# New fixed-parameter algorithms for the minimum quartet inconsistency problem

Maw-Shang Chang[1] Chuang-Chieh Lin (Joseph)[1]
Peter Rossmanith[2]

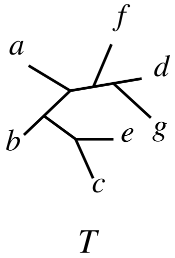
Department of Computer Science and Information Engineering,
National Chung Cheng University, Ming-Hsiung, Chiayi, Taiwan
mschang@cs.ccu.edu.tw; lincc@cs.ccu.edu.tw

Department of Computer Science, RWTH Aachen University, Germany
rossmani@informatik.rwth-aachen.de

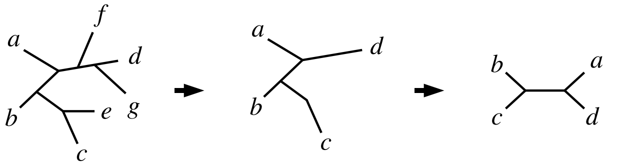
May 16, 2008

# Evolutionary trees

- $S$: a set of taxa; $|S| = n$.
- An evolutionary tree $T$ on $S$:
  - An *unrooted*, *leaf-labeled* tree
  - The leaves are bijectively labeled by the taxa in $S$
  - Each internal node has degree *three*



# Quartet topologies

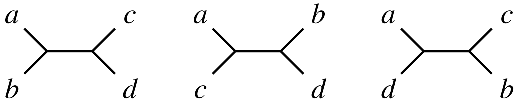



$T$

The path structure connecting $a$, $b$, $c$, $d$ in $T$

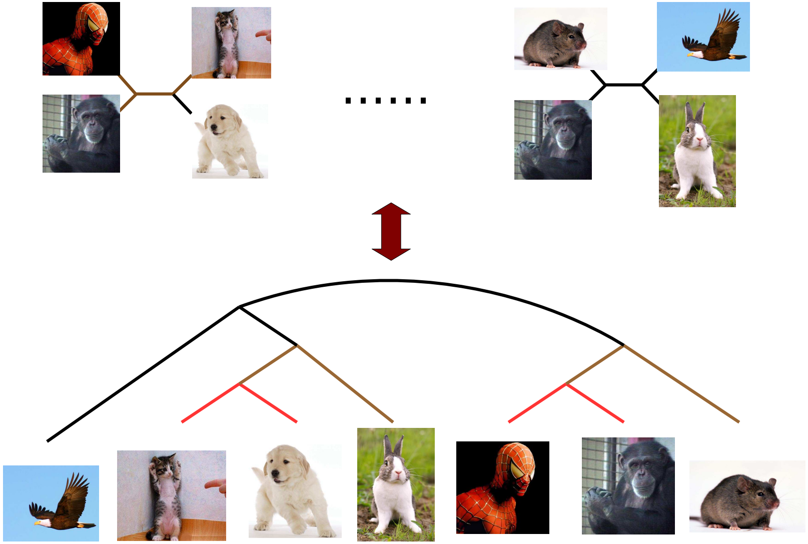
The topology of $\{a, b, c, d\}$

# Quartet topologies (contd.)




$[ab|cd]$ $[ac|bd]$ $[ad|bc]$

# Biological issue



# Tree-consistency

- $Q_T$ : the set of quartet topologies induced by $T$.
  - $|Q_T| = \binom{n}{4}$.
- $Q$ is tree-consistent (with $T$):
  - $\exists T$ s.t. $Q \subseteq Q_T$.

  ▷ tree-like if $Q = Q_T$.
- $Q$ is called complete:
  - Exactly one topology for every quartet;
  - Otherwise, incomplete.

# Tree-consistency

- $Q_T$ : the set of quartet topologies induced by $T$.
  - $|Q_T| = \binom{n}{4}$.
- $Q$ is tree-consistent (with $T$):
  - $\exists T$ s.t. $Q \subseteq Q_T$.

▷ tree-like if $Q = Q_T$.

- $Q$ is called complete:
  - Exactly one topology for every quartet;
  - Otherwise, incomplete.

# Tree-consistency

- $Q_T$ : the set of quartet topologies induced by $T$.
  - $|Q_T| = \binom{n}{4}$.
- $Q$ is tree-consistent (with $T$):
  - $\exists T$ s.t. $Q \subseteq Q_T$.

  ▷ tree-like if $Q = Q_T$.
- $Q$ is called complete:
  - Exactly one topology for every quartet;
  - Otherwise, incomplete.

# Tree-consistency

- $Q_T$ : the set of quartet topologies induced by $T$.
  - $|Q_T| = \binom{n}{4}$.
- $Q$ is tree-consistent (with $T$):
  - $\exists T$ s.t. $Q \subseteq Q_T$.

  ▷ tree-like if $Q = Q_T$.
- $Q$ is called complete:
  - Exactly one topology for every quartet;
  - Otherwise, incomplete.

# Quartet errors

- Given complete $Q$ and $Q^*$ (tree-like).
- **# quartet errors of $Q$ w.r.t. $Q^*$:**
  - $\Delta(Q, Q^*)$.
- **# quartet errors of $Q$:**
  - $\Delta^*(Q) := \min\{\Delta(Q, Q^*) : Q^* \text{ is tree-like}\}$.

# Quartet errors

- Given complete $Q$ and $Q^*$ (tree-like).
- **# quartet errors of $Q$ w.r.t. $Q^*$:**
  - $\Delta(Q, Q^*)$.
- # quartet errors of $Q$:
  - $\Delta^*(Q) := \min\{\Delta(Q, Q^*) : Q^* \text{ is tree-like}\}$.

# Quartet errors

- Given complete $Q$ and $Q^*$ (tree-like).
- **# quartet errors of $Q$ w.r.t. $Q^*$**:
  - $\Delta(Q, Q^*)$.
- **# quartet errors of $Q$**:
  - $\Delta^*(Q) := \min\{\Delta(Q, Q^*) : Q^* \text{ is tree-like}\}$.

# The problem focused in this paper:

**Given:** a complete set of quartet topologies $Q$ and an integer $k$.

- The parameterized minimum quartet inconsistency problem:

> Determine whether there exists an evolutionary tree $T$ such that $\Delta(Q, Q_T) \leq k$.

⋆ **NP**-complete [Berry *et al.* 1999].

⋆ $O(4^k n + n^4)$ [Gramm and Niedermeier 2003].

# The problem focused in this paper:

**Given:** a complete set of quartet topologies $Q$ and an integer $k$.

- The parameterized minimum quartet inconsistency problem:

> Determine whether there exists an evolutionary tree $T$ such that $\Delta(Q, Q_T) \leq k$.

- ⋆ **NP**-complete [Berry *et al.* 1999].
- ⋆ $O(4^k n + n^4)$ [Gramm and Niedermeier 2003].

# Our results

- An $O^*(3.0446^k)$ fixed-parameter algorithm.
- An $O^*(2.0162^k)$ fixed-parameter algorithm.
- An $O^*((1+\epsilon)^k)$ fixed-parameter algorithm.

# Related works (Constructing $T$ and QCP)

- Construct $T$ from a given tree-like $Q$:
  - ⋆ $O(n^4)$ [Berry and Gascuel 2000].
- The Quartet Compatibility Problem (QCP):

  > Determine whether there exists an evolutionary tree $T$ satisfying all quartet topologies in $Q$.

  - ⋆ **NP**-complete [Steel 1992].
  - ⋆ Polynomial time solvable if $Q$ is complete [Erdős *et al.* 1999].
- Consider the cases of **complete** $Q$.

# Related works (Constructing $T$ and QCP)

- Construct $T$ from a given tree-like $Q$:
  - ⋆ $O(n^4)$ [Berry and Gascuel 2000].
- The Quartet Compatibility Problem (QCP):

  > Determine whether there exists an evolutionary tree $T$ satisfying all quartet topologies in $Q$.

  - ⋆ **NP**-complete [Steel 1992].
  - ⋆ Polynomial time solvable if $Q$ is complete [Erdős *et al.* 1999].
- Consider the cases of **complete** $Q$.

# Related works (Constructing $T$ and QCP)

- Construct $T$ from a given tree-like $Q$:
  - ⋆ $O(n^4)$ [Berry and Gascuel 2000].
- The Quartet Compatibility Problem (QCP):

> Determine whether there exists an evolutionary tree $T$ satisfying all quartet topologies in $Q$.

  - ⋆ **NP**-complete [Steel 1992].
  - ⋆ Polynomial time solvable if $Q$ is complete [Erdős *et al.* 1999].
- Consider the cases of **complete** $Q$.

# Related works (MQI & MQC)

## Minimum Quartet Inconsistency Problem (MQI)

Construct an evolutionary tree $T$ s.t. $\Delta(Q, Q_T)$ is minimized.

- ⋆ **NP**-hard [Berry *et al.* 1999].
- ⋆ Approx. ratio: $O(n^2)$ [Jiang *et al.* 2000].
- ⋆ $O(3^n n^4)$ dynamic programming [Ben-Dor *et al.* 1998].
- ⋆ $O(n^4)$ if $\Delta^*(Q) < (n-3)/2$ [Berry *et al.* 1999].
- ⋆ $O(n^5 + 2^{4c} n^{12c+2})$ if $\Delta^*(Q) < cn$ for some constant $c$ [Wu *et al.* 2006].

## Maximum Quartet Consistency Problem (MQC)

Dual problem of MQI.

- ⋆ **NP**-hard [Berry *et al.* 1999].
- ⋆ PTAS [Jiang *et al.* 2001].

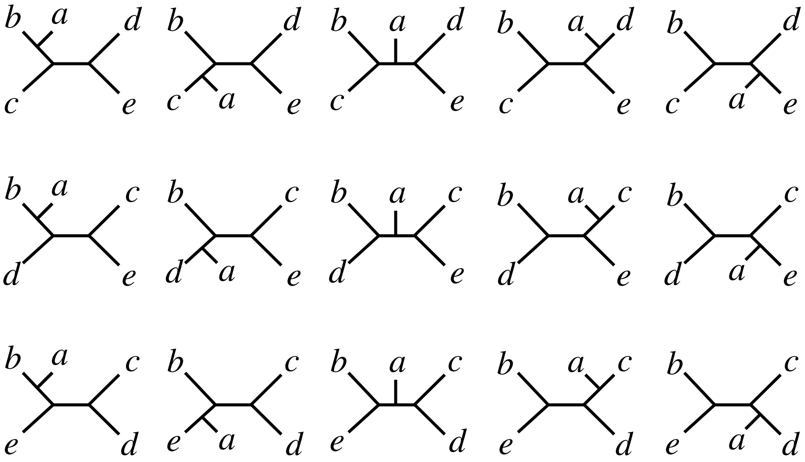
# Quintets

- A quintet is a set of five taxa in $S$.
- Quintet topologies:

# Quintets

- A quintet is a set of five taxa in $S$.
- Quintet topologies:

# Quintet topologies



# Resolved quintets

- What is a resolved quintet?

▷ $[ab|cd], [ab|ce], [ab|de], [ac|de], [bc|de] \in Q$.

# Resolved quintets

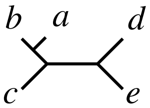
- What is a resolved quintet?

▷ $[ab|cd], [ab|ce], [ab|de], [ac|de], [bc|de] \in Q$.

# Resolved quintets

- What is a resolved quintet?

▷ $[ab|cd], [ab|ce], [ab|de], [ac|de], [bc|de] \in Q$.



# Tree consistency and conflicts

- Local conflict: a set of three quartet topologies which is not tree-consistent.

**Lemma 2.1 (Gramm and Niedermeier 2003)**

*3 quartet topologies with $> 5$ taxa $\Rightarrow$ no local conflict.*

# Tree consistency and conflicts

- Local conflict: a set of three quartet topologies which is not tree-consistent.

**Lemma 2.1 (Gramm and Niedermeier 2003)**

*3 quartet topologies with $> 5$ taxa $\Rightarrow$ no local conflict.*

# Tree consistency and conflicts (contd.)

**Theorem 2.2 (Gramm and Niedermeier 2003)**

*$Q$ is tree-like $\Leftrightarrow$ no local conflict for every set of 3 quartet topologies involving a fixed taxon $f$.*

**Theorem 2.3 (Bandelt and Dress 1986)**

*$Q$ is tree-like $\Leftrightarrow$ every quintet containing $f$ is resolved.*

# Tree consistency and conflicts (contd.)

**Theorem 2.2 (Gramm and Niedermeier 2003)**

*$Q$ is tree-like $\Leftrightarrow$ no local conflict for every set of 3 quartet topologies involving a fixed taxon $f$.*

**Theorem 2.3 (Bandelt and Dress 1986)**

*$Q$ is tree-like $\Leftrightarrow$ every quintet containing $f$ is resolved.*

# Idea of Gramm and Niedermeier's algorithm

- Bounded-depth search tree strategy.
- Eliminate a local conflict $\Rightarrow$ 4 kinds of ways.
- Each branching node has 4 branches.
- Branching vector: $(1, 1, 1, 1)$
  - Branching number: 4, hence the $O^*(4^k)$ complexity.

# Idea of Gramm and Niedermeier's algorithm

- Bounded-depth search tree strategy.
- Eliminate a local conflict $\Rightarrow$ 4 kinds of ways.
- Each branching node has 4 branches.
- Branching vector: $(1, 1, 1, 1)$
  - ▷ Branching number: 4, hence the $O^*(4^k)$ complexity.

# Idea of Gramm and Niedermeier's algorithm

- Bounded-depth search tree strategy.
- Eliminate a local conflict $\Rightarrow$ 4 kinds of ways.
- Each branching node has 4 branches.
- Branching vector: $(1, 1, 1, 1)$
  - ▷ Branching number: 4, hence the $O^*(4^k)$ complexity.

# Idea of Gramm and Niedermeier's algorithm

- Bounded-depth search tree strategy.
- Eliminate a local conflict $\Rightarrow$ 4 kinds of ways.
- Each branching node has 4 branches.
- Branching vector: $(1, 1, 1, 1)$
  - ▷ Branching number: 4, hence the $O^*(4^k)$ complexity.

# Idea of Gramm and Niedermeier's algorithm

- Bounded-depth search tree strategy.
- Eliminate a local conflict $\Rightarrow$ 4 kinds of ways.
- Each branching node has 4 branches.
- Branching vector: $(1, 1, 1, 1)$
  - ▷ Branching number: 4, hence the $O^*(4^k)$ complexity.

# Idea of our first algorithm

- Also bounded-depth search tree strategy
- Eliminate unresolved quintets.
- 15 branches for each node of the search tree

# Idea of our first algorithm

- Also bounded-depth search tree strategy
- Eliminate unresolved quintets.
- 15 branches for each node of the search tree

# Idea of our first algorithm

- Also bounded-depth search tree strategy
- Eliminate unresolved quintets.
- 15 branches for each node of the search tree

# The first algorithm (contd.)

- For the quintet $\{a, b, c, d, e\}$:
  - $[ab|cd]$, $[ac|be]$, $[ae|bd]$, $[ad|ce]$, $[bc|de] \in Q$.
- Consider the (first) quintet topology:
  - $[ab|cd]$, $[ab|ce]$, $[ab|de]$, $[ac|de]$, $[bc|de]$.

| branching vector | branching number |
|---|---|
| $(3, 3, 4, 3, 3, 3, 4, 3, 3, 4, 4, 3, 3, 4, 3)$ | $2.30042\ldots$ |
| $(2, 4, 4, 4, 5, 2, 2, 3, 3, 4, 3, 4, 3, 3, 4)$ | $2.46596\ldots$ |
| $\ldots$ | $\ldots$ |
| $(3, 5, 5, 3, 5, 2, 2, 3, 5, 5, 2, 3, 2, 3, 2)$ | $2.67102\ldots$ |
| $(1, 3, 3, 5, 5, 1, 3, 3, 3, 4, 2, 4, 4, 4, 5)$ | $3.04454\ldots$ |

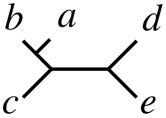
# The first algorithm (contd.)

- For the quintet $\{a, b, c, d, e\}$:
  - $[ab|cd]$, $[ac|be]$, $[ae|bd]$, $[ad|ce]$, $[bc|de] \in Q$.
- Consider the (first) quintet topology:
  - $[ab|cd]$, $[ab|ce]$, $[ab|de]$, $[ac|de]$, $[bc|de]$.

| branching vector | branching number |
|---|---|
| $(3, 3, 4, 3, 3, 3, 4, 3, 3, 4, 4, 3, 3, 4, 3)$ | $2.30042\ldots$ |
| $(2, 4, 4, 4, 5, 2, 2, 3, 3, 4, 3, 4, 3, 3, 4)$ | $2.46596\ldots$ |
| $\ldots$ | $\ldots$ |
| $(3, 5, 5, 3, 5, 2, 2, 3, 5, 5, 2, 3, 2, 3, 2)$ | $2.67102\ldots$ |
| $(1, 3, 3, 5, 5, 1, 3, 3, 3, 4, 2, 4, 4, 4, 5)$ | $3.04454\ldots$ |

# The first algorithm (contd.)

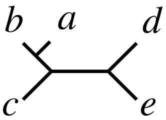
- For the quintet $\{a, b, c, d, e\}$:
  - $[ab|cd]$, $[ac|be]$, $[ae|bd]$, $[ad|ce]$, $[bc|de] \in Q$.
- Consider the (first) quintet topology:
  - $[ab|cd]$, $[ab|ce]$, $[ab|de]$, $[ac|de]$, $[bc|de]$.




| branching vector | branching number |
|---|---|
| $(3, 3, 4, 3, 3, 3, 4, 3, 3, 4, 4, 3, 3, 4, 3)$ | $2.30042\ldots$ |
| $(2, 4, 4, 4, 5, 2, 2, 3, 3, 4, 3, 4, 3, 3, 4)$ | $2.46596\ldots$ |
| $\ldots$ | $\ldots$ |
| $(3, 5, 5, 3, 5, 2, 2, 3, 5, 5, 2, 3, 2, 3, 2)$ | $2.67102\ldots$ |
| $(1, 3, 3, 5, 5, 1, 3, 3, 3, 4, 2, 4, 4, 4, 5)$ | $3.04454\ldots$ |

# The first algorithm (contd.)

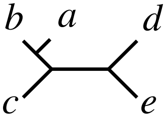
- For the quintet $\{a, b, c, d, e\}$:
  - $[ab|cd]$, $[ac|be]$, $[ae|bd]$, $[ad|ce]$, $[bc|de] \in Q$.
- Consider the (first) quintet topology:
  - $[ab|cd]$, $[ab|ce]$, $[ab|de]$, $[ac|de]$, $[bc|de]$.




| branching vector | branching number |
|---|---|
| (**3**, 3, 4, 3, 3, 3, 4, 3, 3, 4, 4, 3, 3, 4, 3) | 2.30042... |
| (2, 4, 4, 4, 5, 2, 2, 3, 3, 4, 3, 4, 3, 3, 4) | 2.46596... |
| ... | ... |
| (3, 5, 5, 3, 5, 2, 2, 3, 5, 5, 2, 3, 2, 3, 2) | 2.67102... |
| (1, 3, 3, 5, 5, 1, 3, 3, 3, 4, 2, 4, 4, 4, 5) | 3.04454... |

# The first algorithm (contd.)

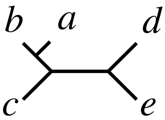
- For the quintet $\{a, b, c, d, e\}$:
  - $[ab|cd]$, $[ac|be]$, $[ae|bd]$, $[ad|ce]$, $[bc|de] \in Q$.
- Consider the (first) quintet topology:
  - $[ab|cd]$, $[ab|ce]$, $[ab|de]$, $[ac|de]$, $[bc|de]$.




| branching vector | branching number |
|---|---|
| (**3**, 3, 4, 3, 3, 3, 4, 3, 3, 4, 4, 3, 3, 4, 3) | 2.30042... |
| (2, 4, 4, 4, 5, 2, 2, 3, 3, 4, 3, 4, 3, 3, 4) | 2.46596... |
| ... | ... |
| (3, 5, 5, 3, 5, 2, 2, 3, 5, 5, 2, 3, 2, 3, 2) | 2.67102... |
| (1, 3, 3, 5, 5, 1, 3, 3, 3, 4, 2, 4, 4, 4, 5) | 3.04454... |

# The first algorithm (contd.)

- For the quintet $\{a, b, c, d, e\}$:
  - ▷ $[ab|cd]$, $[ac|be]$, $[ae|bd]$, $[ad|ce]$, $[bc|de] \in Q$.
- Consider the (first) quintet topology:
  - ▷ $[ab|cd]$, $[ab|ce]$, $[ab|de]$, $[ac|de]$, $[bc|de]$.



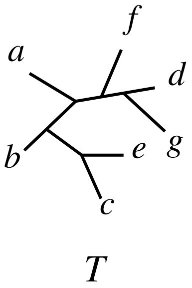

| branching vector | branching number |
|---|---|
| (**3**, 3, 4, 3, 3, 3, 4, 3, 3, 4, 4, 3, 3, 4, 3) | 2.30042... |
| (2, 4, 4, 4, 5, 2, 2, 3, 3, 4, 3, 4, 3, 3, 4) | 2.46596... |
| ... | ... |
| (3, 5, 5, 3, 5, 2, 2, 3, 5, 5, 2, 3, 2, 3, 2) | 2.67102... |
| (1, 3, 3, 5, 5, 1, 3, 3, 3, 4, 2, 4, 4, 4, 5) | 3.04454... |

# The first algorithm (contd.)

**Theorem 3.1**

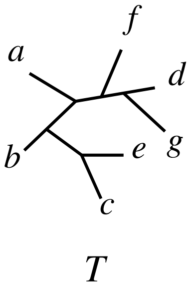
*There exists an $O(3.0446^k n + n^4)$ fixed-parameter algorithm for the parameterized MQI problem.*

# Siblings



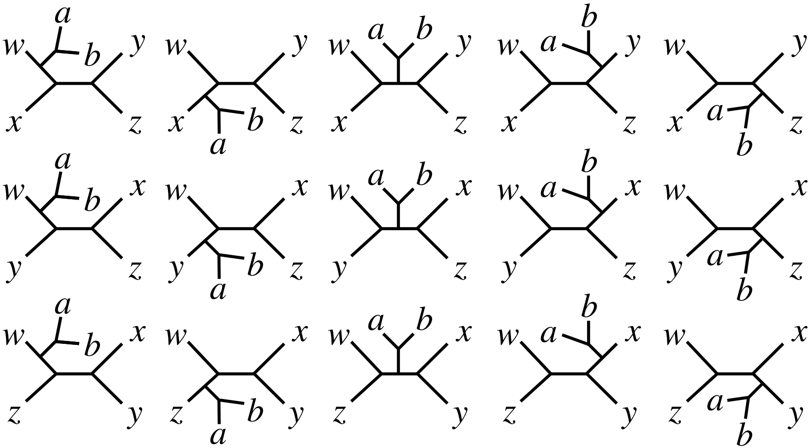

- Siblings: $\{c, e\}$ and $\{d, g\}$.

# Siblings




- Siblings: $\{c, e\}$ and $\{d, g\}$.

# Sextet topologies & a fixed pair of siblings



# Sextet topologies & a fixed pair of siblings (contd.)

- $\{a, b, w, x\}$, $\{a, b, w, y\}$, $\{a, b, w, z\}$, $\{a, b, x, y\}$, $\{a, b, x, z\}$, $\{a, b, y, z\}$ have determined topologies.
  - ▷ $[ab|wx]$, $[ab|wy]$, $[ab|wz]$, $[ab|xy]$, $[ab|xz]$, $[ab|yz]$.
- 9 quartet topologies undetermined.

# Sextet topologies & a fixed pair of siblings (contd.)

- $\{a, b, w, x\}$, $\{a, b, w, y\}$, $\{a, b, w, z\}$, $\{a, b, x, y\}$, $\{a, b, x, z\}$, $\{a, b, y, z\}$ have determined topologies.
  - ▷ $[ab|wx]$, $[ab|wy]$, $[ab|wz]$, $[ab|xy]$, $[ab|xz]$, $[ab|yz]$.
- 9 quartet topologies undetermined.

# The second algorithm

| branching vector | branching number |
|---|---|
| $(6,6,8,6,6,6,6,5,6,6,6,6,5,6,6)$ | $1.58005\ldots$ |
| $(5,6,6,5,6,6,6,5,6,6,7,6,7,6,7)$ | $1.58142\ldots$ |
| $\ldots$ | $\ldots$ |
| $(1,5,5,7,8,2,6,6,8,9,3,7,7,8,8)$ | $2.00904\ldots$ |
| $(1,5,5,9,9,2,6,6,6,8,3,7,7,7,9)$ | $2.01615\ldots$ |

**Theorem 3.2**

*There exists an $O(2.0162^k n^3 + n^5)$ fixed-parameter algorithm for the parameterized MQI problem.*

# The second algorithm

| branching vector | branching number |
|---|---|
| $(6,6,8,6,6,6,6,5,6,6,6,6,5,6,6)$ | $1.58005\ldots$ |
| $(5,6,6,5,6,6,6,5,6,6,7,6,7,6,7)$ | $1.58142\ldots$ |
| $\ldots$ | $\ldots$ |
| $(1,5,5,7,8,2,6,6,8,9,3,7,7,8,8)$ | $2.00904\ldots$ |
| $(1,5,5,9,9,2,6,6,6,8,3,7,7,7,9)$ | $2.01615\ldots$ |

**Theorem 3.2**

*There exists an $O(2.0162^k n^3 + n^5)$ fixed-parameter algorithm for the parameterized MQI problem.*

# Idea of the third algorithm

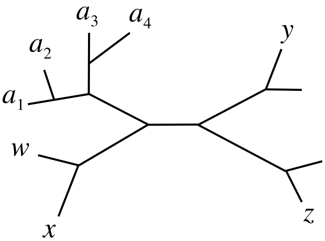
- Generalized from the second algorithm.
- Siblings $\Rightarrow$ adjacent taxa.

# Idea of the third algorithm

- Generalized from the second algorithm.
- Siblings $\Rightarrow$ adjacent taxa.

# Adjacent taxa

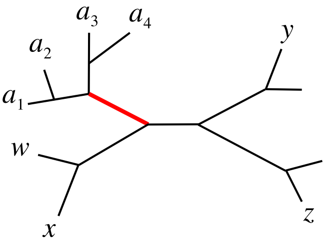
- Adjacent $m \geq 2$ taxa $a_1, \ldots, a_m$:
- $(\{a_1, \ldots, a_m\}, S \setminus \{a_1, \ldots, a_m\})$.




Given a number $2 \leq \omega \leq n/2$, there must be $m$ adjacent taxa, where $\omega \leq m \leq 2\omega - 2$.

# Adjacent taxa

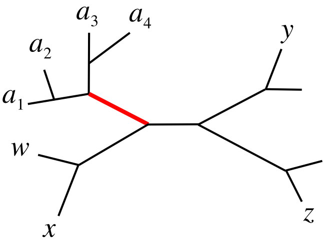
- Adjacent $m \geq 2$ taxa $a_1, \ldots, a_m$:
- $(\{a_1, \ldots, a_m\}, S \setminus \{a_1, \ldots, a_m\})$.




Given a number $2 \leq \omega \leq n/2$, there must be $m$ adjacent taxa, where $\omega \leq m \leq 2\omega - 2$.

# Adjacent taxa

- Adjacent $m \geq 2$ taxa $a_1, \ldots, a_m$:
- $(\{a_1, \ldots, a_m\}, S \setminus \{a_1, \ldots, a_m\})$.

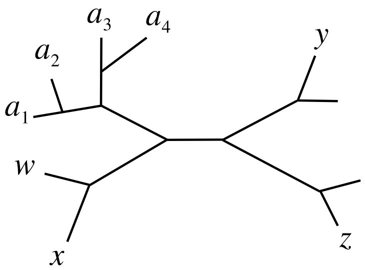



Given a number $2 \leq \omega \leq n/2$, there must be $m$ adjacent taxa, where $\omega \leq m \leq 2\omega - 2$.

# The third algorithm

- Change the topology of $\{a_1, w, x, y\}$.

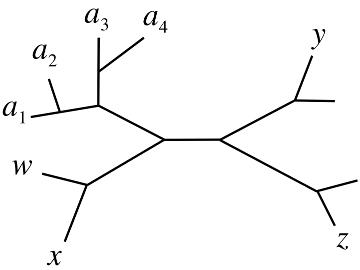
▷ Change the topologies of $\{a_2, w, x, y\}$, $\{a_3, w, x, y\}$, $\{a_4, w, x, y\}$ as well.



# The third algorithm

- Change the topology of $\{a_1, w, x, y\}$.

▷ Change the topologies of $\{a_2, w, x, y\}$, $\{a_3, w, x, y\}$, $\{a_4, w, x, y\}$ as well.



# Concluding theorem

- $O^*((1+2m^{-1/2})^k)$.
- Assume that $1+2m^{-1/2} \leq 1+\epsilon$ for some constant $\epsilon > 0$.
- Time complexity: $O((1+\epsilon)^k n^{8/\epsilon^2-1} + n^{8/\epsilon^2+1} + n^5)$.

**Theorem 3.3**

*There exists an $O^*((1+\epsilon)^k)$ time fixed-parameter algorithm for the parameterized MQI problem, where $\epsilon > 0$ is an arbitrarily small constant.*

# Concluding theorem

- $O^*((1+2m^{-1/2})^k)$.
- Assume that $1+2m^{-1/2} \leq 1+\epsilon$ for some constant $\epsilon > 0$.
- Time complexity: $O((1+\epsilon)^k n^{8/\epsilon^2-1} + n^{8/\epsilon^2+1} + n^5)$.

**Theorem 3.3**

*There exists an $O^*((1+\epsilon)^k)$ time fixed-parameter algorithm for the parameterized MQI problem, where $\epsilon > 0$ is an arbitrarily small constant.*

# Concluding theorem

- $O^*((1+2m^{-1/2})^k)$.
- Assume that $1+2m^{-1/2} \leq 1+\epsilon$ for some constant $\epsilon > 0$.
- Time complexity: $O((1+\epsilon)^k n^{8/\epsilon^2-1} + n^{8/\epsilon^2+1} + n^5)$.

**Theorem 3.3**

*There exists an $O^*((1+\epsilon)^k)$ time fixed-parameter algorithm for the parameterized MQI problem, where $\epsilon > 0$ is an arbitrarily small constant.*

# Thank you!