(\mathscr{C}(T(\mathring{N}))) \leq tw(N) \leq k + 1$.

Then we have an algorithm that runs in $O(n^{k+5})$.

# Tree-width and level in tree-child networks

The restriction to tree-child networks simplify the problem?

No. We can build a tree-child network whose level is quadratic with respect to its tree-width, and remember that the level is unbounded.

# Conclusions

# Future directions

- Find an algorithm that exploits topological features of tree-child networks to count their linear extensions.

- Which kind of posets tree-child networks define? The problem becomes easier to be solved than in general posets?

- Our strategy does not work for orchard networks. The problem remains open.

# Thanks! Time for questions