

BOUNDING THE MAXIMUM SIZE OF A MINIMAL DEFINITIVE SET OF QUARTETS

MAX DIETRICH, CATHERINE MCCARTIN, AND CHARLES SEMPLE

ABSTRACT. Dowden (2010) showed that the maximum size of a minimal definitive set \mathcal{Q} of quartets is at least $2n - 8$, for all $n \geq 4$, where n is the size of the label set of \mathcal{Q} . In this paper, we show that the maximum size of such a set of quartets is at least $\Omega(n^2)$ and, moreover, is strictly less than n^3 , for all $n \geq 4$.

1. INTRODUCTION

Supertree methods are prominent methods for reconstructing phylogenetic (evolutionary) trees. These methods take as input a collection of smaller phylogenetic trees with overlapping taxa (leaf labels) and output a single parent tree whose taxa set is the union of the taxa sets of the smaller trees. For unrooted methods, the basic input is a collection of quartet trees, that is, binary phylogenetic trees with exactly four leaves. For an insight into the scope and variety of supertree methods, we refer the interested to [1]. Ideally, for any reconstruction algorithm, supertree method or otherwise, one would like the input to be compatible, in which case, there is a phylogenetic tree that is consistent with the input. Moreover, if the input is compatible, one would additionally like there to be precisely one such tree, that is, the input is definitive. In this regard, a natural task is to investigate definitive sets of quartets. In this paper, we study the problem of determining the maximum size of a minimal definitive set of quartets. The rest of the introduction formalizes the problem, details previous results, and states the main results of this paper. Throughout the paper, our notation and terminology follows Semple and Steel [4].

Let X be a finite set, where $|X| \geq 3$. A *phylogenetic X -tree* \mathcal{T} is a tree with no degree-two vertices, whose interior vertices are unlabeled and whose leaves are labeled distinctly by the elements of X . We often denote the label set of \mathcal{T} by $\mathcal{L}(\mathcal{T})$. A phylogenetic tree is *binary* if every interior vertex has degree three. A *quartet* is a binary phylogenetic tree with exactly four leaves.

Date: 18 May 2012.

1991 Mathematics Subject Classification. 05C05; 92D15.

Key words and phrases. Phylogenetic tree, quartets, definitiveness.

The first and third authors were supported by the Allan Wilson Centre for Molecular Ecology and Evolution. The third author was supported by the New Zealand Marsden Fund.

If q is a quartet with $\mathcal{L}(q) = \{a, b, c, d\}$ and the path from a to b does not intersect the path from c to d , then we denote q by $ab|cd$.

Let \mathcal{T} be a binary phylogenetic X -tree and let $q = ab|cd$ be a quartet. We say that \mathcal{T} *displays* q if $\{a, b, c, d\} \subseteq X$ and the path from a to b does not intersect the path from c to d in \mathcal{T} . More generally, let \mathcal{Q} be a set of quartets, then \mathcal{T} *displays* \mathcal{Q} if \mathcal{T} displays each quartet in \mathcal{Q} . Let $\mathcal{L}(\mathcal{Q}) = \bigcup_{q \in \mathcal{Q}} \mathcal{L}(q)$. If $\mathcal{L}(\mathcal{Q}) = X$ and \mathcal{T} is the *only* phylogenetic X -tree that displays \mathcal{Q} , then \mathcal{T} is *defined* by \mathcal{Q} and \mathcal{Q} is said to be *definitive*.

A definitive set \mathcal{Q} of quartets is *minimal* if, for each $q \in \mathcal{Q}$, the set $\mathcal{Q} - q$ is *not* definitive, in the sense that, if we let $X = \mathcal{L}(\mathcal{Q})$, then there is more than one phylogenetic X -tree that displays $\mathcal{Q} - q$. It is well-known that the minimum size of such a minimal definitive set is $n - 3$, where n is the size of $\mathcal{L}(\mathcal{Q})$ (for example, see [4]). In contrast, the maximum size of a minimal definitive set of quartets has remained unknown, other than that it is trivially bounded above by $O(n^4)$. Until recently, it was conjectured that the maximum size is $n - 3 + k$, where k is some constant. However, Humphries [3] showed that the maximum size is at least $\frac{3}{2}n - 6$, while Dowden [2] showed the maximum size is at least $2n - 8$, for all $n \geq 4$. In this paper, we show that the maximum size of a minimal definitive set of quartets is at least quadratic in n and, moreover, at most cubic in n , for all $n \geq 4$. In particular, we prove the following two results, whose proofs are in Sections 2 and 3.

Theorem 1.1. *Let X be a finite set such that $|X| = n$. For all $n \geq 4$, there is a minimal definitive set \mathcal{Q} of quartets, with $\mathcal{L}(\mathcal{Q}) = X$, such that $|\mathcal{Q}| = f(n)$, where $f(n)$ is $\Omega(n^2)$. More precisely,*

$$|\mathcal{Q}| = \sum_{i=1}^{n-3} \lceil i/2 \rceil = \begin{cases} \frac{1}{4}(n-3)(n-1) & \text{if } n \text{ is odd;} \\ \frac{1}{4}(n-2)^2 & \text{if } n \text{ is even.} \end{cases}$$

Theorem 1.2. *Let X be a finite set such that $|X| = n$. For all $n \geq 4$, every minimal definitive set \mathcal{Q} of quartets, with $\mathcal{L}(\mathcal{Q}) = X$, has the property that $|\mathcal{Q}| < n^3$. More precisely,*

$$|\mathcal{Q}| \leq \sum_{i=1}^{n-3} i(i+2).$$

The next lemma and its corollary will be used in the proofs of Theorems 1.1 and 1.2.

Lemma 1.3. *Let $\mathcal{P} = \{a_1a_2|yz, a_2a_3|yz, \dots, a_{k-2}a_{k-1}|yz, a_{k-1}a_k|yz\}$ be a set of quartets and let \mathcal{T} be a phylogenetic tree in which $\mathcal{L}(\mathcal{P}) \subseteq \mathcal{L}(\mathcal{T})$. If \mathcal{T} displays \mathcal{P} , then \mathcal{T} displays $a_1a_k|yz$.*

Proof. The proof is by induction on k . If $k = 2$, the lemma trivially holds. Suppose that $k \geq 3$ and the lemma holds for

$$\mathcal{P}' = \{a_1a_2|yz, a_2a_3|yz, \dots, a_{k-2}a_{k-1}|yz\},$$

that is, if \mathcal{T} displays \mathcal{P}' , then \mathcal{T} displays $a_1a_{k-1}|yz$. Now it is easily seen that if a phylogenetic tree displays $a_1a_{k-1}|yz$ and $a_{k-1}a_k|yz$, then it also displays $a_1a_k|yz$ (see [4]). This completes the proof of the lemma. \square

Corollary 1.4. *Let \mathcal{Q} be a set of quartets, and suppose that there is a “cyclic” subset \mathcal{S} of \mathcal{Q} of the form*

$$\mathcal{S} = \{a_1a_2|yz, a_2a_3|yz, \dots, a_{k-1}a_k|yz, a_ka_1|yz\}.$$

Then \mathcal{Q} is not a minimal definitive set of quartets.

Proof. The proof is by contradiction. Suppose that \mathcal{Q} is a minimal definitive set of quartets. Since \mathcal{Q} is definitive, \mathcal{Q} is displayed by a unique phylogenetic X -tree \mathcal{T} , where $X = \mathcal{L}(\mathcal{Q})$. Furthermore, as \mathcal{Q} is minimal, for each $q \in \mathcal{Q}$, the set $\mathcal{Q} - q$ is displayed by more than one phylogenetic X -tree. In particular, if s_i is a quartet in \mathcal{S} , then there is more than one phylogenetic X -tree displaying $\mathcal{Q} - s_i$. By Lemma 1.3, any phylogenetic X -tree that displays $\mathcal{S} - s_i$ also displays \mathcal{S} . Thus, there must be more than one phylogenetic X -tree that displays \mathcal{Q} ; a contradiction. \square

2. PROOF OF THEOREM 1.1

Let X be a finite set such that $|X| = n$, where $n \geq 4$. Without loss of generality, we set $X = \{1, 2, 3, \dots, n\}$. We begin by constructing a set of quartets \mathcal{Q}_n , with $\mathcal{L}(\mathcal{Q}_n) = X$, such that $|\mathcal{Q}_n| = \sum_{i=1}^{n-3} \lceil i/2 \rceil$. We then show that \mathcal{Q}_n is a minimal definitive set of quartets, thereby proving Theorem 1.1.

For each $i \in \{1, 2, \dots, n-3\}$, set E_i as follows. If i is odd, set

$$E_i = \begin{cases} \{12|(i+2)(i+4)\} \cup \bigcup_{j=1}^{\frac{i-1}{2}} \{(2j)(2j+2)|(i+2)(i+4)\}, & \text{if } i < n-3; \\ \{12|(i+2)(i+3)\} \cup \bigcup_{j=1}^{\frac{i-1}{2}} \{(2j)(2j+2)|(i+2)(i+3)\}, & \text{if } i = n-3. \end{cases}$$

If i is even, set

$$E_i = \begin{cases} \bigcup_{j=0}^{\frac{i-2}{2}} \{(2j+1)(2j+3)|(i+2)(i+4)\}, & \text{if } i < n-3; \\ \bigcup_{j=0}^{\frac{i-2}{2}} \{(2j+1)(2j+3)|(i+2)(i+3)\}, & \text{if } i = n-3. \end{cases}$$

For convenience later in the proof, for when i is odd, we view the quartets $12|(i+2)(i+4)$ and $12|(i+2)(i+3)$ as resulting from setting $j = 0$ when $i < n-3$ and $i = n-3$, respectively. Now set

$$\mathcal{Q}_n = E_1 \cup E_2 \cup \dots \cup E_{n-3}.$$

Note that $|\mathcal{Q}_n| = \sum_{i=1}^{n-3} \lceil i/2 \rceil$. To illustrate the construction, if $n = 8$, then $E_1 = \{12|35\}$, $E_2 = \{13|46\}$, $E_3 = \{12|57, 24|57\}$, $E_4 = \{13|68, 35|68\}$, and $E_5 = \{12|78, 24|78, 46|78\}$. Thus

$$\mathcal{Q}_8 = \{12|35, 13|46, 12|57, 24|57, 13|68, 35|68, 12|78, 24|78, 46|78\}.$$

For all $n \geq 4$, let \mathcal{C}_n denote the binary phylogenetic X -tree shown in Fig. 1. For now, ignore the labeling of the interior edges. It is easily checked that

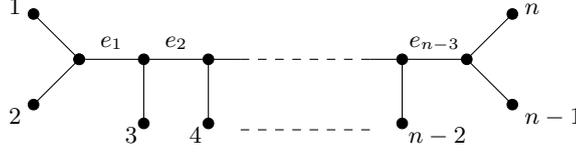


FIGURE 1. The phylogenetic tree \mathcal{C}_n . For each fixed $n \geq 4$, the phylogenetic tree \mathcal{C}_n displays the set \mathcal{Q}_n of quartets constructed at the start of Section 2.

\mathcal{C}_n displays \mathcal{Q}_n for all $n \geq 4$. We will show that \mathcal{C}_n is the *only* phylogenetic X -tree to display \mathcal{Q}_n , thus \mathcal{Q}_n is definitive, and that, for each $q \in \mathcal{Q}_n$, the set $\mathcal{Q}_n - q$ is not definitive, thus \mathcal{Q}_n is minimal.

We require one further definition before showing that \mathcal{Q}_n is definitive for all $n \geq 4$. Let \mathcal{T} be a phylogenetic X -tree and let $q = ab|cd$ be a quartet with $\mathcal{L}(q) \subseteq X$. Suppose that \mathcal{T} displays q . We say that q *distinguishes* an edge e of \mathcal{T} if the path from a to b contains one end vertex of e , while the path from c to d contains the other end vertex of e .

Lemma 2.1. *For all $n \geq 4$, the set \mathcal{Q}_n is definitive.*

Proof. If $n = 4$, then $\mathcal{Q}_4 = \{12|34\}$ and this set is clearly definitive, so assume that $n > 4$. First observe that, for each $i \in \{1, 2, \dots, n-4\}$, we can write E_i , for i odd, in the form

$$\{12|(i+2)(i+4), 24|(i+2)(i+4), \dots, (i-1)(i+1)|(i+2)(i+4)\}$$

and, for i even,

$$\{13|(i+2)(i+4), 35|(i+2)(i+4), \dots, (i-1)(i+1)|(i+2)(i+4)\}.$$

Furthermore, when $i = n-3$, we can write E_i , for i odd, in the form

$$\{12|(i+2)(i+3), 24|(i+2)(i+3), \dots, (i-1)(i+1)|(i+2)(i+3)\},$$

and, for i even,

$$\{13|(i+2)(i+3), 35|(i+2)(i+3), \dots, (i-1)(i+1)|(i+2)(i+3)\}.$$

Now let \mathcal{T} be a phylogenetic X -tree that displays E_i for all $i \in \{1, 2, \dots, n-3\}$. By Lemma 1.3, \mathcal{T} displays $1(i+1)|(i+2)(i+4)$ for each $i \in \{1, 2, \dots, n-4\}$ and $1(i+1)|(i+2)(i+3)$ for $i = n-3$. In particular, every phylogenetic X -tree that displays \mathcal{Q}_n , where $n > 4$, displays each of the quartets $12|35, 13|46, \dots, 1(n-2)|(n-1)n$. Since each interior edge e_i of \mathcal{C}_n is distinguished by one of the quartets in the set $\{12|35, 13|46, \dots, 1(n-2)|(n-1)n\}$, and all of the quartets in this set share the common leaf labeled 1, it follows by [4, Proposition 6.4.4] that \mathcal{C}_n is the only phylogenetic X -tree that displays \mathcal{Q}_n . Hence \mathcal{Q}_n is definitive. \square

The next lemma is used in showing that \mathcal{Q}_n has the desired minimality property. In particular, for each $n \geq 4$, it provides an alternative way to

view \mathcal{Q}_n . We use this “mirror” view of \mathcal{Q}_n as a convenient tool to show that for each quartet q in E_{n-3} , the set $\mathcal{Q}_n - q$ is not definitive.

Lemma 2.2. *For all $n \geq 4$, let*

$$\mathcal{Q}_n^* = \{(n+1-a_4)(n+1-a_3)|(n+1-a_2)(n+1-a_1) : a_1a_2|a_3a_4 \in \mathcal{Q}_n\}.$$

Then $\mathcal{Q}_n^ = \mathcal{Q}_n$.*

Proof. We show that $\mathcal{Q}_n^* \subseteq \mathcal{Q}_n$, by showing that every quartet $q \in \mathcal{Q}_n^*$ is contained in \mathcal{Q}_n . Then, since clearly $|\mathcal{Q}_n^*| = |\mathcal{Q}_n|$, it must be the case that $\mathcal{Q}_n^* = \mathcal{Q}_n$.

Choose a quartet q from \mathcal{Q}_n^* , of the form

$$(n+1-a_4)(n+1-a_3)|(n+1-a_2)(n+1-a_1),$$

where $a_1a_2|a_3a_4$ is a quartet in \mathcal{Q}_n . Recall that $\mathcal{Q}_n = E_1 \cup E_2 \cup \dots \cup E_{n-3}$. For all (i, j) such that $1 \leq i \leq n-3$ and $0 \leq j < |E_i|$, let $q_{i,j}$ denote the j -th element of E_i . It must be the case that $a_1a_2|a_3a_4 = q_{i,j}$ for some $1 \leq i \leq n-3$ and $0 \leq j < |E_i|$.

We have four cases to consider, dependent on the parity of both n and i . We complete the proof for the case that both n and i are odd and note that the other three cases are similar.

Given that n and i are both assumed to be odd, and hence $1 \leq i \leq n-4$ and $0 \leq j \leq \frac{i-1}{2}$, the construction states that

$$q_{i,j} = \begin{cases} 12|(i+2)(i+4), & \text{if } j = 0; \\ (2j)(2j+2)|(i+2)(i+4), & \text{if } j > 0. \end{cases}$$

Thus, q must be of the form

$$q = \begin{cases} (n+1-(i+4))(n+1-(i+2)|(n+1-2)(n+1-1), & \text{if } j = 0; \\ (n+1-(i+4))(n+1-(i+2)|(n+1-(2j+2))(n+1-2j), & \text{if } j > 0. \end{cases}$$

We can rewrite q as

$$q = \begin{cases} (n-i-3)(n-i-1)|(n-1)n, & \text{if } j = 0; \\ (n-i-3)(n-i-1)|(n-1-2j)(n+1-2j), & \text{if } j > 0. \end{cases}$$

Now set $i' = n-3-2j$ and $j' = \frac{n-i-4}{2}$. Note that, when n is odd, i' is even. Let

$$q_{i',j'} = \begin{cases} (2j'+1)(2j'+3)|(i'+2)(i'+4), & \text{if } i' < n-3; \\ (2j'+1)(2j'+3)|(i'+2)(i'+3), & \text{if } i' = n-3. \end{cases}$$

The quartet $q_{i',j'}$ will be in \mathcal{Q}_n if and only if $2 \leq i' \leq n-3$, and $0 \leq j' \leq \frac{i'-2}{2}$. As $0 \leq j \leq \frac{i-1}{2}$, we get $n-i-2 \leq i' \leq n-3$. So, as $1 \leq i \leq n-4$, we get $2 \leq i' \leq n-3$. Now, as $i' \geq n-i-2$, we have $\frac{i'-2}{2} \geq \frac{n-i-4}{2}$. Since $j' = \frac{n-i-4}{2}$ and $1 \leq i \leq n-4$, this gives us $0 \leq j' \leq \frac{i'-2}{2}$. Hence, if both n and i are odd and $q_{i,j} \in \mathcal{Q}_n$, then $q_{i',j'} \in \mathcal{Q}_n$.

Finally, we rewrite $q_{i',j'}$ by substituting $n - 3 - 2j$ for i' and $\frac{n-i-4}{2}$ for j' to get

$$q_{i',j'} = \begin{cases} (n-i-3)(n-i-1)|(n-1-2j)(n+1-2j), & \text{if } j > 0; \\ (n-i-3)(n-i-1)|(n-1)n, & \text{if } j = 0. \end{cases}$$

It is clear that $q = q_{i',j'}$. Thus, if n and i are both odd and $a_1a_2|a_3a_4 = q_{i,j}$ for some $1 \leq i \leq n-3$ and $0 \leq j < |E_i|$, then $q \in \mathcal{Q}_n$.

The other three cases are similar. In particular, for those cases, we select i' and j' as follows:

$$(i', j') = \begin{cases} (n-3-2j, \frac{n-i-4}{2}) & \text{if } n \text{ odd, } i \text{ odd;} \\ (n-3-2j, \frac{n-i-3}{2}) & \text{if } n \text{ even, } i \text{ odd;} \\ (n-4-2j, \frac{n-i-4}{2}) & \text{if } n \text{ even, } i \text{ even;} \\ (n-4-2j, \frac{n-i-3}{2}) & \text{if } n \text{ odd, } i \text{ even.} \end{cases}$$

□

We now complete the proof of Theorem 1.1.

Proof of Theorem 1.1. By Lemma 2.1, \mathcal{Q}_n is definitive for all $n \geq 4$. We establish the minimality part of the theorem by induction on n . Since $\mathcal{Q}_4 = \{12|34\}$ and $\mathcal{Q}_4 - 12|34 = \emptyset$, it follows that any one of the three distinct binary phylogenetic trees with labeled set $\{1, 2, 3, 4\}$ displays $\mathcal{Q}_4 - 12|34$. Therefore, if $n = 4$, then the minimality property holds.

Suppose that the minimality property holds for \mathcal{Q}_{n-1} , where $n > 4$. Then, for $X = \{1, 2, \dots, n-1\}$ and each $q \in \mathcal{Q}_{n-1}$, there is a phylogenetic X -tree, other than \mathcal{C}_{n-1} , that displays $\mathcal{Q}_{n-1} - q$. Now let $q_{i,j} \in \mathcal{Q}_n$, where i denotes that $q_{i,j} \in E_i$ and, if i is odd, j denotes the value in $\{0, 1, 2, \dots, \frac{i-1}{2}\}$ resulting in $q_{i,j}$ in the construction of E_i while, if i is even, j denotes the value in $\{0, 1, 2, \dots, \frac{i-2}{2}\}$ resulting in $q_{i,j}$ in the construction of E_i .

First suppose that $i \in \{1, 2, \dots, n-4\}$ and consider the quartet $q_{i,j}$. If $i \in \{1, 2, \dots, n-5\}$, then, for all j , we have that $q_{i,j}$ is a quartet in \mathcal{Q}_{n-1} , in which case, there is a phylogenetic tree $\mathcal{T}_{i,j}$ with label set $X - n$, other than \mathcal{C}_{n-1} , that displays $\mathcal{Q}_{n-1} - q_{i,j}$. By adjoining a new leaf n to $\mathcal{T}_{i,j}$ via the pendant edge incident with $n-1$, it is easily checked that the resulting tree displays $\mathcal{Q}_n - q_{i,j}$, but it is not \mathcal{C}_n .

For $i = n-4$, we can similarly show that $\mathcal{Q}_n - q_{i,j}$ is not definitive as follows. If i is odd, then \mathcal{Q}_n contains the subset

$$(1) \quad \{12|(i+2)(i+4)\} \cup \bigcup_{j=1}^{\frac{i-1}{2}} \{(2j)(2j+2)|(n-2)n\}$$

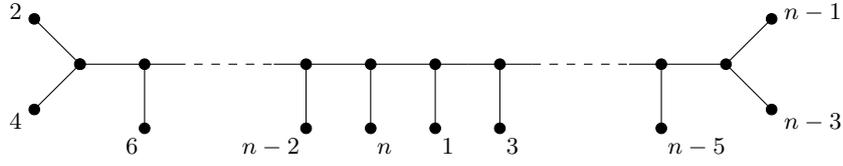


FIGURE 2. For each fixed $n \geq 4$, where n even, this phylogenetic tree displays $\mathcal{Q}_n - 12|(n-1)n$.

of quartets and \mathcal{Q}_{n-1} contains the subset

$$(2) \quad \{12|(i+2)(i+3)\} \cup \bigcup_{j=1}^{\frac{i-1}{2}} \{(2j)(2j+2)|(n-2)(n-1)\}$$

of quartets. If i is even, then \mathcal{Q}_n contains the subset

$$(3) \quad \bigcup_{j=0}^{\frac{i-2}{2}} \{(2j+1)(2j+3)|(n-2)n\}$$

of quartets and \mathcal{Q}_{n-1} contains the subset

$$(4) \quad \bigcup_{j=0}^{\frac{i-2}{2}} \{(2j+1)(2j+3)|(n-2)(n-1)\}$$

of quartets. Let $\mathcal{T}_{i,j}$ be a phylogenetic tree with label set $X - n$, other than \mathcal{C}_{n-1} , that displays $\mathcal{Q}_{n-1} - q_{i,j}^{n-1}$ where, depending upon whether i is odd or even, $q_{i,j}^{n-1}$ is the j -th quartet in (2) or (4), respectively, in the construction of \mathcal{Q}_{n-1} . It is now easily checked that, by adjoining a new leaf n to $\mathcal{T}_{i,j}$ via the pendant edge incident with $n-1$, the resulting tree displays $\mathcal{Q}_n - q_{i,j}$, but it is not \mathcal{C}_n where, depending upon whether i is odd or even, $q_{i,j}$ is the j -th quartet in (1) or (3), respectively, in the construction of \mathcal{Q}_n .

It remains to show how an alternative phylogenetic X -tree displaying $\mathcal{Q}_n - q_{i,j}$ is produced for $q_{i,j}$ when $i = n-3$. By Lemma 2.2, we know that, for each such quartet $q_{i,j} = a_1 a_2 |(n-1)n$, a corresponding quartet $12|(n+1-a_2)(n+1-a_1)$ is an element of \mathcal{Q}_n . This ‘‘mirror quartet’’ corresponds to some quartet $q_{i',j'} \in \mathcal{Q}_n$, where i' is odd and $j' = 0$. If $i' < n-3$, we know by above that there is a phylogenetic X -tree $\mathcal{T}'_{i',j'}$, other than \mathcal{C}_n , that displays $\mathcal{Q}_n - q_{i',j'}$. Replacing each leaf l by $(n+1-l)$ in $\mathcal{T}'_{i',j'}$ will result in a tree $\mathcal{T}'_{i,j}$ that displays $\mathcal{Q}_n - q_{i,j}$, but it is not \mathcal{C}_n .

If $i' = n-3$, it must be the case that $i' = i = n-3$ is odd, and so n is even and $q_{i,j} = q_{i',j'} = 12|(n-1)n$. Consider the phylogenetic X -tree, \mathcal{C}'_n say, shown in Fig 2. Since \mathcal{C}'_n displays $2n|1(n-1)$, it does not display $12|(n-1)n$. However, \mathcal{C}'_n does display each of the remaining quartets in \mathcal{Q}_n since each such quartet takes one of the following forms: $12|a(a+2)$, where a is odd; $a(a+2)|b(b+2)$ or $b(b+2)|a(a+2)$, where a is odd and b is even; $a(a+2)|(n-1)n$, where a is even. This completes the proof of Theorem 1.1. \square

3. PROOF OF THEOREM 1.2

Proof of Theorem 1.2. Let X be a finite set such that $|X| = n$, where $n \geq 4$. Without loss of generality, we set $X = \{1, 2, 3, \dots, n\}$, and let \mathcal{Q} be an arbitrary minimal definitive set of quartets with $\mathcal{L}(\mathcal{Q}) = X$. Fixing a 2-element subset $\{y, z\} \subseteq X$, let G_{yz} be the graph with vertex set $X - \{y, z\}$ and where there is an edge $\{a, b\}$ in G_{yz} precisely if $ab|yz$ is a quartet in \mathcal{Q} . Since \mathcal{Q} is a minimal definitive set, it follows by Corollary 1.4 that G_{yz} is acyclic. Therefore, G_{yz} has at most $n - 3$ edges. Thus, \mathcal{Q} contains at most $n - 3$ quartets of the form $ab|yz$, where $a, b \in X - \{y, z\}$.

Now let P_X be the set of 2-element subsets of X and observe that

$$|P_X| = (n - 1) + (n - 2) + \dots + 2 + 1.$$

Furthermore, for each $i \in \{1, 2, \dots, n\}$ and $j \in \{i + 1, \dots, n\}$, let G_{ij} be the graph with vertex set $X - \{\{1, 2, \dots, i - 1\} \cup \{i, j\}\}$ and where there is an edge $\{k, l\}$ precisely if there is a quartet in \mathcal{Q} of the form $ij|kl$. For each element in P_X of the form $\{1, j\}$, where $2 \leq j \leq n$, the graph G_{1j} has at most $n - 3$ edges by the argument in the previous paragraph. In general, for each $i \in \{1, 2, \dots, n\}$ and $j \in \{i + 1, \dots, n\}$, the graph G_{ij} has at most $n - (i + 2)$ edges. For each $i \in \{1, 2, \dots, n\}$, let \mathcal{Q}_i be the subset of quartets in \mathcal{Q} of the form $ij|kl$, where $i + 1 \leq j \leq n$ and $k, l \in X - \{\{1, 2, \dots, i - 1\} \cup \{i, j\}\}$. Since G_{ij} has at most $n - (i + 2)$ edges, for each $j \in \{i + 1, \dots, n\}$, it must be the case that $|\mathcal{Q}_i| \leq (n - i)(n - (i + 2))$. Since

$$\mathcal{Q} = \mathcal{Q}_1 \cup \mathcal{Q}_2 \cup \dots \cup \mathcal{Q}_{n-3},$$

it follows that $|\mathcal{Q}| \leq \sum_{i=1}^{n-3} (n - i)(n - (i + 2)) = \sum_{i=1}^{n-3} i(i + 2)$. \square

ACKNOWLEDGEMENTS

We thank the anonymous referees for their careful reading of the paper and comments.

REFERENCES

- [1] O.R.P. Bininda-Emonds (ed.), *Phylogenetic supertrees: Combining information to reveal the Tree of Life*, Kluwer, Dordrecht, 2004.
- [2] C. Dowden, On the maximum size of minimal definitive quartet sets, *Discrete Mathematics* 310 (2010) 2546–2549.
- [3] P.J. Humphries, *Combinatorial aspects of leaf-labelled trees*, Ph.D. Thesis, University of Canterbury, 2008.
- [4] C. Semple, M. Steel, *Phylogenetics*, Oxford University Press, Oxford, 2003.

SCHOOL OF ENGINEERING AND ADVANCED TECHNOLOGY, MASSEY UNIVERSITY, NEW ZEALAND

E-mail address: `max.dietrich@massey.ac.nz`

SCHOOL OF ENGINEERING AND ADVANCED TECHNOLOGY, MASSEY UNIVERSITY, NEW ZEALAND

E-mail address: `c.m.mccartin@massey.ac.nz`

BIOMATHEMATICS RESEARCH CENTRE, DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF CANTERBURY, CHRISTCHURCH, NEW ZEALAND

E-mail address: `charles.semple@canterbury.ac.nz`