



'Bureaucratic' set systems, and their role in phylogenetics

David Bryant^a, Mike Steel^{b,*}

^a Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand

^b Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand

ARTICLE INFO

Article history:

Received 9 June 2011

Received in revised form 7 November 2011

Accepted 10 February 2012

Keywords:

Cluster

Hierarchy

Tree

Algorithm

Phylogenetics

ABSTRACT

We say that a collection \mathcal{C} of subsets of X is *bureaucratic* if every maximal hierarchy on X contained in \mathcal{C} is also maximum. We characterize bureaucratic set systems and show how they arise in phylogenetics. This framework has several useful algorithmic consequences: we generalize some earlier results and derive a polynomial-time algorithm for a parsimony problem arising in phylogenetic networks.

© 2012 Elsevier Ltd. All rights reserved.

1. Bureaucratic sets and their characterization

In this work we introduce and study a class of set systems that arise in various ways from trees, graphs and intervals. We are interested in this class because it can provide a setting in which certain hard optimization problems can be solved efficiently, and we provide a particular example of this for a parsimony problem on phylogenetic networks.

We first recall some standard phylogenetic terminology (for more details, the reader can consult [1]). Recall that a *hierarchy* \mathcal{H} on a finite set X is a collection of sets with the property that the intersection of any two sets is either empty or equal to one of the two sets.

A hierarchy is *maximum* if $|\mathcal{H}| = 2|X| - 1$, which is the largest possible cardinality. In this case \mathcal{H} corresponds to the set of clusters $c(T)$ of some rooted binary tree T with leaf set X (a *cluster* of T is the set of leaves that are separated from the root of the tree by any vertex). A maximum hierarchy necessarily contains $\{x\}$ for each $x \in X$, as well as X itself; we will refer to these $|X| + 1$ sets as the *trivial clusters* of X . More generally, any hierarchy containing all the trivial clusters corresponds to the clusters $c(T)$ of a rooted tree T with leaf set X (examples of these concepts are illustrated in Fig. 1(a), (b)). Note that a hierarchy \mathcal{H} is maximum if and only if (i) \mathcal{H} contains all the trivial clusters, and (ii) each set $C \in \mathcal{H}$ of size greater than 1 can be written as a disjoint union $C = A \sqcup B$, for two (disjoint) sets $A, B \in \mathcal{H}$.

We now introduce a new notion.

Definition. We say that a collection \mathcal{C} of subsets of a finite set X is a *bureaucracy* if (i) $\mathcal{C} \neq \emptyset$ and $\emptyset \notin \mathcal{C}$, and (ii) every hierarchy $\mathcal{H} \subseteq \mathcal{C}$ can be extended to a maximum hierarchy \mathcal{H}' such that $\mathcal{H} \subseteq \mathcal{H}' \subseteq \mathcal{C}$. In this case, we also say that \mathcal{C} is *bureaucratic*.

Simple examples of bureaucracies include two extreme cases: the set of clusters of a binary tree, and the set $\mathcal{P}(X)$ of all non-empty subsets of X . Notice that $\{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, b, c\}\}$ and $\{\{a\}, \{b\}, \{c\}, \{b, c\}, \{a, b, c\}\}$ are both bureaucratic subsets of $\mathcal{P}(X)$ for $X = \{a, b, c\}$ but their intersection, $\{\{a\}, \{b\}, \{c\}, \{a, b, c\}\}$, is not. In particular, for an arbitrary subset Y of $\mathcal{P}(X)$ (e.g. $Y = \{\{a\}, \{b\}, \{c\}, \{a, b, c\}\}$), there may not be a unique minimal bureaucratic subset of $\mathcal{P}(X)$ containing Y .

* Corresponding author.

E-mail addresses: david.bryant@otago.ac.nz (D. Bryant), mike.steel@canterbury.ac.nz, mathmomike@gmail.com (M. Steel).

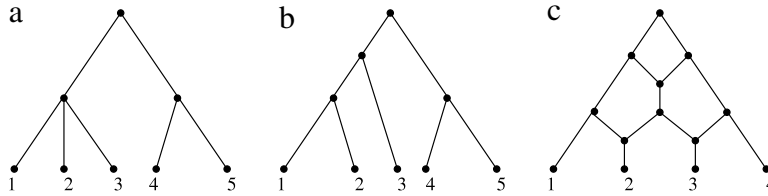


Fig. 1. (a) A rooted tree T with leaf set $X = \{1, 2, 3, 4, 5\}$, and with the cluster set $c(T)$ being equal to the hierarchy \mathcal{H} consisting of the sets $\{1, 2, 3\}$, $\{4, 5\}$ and the trivial clusters. (b) A binary tree T with a cluster set consisting of $\mathcal{H} \cup \{\{1, 2\}\}$. (c) A binary and planar phylogenetic network \mathcal{N} over $X = \{1, 2, 3, 4\}$ with a soft-wired cluster set $sw(\mathcal{N})$ consisting of $\{1, 2\}$, $\{2, 3\}$, $\{3, 4\}$, $\{1, 2, 3\}$, $\{2, 3, 4\}$ and the trivial clusters.

In the next section we describe a more extensive list of examples, but first we describe some properties and provide a characterization of bureaucracies. In the following lemma, given two sets A and B from \mathcal{C} we say that B covers A if $A \subsetneq B$ and there is no set $C \in \mathcal{C}$ with $A \subsetneq C \subsetneq B$.

Lemma 1. *If \mathcal{C} is bureaucratic then:*

- (i) For any pair $A, B \in \mathcal{C}$, if B covers A then $B - A \in \mathcal{C}$.
- (ii) For any $C \in \mathcal{C}$ with $|C| > 1$, we can write $C = A \sqcup B$ for (disjoint) sets $A, B \in \mathcal{C}$.

Proof. For Part (i), suppose that $A, B \in \mathcal{C}$ and that B covers A . Let $\mathcal{H} = \{A, B\}$. Then \mathcal{H} is a hierarchy that is contained within \mathcal{C} and so there exists a maximum hierarchy $\mathcal{H}' \subseteq \mathcal{C}$ that contains \mathcal{H} . Note that A must be a maximal sub-cluster of B in \mathcal{H}' (as otherwise B does not cover A) which requires that $B - A$ is a cluster of \mathcal{H}' and thereby an element of \mathcal{C} .

For Part (ii), observe that the set $\mathcal{H} = \{C\}$ is a hierarchy, and the assumption that \mathcal{C} is bureaucratic ensures the existence of a maximum hierarchy $\mathcal{H}' \subseteq \mathcal{C}$ containing \mathcal{H} , and so \mathcal{H}' contains the required sets A, B . \square

Note that the conditions described in Parts (i) and (ii) of Lemma 1, while they are necessary for \mathcal{C} to be a bureaucracy, are not sufficient. For example, let $X = \{1, 2, 3, 4, 5, 6\}$ and let \mathcal{C} be the union of

$$\{\{1, 2\}, \{3, 4\}, \{5, 6\}, \{1, 2, 3\}, \{4, 5, 6\}, \{3, 4, 5\}, \{1, 2, 6\}, \{1, 5, 6\}, \{2, 3, 4\}\}$$

with the set of the seven trivial clusters. Then \mathcal{C} satisfies Parts (i) and (ii) of Lemma 1, yet \mathcal{C} is not bureaucratic since $\mathcal{H} = \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$ does not extend to a maximum hierarchy on X using just elements from \mathcal{C} .

Theorem 2. *A collection \mathcal{C} of subsets of X is bureaucratic if and only if it satisfies the following two properties:*

- (P1) \mathcal{C} contains all trivial clusters of X .
- (P2) If $\{C_1, C_2, \dots, C_k\} \subseteq \mathcal{C}$ are disjoint and have union $\cup_i C_i$ in \mathcal{C} then there are distinct i, j such that $C_i \cup C_j \in \mathcal{C}$.

Proof. First suppose that \mathcal{C} is bureaucratic. Then \mathcal{C} contains a maximum hierarchy; in particular, it contains all the trivial clusters, and so (P1) holds. For (P2), suppose that \mathcal{C}' is a collection of $k \geq 3$ disjoint subsets of X , each an element of \mathcal{C} , and $\cup \mathcal{C}' \in \mathcal{C}$. Then $\mathcal{H} = \mathcal{C}' \cup \{\cup \mathcal{C}'\}$ is a hierarchy. Let $\mathcal{H}' \subseteq \mathcal{C}$ be a maximum hierarchy on X that contains \mathcal{H} (this exists, since \mathcal{C} is bureaucratic) and let C be a minimal subset of X in \mathcal{H}' that contains the union of at least two elements of \mathcal{C}' . Since \mathcal{H}' is a maximum hierarchy, and $\cup \mathcal{C}' \in \mathcal{H}'$, C is precisely the union of exactly two elements of \mathcal{C}' ; since $C \in \mathcal{H}' \subseteq \mathcal{C}$, this establishes (P2).

Conversely, suppose that a collection \mathcal{C} of subsets of X satisfies (P1) and (P2), and that $\mathcal{H} \subseteq \mathcal{C}$ is a maximal hierarchy which is contained within \mathcal{C} . Suppose that \mathcal{H} is not maximum (we will derive a contradiction). Then \mathcal{H} contains a set C that is the disjoint union of $k \geq 3$ maximal proper subsets A_1, \dots, A_k , each belonging to \mathcal{H} (and thereby \mathcal{C}). Applying (P2) to $\mathcal{C}' = \{A_1, \dots, A_k\}$, there exist two sets, say A_i, A_j for which $A_i \cup A_j \in \mathcal{C}$. So, if we let $\mathcal{H}' = \mathcal{H} \cup \{A_i \cup A_j\}$, then we obtain a larger hierarchy containing \mathcal{H} that is still contained within \mathcal{C} , which is a contradiction. This completes the proof. \square

2. Examples of bureaucracies

We have mentioned two extreme cases of bureaucracies, namely the set of clusters of a rooted binary tree having leaf set X , and the full power set $\mathcal{P}(X)$. Here are some further examples.

- (1) The set of intervals of $[n] = \{1, 2, \dots, n\}$ is a bureaucracy where an *interval* is a set $[i, j] = \{k : i \leq k \leq j\}$, $1 \leq i \leq j \leq n$.

Proof. Let \mathcal{C} be the set of intervals of $[n]$. Then \mathcal{C} contains the trivial clusters. Also, a disjoint collection I_1, \dots, I_k , $k > 2$, of intervals has union an interval if and only if every element of $[n]$ between $\min \cup I_j$ and $\max \cup I_j$ lies in (exactly) one interval, in which case the union of any pair of consecutive intervals is an interval, so (P2) holds. By Theorem 2, \mathcal{C} is bureaucratic. \square

Similarly, if we order the elements of X in any fashion, we can define the set of *intervals on X* for that ordering by this construction (associating x_i with i), and can thus obtain a bureaucracy.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات