Applied Mathematics Letters 22 (2009) 1604–1608

Contents lists available at ScienceDirect

Applied Mathematics Letters

journal homepage: www.elsevier.com/locate/aml

Quartets in maximal weakly compatible split systems

S. Grünewald [a,b], J.H. Koolen [c,d,*], W.S. Lee [d]

[a] Institute for Computational Biology, Shanghai Institutes for Biological Sciences, 320 Yue Yang Road, Shanghai 200031, China
[b] Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
[c] Pohang Mathematics Institute, POSTECH, Hyoja-dong, Namgu, Pohang 790-784, South Korea
[d] Department of Mathematics, POSTECH, Hyoja-dong, Namgu, Pohang 790-784, South Korea

ARTICLE INFO

Article history:
Received 5 May 2009
Accepted 5 May 2009

Keywords:
Split system
Quartet
Weakly compatible
Full weakly compatible split system
k-compatible
Full circular split system
Maximal weakly compatible split system

ABSTRACT

Weakly compatible split systems are a generalization of unrooted evolutionary trees and are commonly used to display reticulate evolution or ambiguity in biological data. They are collections of bipartitions of a finite set X of taxa (e.g. species) with the property that, for every four taxa, at least one of the three bipartitions into two pairs (quartets) is not induced by any of the X-splits. We characterize all split systems where exactly two quartets from every quadruple are induced by some split. On the other hand, we construct maximal weakly compatible split systems where the number of induced quartets per quadruple tends to 0 with the number of taxa going to infinity.

1. Introduction

In biology, phylogenetics is the study of the evolutionary history of a collection of taxonomic units like species or populations. Usually, a phylogenetic analysis starts with morphological or molecular data, e.g. DNA sequences, and aims to construct a leaf-labeled tree that shows how the taxa evolved from their last common ancestor. However, for many data sets there are contradicting signals that cannot be displayed by a single tree, and sometimes, due to reticulate evolution events like recombination or lateral gene transfer, a correct tree does not even exist. In those cases it is often useful to consider structures more general than trees.

The information of an unlabeled unweighted phylogenetic tree with taxa set X can be formulated in terms of *splits*, that is bipartitions of X: Removing any edge from the tree yields two connected components and the sets of taxa in each component define a bipartition of X. Conversely, a set of splits of X corresponds to a phylogenetic tree if and only if, for each two splits, we can select one part from each split such that their intersection is empty [1]. Collections of splits that satisfy this condition are called *compatible*. By relaxing the compatibility condition, we get a generalization of phylogenetic trees. We will give its definition after introducing some notation.

The set of all splits of X is denoted by $\Sigma(X)$. Any subset Σ of $\Sigma(X)$ is called a *split system* of X. Given a split $A|B \in \Sigma(X)$, the number $\min\{|A|, |B|\}$ is also called the *size* of that split, and denoted by $\|A|B\|$. A split $A|B$ of X of size k is called a k-split and a 1-split is called a *trivial* split, and the set of all non-trivial splits in $\Sigma(X)$ is denoted by $\Sigma^*(X)$. We will not distinguish between $A|B$ and $B|A$, as both terms stand for the same bipartition $\{A, B\}$ of X. A split system Σ is called *weakly compatible* if for any $\{A_i, B_i\} \in \Sigma$ $(i = 1, 2, 3)$ at least one of the intersections $A_1 \cap A_2 \cap A_3$, $A_1 \cap B_2 \cap B_3$, $B_1 \cap A_2 \cap B_3$, $B_1 \cap B_2 \cap A_3$ is empty. A split system Σ is called *maximal weakly compatible* if it is weakly compatible and there does not exist a weakly

* Corresponding author at: Department of Mathematics, POSTECH, Hyoja-dong, Namgu, Pohang 790-784, South Korea.
E-mail addresses: stefan@picb.ac.cn (S. Grünewald), koolen@postech.ac.kr (J.H. Koolen), lws82@postech.ac.kr (W.S. Lee).

0893-9659/$ – see front matter
doi:10.1016/j.aml.2009.05.006

compatible split system Σ' such that $\Sigma \subset \Sigma'$. Equivalently, weakly compatible split systems can be defined in terms of their quartet sets: A *quartet q* on X is a partition of a four-element subset $\{a_1, a_2, b_1, b_2\}$ of X into two two-element subsets $\{a_1, a_2\}$ and $\{b_1, b_2\}$. Any such partition is denoted, for short, by $a_1a_2|b_1b_2$. We denote the set of all quartets on X by $\mathcal{Q}(X)$. A subset of $\mathcal{Q}(X)$ is called a *quartet system.* A quartet $a_1a_2|b_1b_2$ is *displayed* by split $A|B$ if $a_1, a_2 \in A$ and $b_1, b_2 \in B$ or $a_1, a_2 \in B$ and $b_1, b_2 \in A$.
For $\Sigma \subseteq \Sigma^*(X)$,

$$\mathcal{Q}_\Sigma := \{a_1a_2|b_1b_2 : \exists S = A|B \in \Sigma \text{ such that } \{a_1, a_2\} \subseteq A, \{b_1, b_2\} \subseteq B\}.$$

For $\mathcal{Q} \subseteq \mathcal{Q}(X)$,

$$\Sigma_\mathcal{Q} := \{A|B \in \Sigma^*(X) : \text{ for all } a_1 \neq a_2 \in A, b_1 \neq b_2 \in B, a_1a_2|b_1b_2 \in \mathcal{Q}\}.$$

The following result is easy to prove and was established in [2].

Proposition 1.1. *Let X be a finite and non-empty set and $\Sigma \subseteq \Sigma^*(X)$. Then $\mathcal{Q}_\Sigma$ has the property that for all x, y, u, v in X, there are at most two quartets on x, y, u, v if and only if Σ is weakly compatible.*

A *maximal circular split system* is a special weakly compatible split system that is defined by a circular ordering of X. Let C be a cycle (graph) with vertex set X. We define $\Sigma(C)$ to be the split system containing precisely all splits $A|B$ for which there are two edges e, f of C such that A and B are the vertex sets of the two components of $C - e - f$. A *circular split system* is a subset of any maximal circular split system. Note that compatible split systems are circular.

The methods most widely used to construct weakly compatible split systems are split decompositions [3] as implemented in SplitsTree [4]and NeighborNet [5]. While split decomposition often produces too few splits for real data, NeighborNet and its quartet-based analogue QNet [6] can only reconstruct circular split systems. In order to develop variants of NeighborNet or QNet that can reconstruct more general split systems, a better understanding of weakly compatible split systems and their quartet systems would be desirable.

In this note we consider the cardinality of $\mathcal{Q}_\Sigma$ for a weakly compatible split system Σ of X. We show that, for $|X| \geq 7$, $|\mathcal{Q}_\Sigma| = 2\binom{|X|}{4}$ if and only if Σ is maximal circular and we construct maximal weakly compatible split systems where the number of induced quartets per quadruple tends to 0 with the number of taxa going to infinity.

2. Full weakly compatible split systems

We recall that a weakly compatible split system without trivial splits can be reconstructed from the quartet set.

Lemma 2.1. *Let X be a finite and non-empty set and let $\Sigma \subseteq \Sigma^*(X)$. If Σ is weakly compatible then $\Sigma_{\mathcal{Q}_\Sigma} = \Sigma$.*

Let Σ be a split system on X. Define a 2-split graph $G_2(\Sigma)$ with X as its vertices and with a_1 adjacent to a_2 if and only if $\{a_1, a_2\}|(X - a_1 - a_2) \in \Sigma$. We say that Σ is a *full weakly compatible split system* if, for all four taxa $x, y, u, v \in X$, there are exactly two quartets on x, y, u, v in $\mathcal{Q}_\Sigma$. Note that such a split system is weakly compatible by Proposition 1.1.

Lemma 2.2. *Let X be a finite and non-empty set with $|X| = n$ and let Σ be a full weakly compatible split system. Then the degree of any vertex of $G_2(\Sigma)$ is at most 2. In particular, there are at most n 2-splits in Σ.*

Proof. Let $X = \{1, 2, \ldots, n\}$. Then the degree of any vertex of $G_2(\Sigma)$ is at most 2. Otherwise, $G_2(\Sigma)$ has $K_{1,3}$ as a subgraph and this implies that there is a quadruple which has three quartets, a contradiction. The number of edges is maximal if $G_2(\Sigma)$ is 2-regular; hence the result follows. ■

The main result of this section is a characterization of full weakly compatible split systems. Obviously, such a split system contains exactly two out of three possible quartets for the four-taxa case. We distinguish split systems with five, six, seven, or more taxa.

Proposition 2.3. *Let $X = \{1, 2, 3, 4, 5\}$ and let Σ be a full weakly compatible split system. Then Σ is isomorphic to one of the following:*

1. $\Sigma = \{12|345, 13|245, 25|134, 35|124\}$*: a 4-cycle and an isolated vertex;*
2. $\Sigma = \{12|345, 13|245, 25|134, 34|125, 45|123\}$*: a pentagon.*

Proof. All splits in Σ have to be 2-splits so the split system is defined by $G_2(\Sigma)$. In view of Lemma 2.2, the maximum degree is at most 2. Since every 2-split displays only three quartets and $\mathcal{Q}_\Sigma$ contains ten quartets, Σ contains at least four splits. There are only three graphs with at least four edges and maximum degree at most 2, a 4-cycle, a 5-cycle, and a path of length 4. Only the 4-cycle and the 5-cycle yield a full weakly compatible split system. ■