Applied Mathematics Letters 22 (2009) 1789–1792

Contents lists available at ScienceDirect

Applied Mathematics Letters

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

A Hall-type theorem for triplet set systems based on medians in trees

Andreas Dress [a], Mike Steel [b,*]

[a] CAS-MPG Partner Institute for Computational Biology, 320 Yue Yang Road, 200031 Shanghai, China

[b] Biomathematics Research Centre, University of Canterbury, Christchurch, New Zealand

ARTICLE INFO

Article history:
Received 26 June 2009
Accepted 1 July 2009

Keywords:
Trees
Median vertices
Systems of distinct representatives
Patchworks

ABSTRACT

Given a collection $\mathscr{C}$ of subsets of a finite set X, let $\bigcup \mathscr{C} = \cup_{S \in \mathscr{C}} S$. Philip Hall's celebrated theorem [P. Hall, On representatives of subsets, J. London Math. Soc. 10 (1935) 26–30] concerning 'systems of distinct representatives' tells us that for any collection $\mathscr{C}$ of subsets of X there exists an injective (i.e. one-to-one) function $f : \mathscr{C} \to X$ with $f(S) \in S$ for all $S \in \mathscr{C}$ if and and only if $\mathscr{C}$ satisfies the property that for all non-empty subsets $\mathscr{C}'$ of $\mathscr{C}$, we have $|\bigcup \mathscr{C}'| \geq |\mathscr{C}|$. Here, we show that if the condition $|\bigcup \mathscr{C}'| \geq |\mathscr{C}'|$ is replaced by the stronger condition $|\bigcup \mathscr{C}'| \geq |\mathscr{C}'| + 2$, then we obtain a characterization of this condition for a collection of 3-element subsets of X in terms of the existence of an injective function from $\mathscr{C}$ to the vertices of a tree whose vertex set includes X and which satisfies a certain median condition. We then describe an extension of this result to collections of arbitrary-cardinality subsets of X.

1. First result

Given a tree $T = (V, E)$ and a subset S of V of size 3, say $S = \{x, y, z\}$, consider the path in T connecting x, y, the path connecting x, z and the path connecting y, z. There is a unique vertex that is shared by these three paths, the *median vertex* of S in T, denoted $\text{med}_T(S)$. Our first result provides an analogue of Hall's theorem [3] described in the abstract.

Theorem 1.1. *Let X be a finite set, and suppose that $\mathscr{C} \subseteq \binom{X}{3}$, and $\bigcup \mathscr{C} = X$. The following are equivalent:*

(1) *There exists a tree $T = (V, E)$ with $X \subseteq V$ for which the function $S \mapsto \text{med}_T(S)$ from $\mathscr{C}$ to V is injective.*
(2) *There exists a tree $T = (V, E)$ with X as its set of leaves, and all its other vertices of degree 3, for which the function $S \mapsto \text{med}_T(S)$ from $\mathscr{C}$ to the set of interior vertices of T is injective.*
(3) *$\mathscr{C}$ satisfies the following property. For all non-empty subsets $\mathscr{C}'$ of $\mathscr{C}$, we have:*

$$\left|\bigcup \mathscr{C}'\right| \geq |\mathscr{C}'| + 2. \qquad (1)$$

In order to establish Theorem 1.1, we first require a lemma.

Recall from [1] that a collection $\mathscr{P}$ of subsets of a set M forms a *patchwork* if it satisfies the following property:

$$A, B \in \mathscr{P} \text{ and } A \cap B \neq \emptyset \Longrightarrow A \cap B, \quad A \cup B \in \mathscr{P}.$$

Lemma 1.2. *Let X be a finite set, and suppose that $\mathscr{C} \subseteq \binom{X}{3}$, and $\bigcup \mathscr{C} = X$. If $\mathscr{C}$ satisfies the condition described in Part (3) of Theorem 1.1 then the collection $\mathscr{P}$ of non-empty subsets $\mathscr{C}'$ of $\mathscr{C}$ that satisfy $|\bigcup \mathscr{C}'| = |\mathscr{C}'| + 2$ forms a patchwork.*

* Corresponding author.
E-mail addresses: andreas@picb.ac.cn (A. Dress), m.steel@math.canterbury.ac.nz (M. Steel).

0893-9659/$ – see front matter
doi:10.1016/j.aml.2009.07.001

Proof. Suppose $\mathscr{C}_1, \mathscr{C}_2 \subseteq \mathscr{C}$, and that $\mathscr{C}_1 \cap \mathscr{C}_2 \neq \emptyset$. Consider

$$K := \left|\bigcup(\mathscr{C}_1 \cap \mathscr{C}_2)\right| + \left|\bigcup(\mathscr{C}_1 \cup \mathscr{C}_2)\right|.$$

By (1), we have:

$$K \geq (|\mathscr{C}_1 \cap \mathscr{C}_2| + 2) + (|\mathscr{C}_1 \cup \mathscr{C}_2| + 2) = |\mathscr{C}_1| + |\mathscr{C}_2| + 4, \tag{2}$$

and we also have:

$$K \leq \left|\left(\bigcup \mathscr{C}_1\right) \cap \left(\bigcup \mathscr{C}_2\right)\right| + \left|\left(\bigcup \mathscr{C}_1\right) \cup \left(\bigcup \mathscr{C}_2\right)\right| = \left|\bigcup \mathscr{C}_1\right| + \left|\bigcup \mathscr{C}_2\right|. \tag{3}$$

Notice that the right-hand term in (2) and (3) are equal, since $|\bigcup \mathscr{C}_i| = |\mathscr{C}_i| + 2$ as $\mathscr{C}_i \in \mathscr{P}$ for $i = 1, 2$, and thus the inequality in (2) is an equality. Therefore $|\bigcup(\mathscr{C}_1 \cap \mathscr{C}_2)| = |\mathscr{C}_1 \cap \mathscr{C}_2| + 2$ and $|\bigcup(\mathscr{C}_1 \cup \mathscr{C}_2)| = |\mathscr{C}_1 \cup \mathscr{C}_2| + 2$, as required. □

Proof of Theorem 1.1. The implication (2) ⇒ (1) is trivial. For the reverse implication suppose that T satisfies the property described in (2). First delete from T any vertices and edges that are not on a path between two vertices in X. Next attach to every interior (non-leaf) vertex $v \in X$ a new edge for which the adjacent new leaf is assigned the label x, and henceforth do not regard v as an element of X. Next replace each maximal path of degree 2 vertices by a single edge. Finally, replace each vertex v of degree $d > 3$ by an arbitrary tree that has d leaves that we identify with the neighboring vertices of v and whose remaining vertices have degree 3. These four processes result in a tree T' that has X as its set of leaves, and which has all its remaining vertices of degree 3 (i.e. a 'binary phylogenetic X-tree' [2]) and for which the median vertices of the elements of $\mathscr{C}$ remain distinct. Thus (1) and (2) are equivalent.

Next we show that (2) ⇒ (3). Suppose T satisfies the condition (2) and that $\mathscr{C}'$ is a non-empty subset of $\mathscr{C}$. Consider the minimal subtree of T that connects the leaves in $\bigcup \mathscr{C}'$. This tree has at least $|\mathscr{C}'|$ interior vertices that are of degree 3. However, by a simple counting argument, any tree that has k interior vertices of degree 3 must have at least $k + 2$ leaves, and so (1) holds.

The remainder of the proof is devoted to establishing that (3) ⇒ (2). We use induction on $n := |X|$. The result clearly holds for $n = 3$, so suppose it holds whenever $|X| < n$, $n \geq 4$ and that X is a set of size n. For $x \in X$, let $n_{\mathscr{C}}(x)$ be the number of triples in $\mathscr{C}$ that contain x. If there exists $x \in X$ with $n_{\mathscr{C}}(x) = 1$, then select the unique triple in $\mathscr{C}$ containing x, say $\{a, b, x\}$ and let $X' = X - \{x\}$, $\mathscr{C}' = \mathscr{C} - \{\{a, b, x\}\}$. Then $\bigcup \mathscr{C}' = X'$ and $\mathscr{C}'$ satisfies condition (1) and so, by induction, there is a tree T' with leaf set X' for which the median vertices of elements in $\mathscr{C}'$ are all distinct vertices of T'. Let T be the tree obtained from T' by subdividing one of the edges in the path in T' connecting a and b, and making the newly-created vertex of degree 2 adjacent to x by a new edge. Then T satisfies the requirements of Theorem 1.1(2), and thereby establishes the induction step in this case.

Thus we may suppose that $n_{\mathscr{C}}(x) > 1$ holds for all $x \in X$. In this case, we claim that there exists $x \in X$ with $n_{\mathscr{C}}(x) = 2$. Let us count the set $\Omega := \{(x, S) : x \in S \in \mathscr{C}\}$ in two different ways. We have:

$$|\Omega| = \sum_{x \in X} n_{\mathscr{C}}(x) \geq 2k + 3(n - k), \tag{4}$$

where $k = |\{x \in X : n_{\mathscr{C}}(x) = 2\}|$.

On the other hand:

$$|\Omega| = 3|\mathscr{C}| \leq 3(n - 2), \tag{5}$$

where the latter inequality follows from Inequality (1) applied to $\mathscr{C}' = \mathscr{C}$. Combining (4) and (5) gives $2k + 3(n-k) \leq 3n - 6$, and so $k \geq 6$. Thus, since $k > 0$, there exists $x \in X$ with $n_{\mathscr{C}}(x) = 2$, as claimed.

For any such $x \in X$ with $n_{\mathscr{C}}(x) = 2$, let $\{a, b, x\}$ and $\{a', b', x\}$ be the two elements of $\mathscr{C}$ containing x. Without loss of generality there are two cases:

(i) $a = a'$, $b \neq b'$; or
(ii) $\{a, b\} \cap \{a', b'\} = \emptyset$.

In case (i), let:

$$X' := X - \{x\}, \qquad \mathscr{C}' := \mathscr{C} - \left\{\{a, b, x\}, \{a, b', x\}\right\}, \qquad \mathscr{C}_1 := \mathscr{C}' \cup \left\{\{a, b, b'\}\right\}.$$

Note that $\bigcup \mathscr{C}_1 = X'$. Suppose that $\mathscr{C}_1$ fails to satisfy the condition described in Part (3) of Theorem 1.1. Then there is a subset of $\mathscr{C}_1$ that violates Inequality (1) of the form $\mathscr{C}^1 \cup \left\{\{a, b, b'\}\right\}$ where $a, b, b' \in \bigcup \mathscr{C}^1$ and $\mathscr{C}^1 \subseteq \mathscr{C}'$. But in that case $\mathscr{C}^1 \cup \left\{\{a, b, x\}, \{a, b', x\}\right\}$ would violate Inequality (1), which is impossible since Inequality (1) applies to this set, being a non-empty subset of $\mathscr{C}$. Thus, $\mathscr{C}_1$ satisfies Part (3) of Theorem 1.1. Since $\bigcup \mathscr{C}_1 = X'$, which has one less element than X, the inductive hypothesis furnishes a tree T' with leaf set X' that satisfies the requirements of Theorem 1.1(2). Now consider the edge of T' that is incident with leaf b'. Subdivide this edge and make the newly-created midpoint vertex adjacent to a leaf labelled x. This gives a tree T that has X as its set of leaves, and with all its interior vertices of degree 3; moreover, the medians of the elements of $\mathscr{C}$ are all distinct (note that the median of $\{x, a, b'\}$ is the newly-created vertex adjacent to x, while the median of $\{x, a, b\}$ corresponds to the median vertex of $\{a, b, b'\}$ in T' and therefore is a different vertex in T to any other median vertex of an element of $\mathscr{C}$).

In case (ii), let:

$$X' := X - \{x\}, \qquad \mathscr{C}' := \mathscr{C} - \left\{\{a, b, x\}, \{a', b', x\}\right\},$$

and let:

$$\mathscr{C}_1 := \mathscr{C}' \cup \{\{a, a', b\}\}, \qquad \mathscr{C}_2 := \mathscr{C}' \cup \left\{\{a, a', b'\}\right\}.$$

Note that $\bigcup \mathscr{C}_1 = \bigcup \mathscr{C}_2 = X'$. We will establish the following:

Claim: One or both of $\mathscr{C}_1$ or $\mathscr{C}_2$ satisfies the condition described in Part (3) of Theorem 1.1.

Suppose to the contrary that both sets fail the condition described in Theorem 1.1(3). Then there is a subset of $\mathscr{C}_1$ that violates Inequality (1), and it must be of the form $\mathscr{C}^1 \cup \{\{a, a', b\}\}$ where $\mathscr{C}^1 \subseteq \mathscr{C}'$, $a, a', b \in \bigcup \mathscr{C}^1$ and $b' \notin \bigcup \mathscr{C}^1$ (the last claim is justified by the observation that if $b' \in \bigcup \mathscr{C}^1$ then $\mathscr{C}^1 \cup \left\{\{a, b, x\}, \{a', b', x\}\right\}$ would violate the condition described in Part (3) of Theorem 1.1). Similarly a subset of $\mathscr{C}_2$ that violates Inequality (1) is of the form $\mathscr{C}^2 \cup \left\{\{a, a', b'\}\right\}$ where $\mathscr{C}^2 \subseteq \mathscr{C}'$, $a, a', b' \in \bigcup \mathscr{C}^2$ and $b \notin \bigcup \mathscr{C}^2$. Now, let $\mathscr{P}$ be the subset of $\mathscr{C}$ defined in the statement of Lemma 1.2. Then the sets

$$\mathscr{C}_1 := \mathscr{C}^1 \cup \left\{\{x, a, b\}, \{x, a', b'\}\right\}; \quad \text{and} \quad \mathscr{C}_2 := \mathscr{C}^2 \cup \left\{\{x, a, b\}, \{x, a', b'\}\right\}$$

are both elements of $\mathscr{P}$ and they have non-empty intersection, since they both contain $\{x, a, b\}$ (indeed, they also share $\{x, a', b'\}$). Thus, Lemma 1.2 ensures that $\mathscr{C}_1 \cap \mathscr{C}_2$ is also an element of $\mathscr{P}$. However $\mathscr{C}_1 \cap \mathscr{C}_2$ is of the form $\mathscr{C}^3 \cup \left\{\{x, a, b\}, \{x, a', b'\}\right\}$ where $\mathscr{C}^3 \subseteq \mathscr{C}'$, and neither x, b, nor b' is an element of $\bigcup \mathscr{C}^3$ because, by our choice of x, x only occurs in the two triples $\{x, a, b\}$ and $\{x, a', b'\}$, and because $b' \notin \bigcup \mathscr{C}^1$ and $b \notin \bigcup \mathscr{C}^2$. Since $\mathscr{C}^3$ is a subset of $\mathscr{C}$, $\mathscr{C}^3$ satisfies Inequality (1), which implies that (1) must be a strict inequality for $\mathscr{C}_1 \cap \mathscr{C}_2$, contradicting our assertion that $\mathscr{C}_1 \cap \mathscr{C}_2 \in \mathscr{P}$. This justifies our claim that either $\mathscr{C}_1$ or $\mathscr{C}_2$ satisfies part (3) of Theorem 1.1.

We may suppose then, without loss of generality, that $\mathscr{C}_1$ satisfies part (3) of Theorem 1.1. Since $\bigcup \mathscr{C}_1 = X'$, which has one less element than X, the inductive hypothesis furnishes a tree T' with leaf set X' that satisfies the requirements of Theorem 1.1(2). Now, consider the edge of T' that is incident with leaf a'. Subdivide this edge and make the newly-created midpoint vertex adjacent to a leaf labelled x by a new edge. This gives a tree T that has X as its set of leaves, and with all vertices of degree 3; moreover, regardless of where b' attaches in T, the medians of the elements of $\mathscr{C}$ are all distinct (note that the median of $\{x, a', b'\}$ is the newly-created vertex adjacent to x, while the median of $\{x, a, b\}$ corresponds to the median vertex of $\{a, a', b\}$ in T' and therefore is a different vertex in T from any other median vertex of an element of $\mathscr{C}$). This completes the proof. □

2. An extension

For a subset Y of X of size at least 3, and a tree $T = (V, E)$, with $X \subseteq V$, let

$$\mathrm{med}_T(Y) := \{\mathrm{med}_T(S) : S \subseteq Y, |S| = 3\}.$$

Thus, $\mathrm{med}_T(Y)$ is a subset of the vertices of T. Moreover, if X is the set of leaves of T then $\mathrm{med}_T(Y)$ is a subset of the interior vertices of T.

Theorem 2.1. *Let X be a finite set, and suppose that $\mathscr{C}$ is a collection of subsets of X, each of size at least 3, and with $\bigcup \mathscr{C} = X$. The following are equivalent:*

(1) *There exists a tree $T = (V, E)$ with X as its set of leaves, and all its other vertices of degree 3, for which $\{\mathrm{med}_T(Y) : Y \in \mathscr{C}\}$ is a partition of the set of interior vertices of T.*

(2) *$\mathscr{C}$ satisfies the following property. For all non-empty subsets $\mathscr{C}'$ of $\mathscr{C}$, we have:*

$$\left|\bigcup \mathscr{C}'\right| - 2 \geq \sum_{Y \in \mathscr{C}'} (|Y| - 2), \tag{6}$$

and this last inequality is an equality when $\mathscr{C}' = \mathscr{C}$.

Proof. We first show that (1) ⇒ (2). Select a tree T satisfying the requirements of Part (1) of Theorem 2.1. For a non-empty subset $\mathscr{C}'$ of $\mathscr{C}$, the minimal subtree T' of T connecting the leaves in $\bigcup \mathscr{C}'$ has $k := |\bigcup \mathscr{C}'|$ leaves, and $k - 2$ vertices that are of degree 3. By the partitioning assumption, each element $Y \in \mathscr{C}'$ generates $|Y| - 2$ median vertices in T and these sets of median vertices are pairwise disjoint for different choices of $Y \in \mathscr{C}'$. Moreover, distinct interior vertices of T correspond to different degree 3 vertices in T', and so the number of degree 3 vertices in T' can be no smaller than the sum of $|Y| - 2$ over all $Y \in \mathscr{C}'$. This establishes Inequality (6). For the case where $\mathscr{C}' = \mathscr{C}$, note that T has $\bigcup \mathscr{C} = X$ as its leaf set and, by the partitioning assumption, each of its $|X| - 2$ interior vertices occurs in one set $\mathrm{med}_T(Y)$ for some $Y \in \mathscr{C}$, and so $|X| - 2 \leq \sum_{Y \in \mathscr{C}'} (|Y| - 2)$ which, combined with (6), provides the desired equality.

To show (2) ⇒ (1), select for each set $Y \in \mathscr{C}$ a collection $\mathscr{C}_Y$ of 3-element subsets of X of cardinality $|Y| - 2$ for which $\bigcup C_Y = Y$ and which satisfies the condition that for every non-empty subset $\mathscr{C}'$ of $\mathscr{C}_Y$, we have $\bigcup \mathscr{C}' \geq |\mathscr{C}'| + 2$; such a selection is straightforward — for example, if $Y = \{y_1, \ldots, y_m\}$ then we can take:

$$\mathscr{C}_Y = \big\{\{y_1, y_2, y_3\}, \{y_1, y_2, y_4\}, \ldots, \{y_1, y_2, y_m\}\big\}. \tag{7}$$

We first establish the following:

Claim: $\mathscr{C}_* := \cup_{Y \in \mathscr{C}} \mathscr{C}_Y$ is a collection of 3-element subsets of X that satisfies Inequality (1) in Theorem 1.1.

To see this, suppose to the contrary that there exists a subset $\mathscr{C}''$ of $\mathscr{C}_*$ for which Inequality (1) fails. Write $\mathscr{C}'' = S_1 \cup S_2 \cup \cdots \cup S_k$ where $1 \leq k \leq |\mathscr{C}|$ and where S_i is a non-empty set of 3-element subsets of X that are selected from the same set (let us call it Y_i) from $\mathscr{C}$ (note that the fact that $|Y_1 \cup Y_2| - 2 \geq |Y_1| - 2 + |Y_2| - 2$ must hold for all Y_1, Y_2 in $\mathscr{C}$ implies that $|Y_1| + |Y_2| - |Y_1 \cap Y_2| - 2 \geq |Y_1| - 2 + |Y_2| - 2$ and, hence, $2 \geq |Y_1 \cap Y_2|$ must hold for all Y_1, Y_2 in $\mathscr{C}$). By our assumption regarding the set of triples $\mathscr{C}''$ we have $|\bigcup \mathscr{C}''| \leq |\mathscr{C}''| + 1$ and so, if we let $W_i := \bigcup S_i$ we have $\bigcup \mathscr{C}'' = \bigcup_{i=1}^k W_i$, and consequently:

$$\left|\bigcup_{i=1}^k W_i\right| \leq \sum_{i=1}^k |S_i| + 1. \tag{8}$$

For $\mathscr{C}' := \{Y_1, \ldots, Y_k\} \subseteq \mathscr{C}$, we have:

$$\left|\bigcup \mathscr{C}'\right| \geq \sum_{i=1}^k (|Y_i| - 2) + 2 = \sum_{i=1}^k |Y_i| - 2k + 2. \tag{9}$$

On the other hand:

$$\left|\bigcup \mathscr{C}'\right| \leq \left|\bigcup_{i=1}^k W_i\right| + \sum_{i=1}^k (|Y_i - W_i|) = \left|\bigcup_{i=1}^k W_i\right| + \sum_{i=1}^k (|Y_i| - |W_i|),$$

since $W_i \subseteq Y_i$. By the condition imposed on the construction of $\mathscr{C}_Y$, we have $|W_i| \geq |S_i| + 2$ for each i, and so substituting this and (8) into the previous inequality gives:

$$\left|\bigcup \mathscr{C}'\right| \leq \sum_{i=1}^k |S_i| + 1 + \sum_{i=1}^k |Y_i| - \sum_{i=1}^k (|S_i| + 2) = \sum_{i=1}^k |Y_i| - 2k + 1,$$

which, compared with (9), gives $1 \geq 2$, a contradiction. This establishes that C_* satisfies Inequality (1) in Theorem 1.1.

By Theorem 1.1 it now follows that there is a tree $T = (V, E)$ with leaf set X for which the function $S \mapsto \mathrm{med}_T(S)$ is injective from $\mathscr{C}_*$ to the set of interior vertices of T. Now for $Y \in \mathscr{C}$, we have:

$$\mathrm{med}_T(Y) = \{\mathrm{med}_T(S) : S \subseteq Y, |S| = 3\} = \{\mathrm{med}_T(S) : S \in C_Y\}. \tag{10}$$

The second equality in (10) requires some justification. Recalling our particular choice of C_Y from (7), and noting that the medians of the triples in C_Y are distinct vertices of T, it follows that $T|Y$ has the structure of a path connecting y_1, y_2 with each of the remaining leaves $y \in Y - \{y_1, y_2\}$ separated from this path by just one edge. Consequently, if a vertex v of T is the median of three leaves in Y then it is also the median of a triple $\{y_1, y_2, y\}$ for some $y \in Y - \{y_1, y_2\}$; that is, it is an element of $\{\mathrm{med}_T(S) : S \in C_Y\}$.

Consequently, $\{\mathrm{med}_T(Y) : Y \in \mathscr{C}\}$ are disjoint subsets of the set of interior vertices of T. Moreover, each interior vertex of T is covered by $\{\mathrm{med}_T(Y) : Y \in \mathscr{C}\}$ since the number of interior vertices is $|X| - 2$ and, by assumption, $|X| - 2 = \sum_{Y \in \mathscr{C}} (|Y| - 2) = |\mathscr{C}_*|$. This establishes the implication (2) ⇒ (1) and thereby completes the proof. □

Acknowledgement

We thank the New Zealand Marsden Fund for supporting this work.

References

[1] S. Böcker, A.W.M. Dress, Patchworks, Advances in Mathematics 157 (1) (2001) 1–21.
[2] C. Semple, M. Steel, Phylogenetics, Oxford University Press, 2003.
[3] P. Hall, On representatives of subsets, J. London Math. Soc. 10 (1935) 26–30.