Consistency of Topological Moves

Gillian Chu
CS 581 Nov 10, 2020
Overview

1. Introduction
2. Main Conclusions
3. SPR Sequence (RF Distance)
4. BSPPR consistency
5. Q & A
Introduction

Introduction

Balanced Minimum Evolution

- Pauplin’s Formula
  \[ \hat{i}(T) = \sum_{x,y \in X} 2^{1-p_{xy}} \delta_{xy}, \]

How are BNNI/BSPR different from NNI and SPR?

- Inputs: input distance matrix \( \sigma \), phylogenetic tree \( T \)
- Output: Improved Tree \( T' \) s.t. \( \hat{i}(T) - \hat{i}(T') > 0. \)

Paper’s Goal

Input: tree metric distance matrix $\sigma^*$ for tree $T^*$

Main Q: If we apply the BSPR (BNNI) algorithm starting with distance $\sigma^*$ and initial tree $T$, are we guaranteed to output $T^*$?

Input: $T$  Goal: $T^*$
Overview

1. Introduction
2. Main Conclusions
3. SPR Sequence (RF Distance)
4. BSPR consistency
5. Q&A
Main Conclusions

1. “For two distinct phylogenetic trees T and T*, there is a sequence of SPR operations that transforms T into T* and decreases the RF distance at every step.”

2. “For two distinct phylogenetic trees T and T*, there is a sequence of SPR operations that transforms T into T* and decreases the quartet distance at every step.”

3. “BSPR algorithm is consistent and has safety radius of at least $\frac{1}{3}$. “
Overview

1. Introduction
2. Main Conclusions
3. SPR Sequence (RF Distance)
4. BSPPR consistency
5. Q & A
Theorem 3.1. If $T^*$ is a fixed tree and $T$ is any other tree, then there is a sequence s.t.

$$d_{RF}(T_i, T^*) - d_{RF}(T_{i+1}, T^*) > 0,$$

and each arrow is a single SPR move.

Lemma 3.2 Suppose $T$ and $T^*$ are two trees with distinct topologies. Then there exists $T'$ s.t.

$$d_{SPR}(T, T') = 1$$

$$d_{RF}(T^*, T') < d_{RF}(T^*, T)$$

Lemma 2.1 Suppose $T$ and $T^*$ have distinct topologies. Then there exist disjoint subtrees $B$, $D$ in $T$ s.t. $B$, $D$, and $B \cup D$ are subtrees of $T^*$, but $B \cup D$ is not a subtree of $T$. 

**SPR Sequence (RF Distance)**

1. “For two distinct phylogenetic trees $T$ and $T^*$, there is a sequence of SPR operations that transforms $T$ into $T^*$ and decreases the RF distance at every step.”
Lemma 2.1 Suppose $T$ and $T^*$ have distinct topologies. Then there exist disjoint subtrees $B, D$ in $T$ s.t. $B, D,$ and $B \cup D$ are subtrees of $T^*$, but $B \cup D$ is not a subtree of $T$.

Proof: Given $T$ and $T^*$, we have

1) $x, y$ are a cherry in $T^*$ but not in $T$ 
2) $C(T^*) \subseteq C(T)$
**SPR Sequence (RF Distance)**

Lemma 3.2 Suppose $T$ and $T^*$ are two trees with distinct topologies. Then there exists $T'$ s.t.

$$d_{SPR}(T, T') = 1 \quad \text{and} \quad d_{RF}(T^*, T') < d_{RF}(T^*, T)$$

Lemma 2.1 Suppose $T$ and $T^*$ have distinct topologies. Then there exist disjoint subtrees $B, D$ in $T$ s.t. $B, D,$ and $B \cup D$ are subtrees of $T^*$, but $B \cup D$ is not a subtree of $T$.

![Diagram showing two trees $T$ and $T'$ with disjoint subtrees $B$ and $D$.](image)

$$S_{nb}(T) = S_{nb}(T')$$

Input: $T, \sigma^*$  
Goal: $T^*$
Lemma 3.2 Suppose $T$ and $T^*$ are two trees with distinct topologies. Then there exists $T'$ s.t.

\[ d_{\text{SFR}}(T,T') = 1 , \quad d_{\text{RF}}(T',T^*) < d_{\text{RF}}(T^*,T) \]

1. $S_{\text{nb}}(T) = S_{\text{nb}}(T')$, 
2. $S_b(T) \cap S(T^*) = \emptyset$, since the only splits of $T^*$ which separate $B$ and $D$ are $S_1$ and $S_2$, and 
3. $S_b(T') \cap S(T^*) \neq \emptyset$ since $S_c$ is a split of $T'$ and $T^*$. 

SPR Sequence (RF Distance)
Theorem 3.1. If $T^*$ is a fixed tree and $T$ is any other tree, then there is a sequence s.t.

\[ \text{and each arrow is a single SPR move.} \]

SPR Sequence (RF Distance)

1. “For two distinct phylogenetic trees $T$ and $T^*$, there is a sequence of SPR operations that transforms $T$ into $T^*$ and decreases the RF distance at every step.”

Theorem 3.1. If $T^*$ is a fixed tree and $T$ is any other tree, then there is a sequence s.t.

\[ d_{RF}(T_i, T^*) - d_{RF}(T_{i+1}, T^*) > 0, \]

and each arrow is a single SPR move.

Lemma 3.2 Suppose $T$ and $T^*$ are two trees with distinct topologies. Then there exists $T'$ s.t.

\[ d_{SPR}(T, T') = 1 \quad \text{and} \quad d_{RF}(T^*, T') < d_{RF}(T^*, T) \]

Lemma 2.1 Suppose $T$ and $T^*$ have distinct topologies. Then there exist disjoint subtrees $B, D$ in $T$ s.t. $B, D,$ and $B \cup D$ are subtrees of $T^*$, but $B \cup D$ is not a subtree of $T$. 

Input: $T$, $\sigma^*$  
Goal: $T^*$
Overview

1. Introduction
2. Main Conclusions
3. SPR Sequence (RF Distance)
4. BSPPR consistency
5. Q & A
Theorem 5.2 Let $T$ be a tree with distinct topology from $T^*$. Provided then

$$\bar{i}(T) - \bar{i}(T^*) > 0$$

Lemma 2.1 Suppose $T$ and $T^*$ have distinct topologies. Then there exist disjoint subtrees $B, D$ in $T$ s.t. $B, D,$ and $B \cup D$ are subtrees of $T^*$, but $B \cup D$ is not a subtree of $T$.

Lemma 5.1 Produces a definition for

$$\bar{i}(e) - \bar{i}(e') > 0 = \bar{i}(T) - \bar{i}(T')$$

BSPR Consistency

3. “BSPR algorithm is consistent and has safety radius of at least $\frac{1}{3}$.”
Theorem 5.2 Let \( T \) be a tree with distinct topology from \( T^* \). Provided then:

\[
\tilde{i}(T) - \tilde{i}(T^*) > 0
\]

Lemma 5.1 Produces a definition for

\[
\tilde{i}(e) - \tilde{i}(e') > 0 = \tilde{i}(T) - \tilde{i}(T')
\]

|\( |\sigma_{ab} - \sigma'_{ab}| < \varepsilon \) : \( \frac{1}{2} \min_{e \in E(T^*)} l(e) \forall a, b \in X \) |

\[
(\delta_{C_0b} - \delta_{C_0C_i} + \delta_{C_iB}) \geq 2l(e_x) + 2\delta^*_x - 3\varepsilon
\]
BSPR Consistency

3. "BSPR algorithm is consistent and has safety radius of at least $\frac{1}{3}$."

If $|\sigma_{ab} - \sigma_{ab}^*| < \epsilon := \frac{1}{3} \min_{e \in E(T^*)} l(e) \forall a, b \in X$ then the unique tree that achieves the minimal tree length is the correct tree $T^*$. The $\frac{1}{3}$ minimum edge length radius holds regardless of the method used to find the shortest tree (i.e. Ordinary Least Squares ME can be used).

Lemma 2.1 Suppose $T'$ and $T^*$ have distinct topologies. Then there exist disjoint subtrees $B$, $D$ in $T'$ s.t. $B$, $D$, and $B \cup D$ are subtrees of $T^*$, but $B \cup D$ is not a subtree of $T'$.

Theorem 5.2 Let $T''$ be a tree with distinct topology from $T'$. Provided $|\sigma_{ab} - \sigma_{ab}^*| < \epsilon := \frac{1}{3} \min_{e \in E(T^*)} l(e) \forall a, b \in X$ then $\hat{i}(T) - \hat{i}(T') > 0$. 

Overview

1. Introduction
2. Main Conclusions
3. SPR Sequence (RF Distance)
4. BSPPR consistency
5. Q & A
Overview

1. Introduction
2. Main Conclusions
3. SPR Sequence (RF Distance)
4. BSPPR consistency
5. Q & A
Thanks for listening!

Advantages:
- Language readable, provided examples

Disadvantages:
- Notation was inconsistent, diagrams’ captions were not always very helpful, no intuition provided, provided context

Bonus: Open Questions

1. They didn’t prove BNNI
2. Authors believe BME safety radius should be \( \frac{1}{2} \), but proof remains open