Supplementary Material Quartet-Net: A Quartet Based Method to Reconstruct Phylogenetic Networks

Jialiang Yang, Stefan Grünewald and Xiu-Feng Wan

1 k-weakly Compatible System

A split system $\mathcal{S}$ is *k-weakly compatible* if any $k+2$ splits $A_i|B_i \in \mathcal{S}$ with $i = 1, 2, \cdots, k+2$ are k-weakly compatible in the sense that if $|\bigcap_{i=1}^{k+2} A_i| > k-1$, then at least one of the $k+2$ intersections $A_i \cap \bigcup_{j \neq i, 1 \leq j \leq k+2} B_j$, $i = 1, 2, \cdots k+2$ is empty. It is clear from the definition that weakly compatible is the same as 1-weakly compatible.

Lemma 1. *$k-1$-weakly compatible split systems are also k-weakly compatible.*

Proof. Suppose $\mathcal{S}$ is a $k-1$-weakly compatible split system. For any $k+2$ splits $A_i|B_i \in \mathcal{S}$ with $i = 1, 2, \cdots, k+2$, if $|\bigcap_{i=1}^{k+2} A_i| > k-1$, then $|\bigcap_{i=1}^{k+1} A_i| > k-1 > k-2$ since $|\bigcap_{i=1}^{k+1} A_i| \geq |\bigcap_{i=1}^{k+2} A_i|$. Since $\mathcal{S}$ is $k-1$-weakly compatible, at least one of the $k+1$ intersections $A_i \cap \bigcup_{j \neq i, 1 \leq j \leq k+1} B_j$, $i = 1, 2, \cdots k+1$ is empty. Intersecting with B_{k+2}, one has that at least one of the $k+2$ intersections $A_i \cap \bigcup_{j \neq i, 1 \leq j \leq k+2} B_j$, $i = 1, 2, \cdots k+2$ is empty. By definition, $\mathcal{S}$ is also k-weakly compatible. $\square$

As a proposition, any weakly compatible system is also 2-weakly compatible. It is worth noting that the split system $\{xa|bc, xb|ac, xc|ab\}$ is 2-weakly compatible but not weakly compatible. So 2-weakly compatible is indeed a proper generalization of weakly compatible.

2 Proof of the equivalence between the recurrence system Eqn. (1) ~ (4) and Split Decomposition

In Split-Decomposition paper (Bandelt and Dress 1992), the split weight (isolation index) of any split $A|B$, denoted by $\alpha(A|B)$, is defined as

$$\begin{aligned}\alpha(A|B) \;=\; & \frac{1}{2}\min_{a,a'\in A;b,b'\in B}\{\max\{w(a|b)+w(a'|b'),w(a'|b)+w(a|b'),\\ & w(a|a')+w(b|b')\}-w(a|a')-w(b|b')\}.\end{aligned}$$

To prove the equivalence of Equations (1) to (4) and Split Decomposition, we next show that the function α and w are equivalent for any split $A|B$.

Proof. The objective is to show that $\alpha(A|B) = w(A|B)$ for any split $A|B$. We prove it case by case on the cardinalities $|A| = m$ and $|B| = n$.

(1) $m = 1$ and $n \geq 1$, that is, the case $a|A$. By definition,

$$\begin{aligned} & \alpha(a|A)\\ = \; & \frac{1}{2}\min_{b,b'\in A}\{\max\{ab+ab',aa+bb'\}-aa-bb'\}\\ = \; & \frac{1}{2}\min_{b,b'\in A}\{\max\{ab+ab'-bb',0\}\}\\ = \; & \max\left\{0,\frac{1}{2}\min_{b,b'\in A}\{ab+ab'-bb'\}\right\}\\ = \; & w(a|A)\end{aligned}$$

By symmetry, The equality also holds for $n = 1$ and $m \geq 1$.

(2) $m = 2$ and $n = 2$, that is, the case $aa'|bb'$. The following lemma is presented in Bandelt and Dress (1992a).

Lemma 2 (Bandelt and Dress, 1992a)**.** *Let $A_0|B_0$ be a partial split on X, then for all $x \in X - (A_0 \cup B_0)$*

$$\alpha(A_0x|B_0) + \alpha(A_0|B_0x) \leq \alpha(A_0|B_0).$$

Setting $A_0 = a$, $x = a'$, $B_0 = bb'$, one has

$$\alpha(aa'|bb') \leq \alpha(a|bb') - \alpha(a|a'bb').$$

Incorporating with $w(a|bb') = \alpha(a|bb')$ and $w(a|a'bb') = \alpha(a|a'bb')$, one has

$$\alpha(aa'|bb') \leq w(a|bb') - w(a|a'bb')$$

Similarly,

$$\alpha(aa'|bb') \leq w(a'|bb') - w(a'|abb'),$$
$$\alpha(aa'|bb') \leq w(aa'|b) - w(aa'b'|b),$$
$$\alpha(aa'|bb') \leq w(aa'|b') - w(aa'b|b').$$

By definition, $\alpha(aa'|bb') \leq w(aa'|bb')$.

We next prove $\alpha(aa'|bb') \geq w(aa|bb')$. For convenience, we let $\beta(aa'|bb') = \frac{1}{2}(\max\{ab + a'b', a'b + ab', aa' + bb'\} - aa' - bb')$, and denote $\beta(aa|bb)$ and $\beta(aa|bb')$ by $\beta(a|b)$ and $\beta(a|bb')$ respectively. Then, by definition

$$\alpha(aa'|bb') = \min\{\beta(a|b), \beta(a|b'), \beta(a'|b), \beta(a'|b'),$$
$$\beta(aa'|b), \beta(aa'|b'), \beta(a|bb'), \beta(a'|ab'), \beta(a|a'bb')\}.$$

If $\alpha(aa'|bb') = \beta(a|b)$, then $\alpha(aa'|bb') \geq w(aa'|bb')$ since $\beta(a|b) \geq \alpha(a|bb') \geq \alpha(a|bb') - \alpha(a|a'bb') \geq w(aa'|bb')$. Similarly, $\alpha(aa'|bb') \geq w(aa'|bb')$ if $\alpha(aa'|bb')$ equals to $\beta(a|b')$, $\beta(a'|b)$, $\beta(a'|b')$, $\beta(a|bb')$, $\beta(a'|bb')$, $\beta(a, a'|b)$ or $\beta(aa'|b')$.

Thus, without loss of generality, we assume

$$\alpha(aa'|bb') = \beta(aa'|bb') = \frac{1}{2}(\max\{ab + a'b', a'b + ab', aa' + bb'\} - aa' - bb').$$

Noting that $\alpha(a|a'bb') = \frac{1}{2}\min\{2aa', 2ab, 2ab', aa' + ab - a'b, aa' + ab' - a'b', ab + ab' - bb'\}$, we divide into 3 cases.

Case 1: $\alpha(a|a'bb') = ab$, then $\alpha(a|bb') = ab$, and thus $w(aa'|bb') \leq \alpha(a|bb') - \alpha(a|a'bb') = 0 \leq \alpha(aa'|bb')$. Similarly, we can prove $w(aa'|bb') \leq \alpha(aa'|bb')$ if $\alpha(a|a'bb')$ equals to $2ab'$ or $ab + ab' - bb'$.

Case 2: $\alpha(a|a'bb') = \frac{1}{2}\min\{aa'+ab-a'b, aa'+ab'-a'b'\}$, then

$$\begin{aligned}
& w(aa'|bb') \\
\leq\ & \alpha(a|bb') - \alpha(a|a'bb') \\
=\ & \frac{1}{2}\min\{2ab, 2ab', ab+ab'-bb'\} - \frac{1}{2}\min\{aa'+ab-a'b, aa'+ab'-a'b'\} \\
\leq\ & \frac{1}{2}(ab+ab'-bb') - \frac{1}{2}\min\{aa'+ab-a'b, aa'+ab'-a'b'\} \\
=\ & \frac{1}{2}(\max\{ab+a'b', a'b+ab'\} - aa' - bb') \\
=\ & \alpha(aa'|bb').
\end{aligned}$$

The last equality holds since $w(aa'|bb') \geq 0$ by definition.

Case 3: $\alpha(a|a'bb') = 2aa'$. Then, $2aa' \leq aa'+ab-a'b$, $2aa' \leq aa' + ab' - a'b'$, i.e.

$$\begin{aligned}
a'b &\leq ab - aa' \\
a'b' &\leq ab' - aa',
\end{aligned}$$

which implies

$$\begin{aligned}
a'b + a'b' - bb' &\leq ab + a'b' - aa' - bb' \\
a'b' + a'b - bb' &\leq ab' + a'b - aa' - bb'.
\end{aligned}$$

That is, $\beta(a'|bb') \leq \beta(aa'|bb')$. Thus, $\alpha(aa'|bb') \neq \beta(aa'|bb')$, but we have prove that $w(aa'|bb') \leq \alpha(aa'|bb')$ in this case.

In summary, $w(aa'|bb') = \alpha(aa'|bb')$ for $m = 2$ and $n = 2$.

(3) For $m \geq 2$ and $n \geq 2$, since both methods take the minimum of $w(a, a'|b, b')$ where $a, a' \in A$ and $b, b' \in B$, by induction, the weights calculated are the same.

□

3 Consistency of Quartet-Net

We assume that the quartet and triplet weights that are used as the input for Quartet-Net are induced by a weighted 2-weakly compatible split system.

More precisely, let $\mathcal{S}$ be a split system and let w be a weight function that associates a positive weight $w(S)$ to every full split S in $\mathcal{S}$. Then the induced weight $w(A'|B')$ of a (partial) split $A'|B'$ is defined by

$$w(A'|B') = \sum_{A|B \in \mathcal{S}: A|B \text{ displays } A'|B'} w(A|B).$$

In this section we will prove that Quartet-Net, run with the triplet and quartet weights induced by a 2-weakly compatible split system $\mathcal{S}$ with weight function w as its input, will output precisely all splits $A|B$ in $\mathcal{S}$ with weight $w(A|B)$. Note that this notion of induced weights for partial splits and consistency directly corresponds to the distance induced by a weighted split system and the consistency of a distance-based method.

We need the following property of 2-weakly compatible split systems.

Lemma 3. *Let $\mathcal{S}$ be a 2-weakly compatible split system and let $A|B$ be a non-trivial full split in $\mathcal{S}$. Then there is a non-trivial partial split $A^{'}|B'$ with $|A'|, |B'| \leq 3$ such that $A'|B'$ is displayed by $A|B$ and by no other full split in $\mathcal{S}$.*

Proof. Let $A'|B'$ be a maximal non-trivial split such that $A'|B'$ is displayed by $A|B$ and by no other full split in $\mathcal{S}$. Maximal means that every partial split that is displayed by $A'|B'$ but not equal to $A'|B'$ is displayed by at least two different split in $\mathcal{S}$. Such a split exists because $A|B$ is displayed by itself and by no other full split in $\mathcal{S}$.

Assume that A' or B', say B', has at least four different elements b_1, b_2, b_3, b_4 and let a_1, a_2 be different elements of A'. Then, for $i = 1, ..., 4$, there is a split $A_i|B_i$ in $\mathcal{S}$ such that $A_i|B_i$ is not equal to $A|B$ and displays $A'|B' - b_i$. Since $A|B$ is the only split in $\mathcal{S}$ that displays $A'|B'$, the split $A_i|B_i$ must display $A_i + b_i|B - b_i$. But then the four splits $A_i|B_i$ for $i = 1, ..., 4$ can not be 2-weakly compatible, a contradiction. □

We define $A|B$ to be an (i, j)-split if $|A| = i$ and $|B| = j$, or $|A| = j$ and $|B| = i$, and to be a 2-split if $|A| = 2$ or $|B| = 2$.

Proposition 1. *Let $\mathcal{S}$ be a 2-weakly compatible split system and let $A|B$ be a non-trivial 2-split in $\mathcal{S}$. Then there is a quartet or $(2, 3)$-split $A'|B'$ such that $A'|B'$ is displayed by a full split $A|B$ and by no other full split in $\mathcal{S}$.*

Proof. By Lemma 3, there is a non-trivial partial split $A'|B'$ that is displayed by $A|B$ and by no other full split in $\mathcal{S}$. Since $A'|B'$ is non-trivial, $|A'| \geq 2$ and $|B'| \geq 2$. In addition, $A|B$ is a 2-split and $A'|B'$ is displayed by $A|B$, which forces $A'|B'$ to be a 2-split, say $|A'| = 2$. By Lemma 3, $|B'| \leq 3$. Thus, $A'|B'$ is a quartet or $(2,3)$-split. □

The following lemma follows directly from the definition of the induced weight for partial splits.

Lemma 4. *Given a split system $\mathcal{S}$ with weight function w, let $A'|B'$ be a partial split that is displayed by a full split $A|B \in \mathcal{S}$ and by no other full split in $\mathcal{S}$, then $w(A'|B') = w(A|B)$.*

Theorem 1. *If the Quartet-Net algorithm is applied to triplet and quartet weights that are induced by a weighted 2-weakly compatible split system $\mathcal{S}$ on X, then it will output the splits in $\mathcal{S}$ with correct weights.*

Proof of Theorem 1. In view of the identities that were used to define the weights that Quartet-Net assigns to the $(2,3)$-splits, their weights are computed correctly. Note that the equality

$$w(aa'|bb'b'') = \frac{1}{2}\big(w(aa'|bb') - w(bb'|b''a) + w(b''a|a'b) - w(a'b|b'b'') + w(b'b''|aa')\big)$$

always holds (even after permuting a, a' resp. b, b', b'') since the quartet weights are induced by a weighted split system on $\{a, a', b, b', b''\}$. Hence, taking the minimum or the average of these six identical numbers does not effect the consistency of the algorithm. Further, Proposition 1 implies that for every 2-split $A|B$ in $\mathcal{S}$ there is a quartet or $(2,3)$-split that is displayed by $A|B$ and no other full split in $\mathcal{S}$, thus the weights of all 2-splits are computed correctly. Applying the same argument to subsets of X with six elements, we have that the weights of all (2,4)-splits are computed correctly as well. Since for every $(3,3)$-split $A|B$ with $a \in A$ the weight of $A|B$ equals the weight of $A - a|B$ minus the weight of $A - a|B + a$, the weights of all $(3,3)$-splits are computed correctly. We have proved that the weights of all non-trivial (i,j)-split with $i \leq 3$ and $j \leq 3$ are computed correctly. Lemma 3 implies that for every split $A|B$ in $\mathcal{S}$ with $|A|, |B| \geq 3$ there is an (i,j)-split with $i \leq 3$ and $j \leq 3$ that is displayed by $A|B$ and no other full split in $\mathcal{S}$, thus the weight of every such split is also computed correctly. Finally, the weight of every trivial split $x|X - x$ equals the weight of the triplet $x|yz$ minus the sum of the weights of all non-trivial splits in $\mathcal{S}$ that display $x|yz$ for every pair $\{y, z\} \subset X - x$. Hence, the weights of all trivial splits are computed correctly. $\square$

References

[1] Bandelt HJ, Dress AWM. 1992. A canonical decomposition theory for metrics on a finite set. Advances in mathematics. 92: 47-105.