Applied Mathematics Letters 22 (2009) 1496–1499

Contents lists available at ScienceDirect

Applied Mathematics Letters

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

Phylogenetic diversity and the maximum coverage problem

Vincent Moulton [a,*], Andreas Spillner [b]

[a] School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK

[b] Department of Mathematics and Computer Science, Greifswald University, 17489 Greifswald, Germany

ARTICLE INFO

Article history:
Received 23 March 2009
Accepted 23 March 2009

Keywords:
Phylogenetic diversity
Greedoid
Hypergraph
Maximum coverage

ABSTRACT

For a weighted hypergraph (H, ω), with vertex set X, edge set E, and weighting $\omega : E \to \mathbb{R}_{\geq 0}$, the maximum coverage problem is to find a k-element subset $Y \subseteq X$ that maximizes the total weight of those edges that have non-empty intersection with Y among all k-element subsets of X. Such a subset Y is called optimal. Recently, within the field of phylogenetics it has been shown that for certain weighted hypergraphs coming from phylogenetic trees the collection of optimal subsets of X forms a so-called strong greedoid. We call hypergraphs having this latter property *strongly greedy*. In this note we characterize the r-uniform hypergraphs H with unit edge weights that are strongly greedy in the case where r is a prime number.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Recall that a hypergraph $H = (X, E)$ consists of a finite non-empty set X of *vertices* and a collection E of subsets of X, called *hyperedges* or *edges* (cf. [1]). Let $\sim$ be the binary relation on X defined by taking the transitive closure of the relation consisting of those pairs (x, y), $x, y \in X$, for which there exists some edge $e \in E$ with $\{x, y\} \subseteq e$ or $x = y$. The *connected components* of H are the equivalence classes of $\sim$. A *weighted* hypergraph (H, ω) is a hypergraph $H = (X, E)$ together with a map ω that assigns a non-negative real number $\omega(e)$ to every edge $e \in E$. The *score*, $\sigma(Y) = \sigma_{(H,\omega)}(Y)$, of any subset $Y \subseteq X$ relative to (H, ω) is the total weight of those edges that have a non-empty intersection with Y.

In the *maximum coverage* problem, one aims to find subsets $Y \subseteq X$ relative to a weighted hypergraph $(H = (X, E), \omega)$ that are *optimal*, that is, subsets Y of X having maximum score amongst all $|Y|$-element subsets of X [2]. The maximum coverage problem is a well-studied problem in combinatorial optimization and appears in various applications, for example, in circuit layout, scheduling and facility location (see e.g. [3]). Various algorithms have been devised for its solution, among them a simple greedy algorithm that starts with an optimal subset of size 1, and, at each step, adds a vertex such that the increase in the score of the resulting subset is maximum. Although even very restricted versions of the maximum coverage problem are NP-hard [4], in [2] it is shown that this greedy algorithm is guaranteed to yield a solution that is within $(1 - 1/e)$ of the optimal score, and it appears unlikely that any algorithm can do significantly better [5].

Recently, in the field of phylogenetics [6], studies have appeared on applying the greedy algorithm to a special class of hypergraphs called *hierarchies*, hypergraphs $H = (X, E)$ for which $e_1 \cap e_2 \in \{\emptyset, e_1, e_2\}$ holds for all $e_1, e_2 \in E$. Weighted hierarchies are of interest in phylogenetics as they correspond to edge-weighted rooted phylogenetic trees (see [6, p. 52] for more details, and Fig. 1(a) for a simple example). Moreover, in this setting, the score $\sigma(Y)$ of any subset $Y \subseteq X$ is known as the *phylogenetic diversity* of Y (which is simply the length of the subtree spanned by the elements in Y and the root in the corresponding tree — see e.g. Fig. 1(b)), a quantity that has applications in biodiversity conservation [7].

* Corresponding author.
E-mail addresses: vincent.moulton@cmp.uea.ac.uk (V. Moulton), anspillner@googlemail.com (A. Spillner).

0893-9659/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aml.2009.03.017

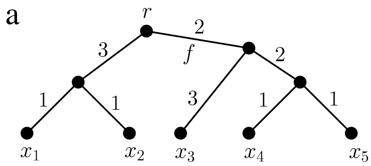

Fig. 1. (a) A rooted phylogenetic tree T corresponding to a weighted hierarchy on the set $\{x_1, \ldots, x_5\}$ with root r. Each edge of T corresponds to an edge in the hierarchy (e.g. f corresponds to $\{x_3, x_4, x_5\}$, having weight 2). (b) The phylogenetic diversity of the set $\{x_1, x_3, x_4\}$ equals 12, the total weight of the induced rooted subtree.

In [8,9] it is shown that the greedy algorithm always yields optimal solutions to the maximum coverage problem for weighted hierarchies, and in [10, Theorem 3.2] it is shown that the collection $\mathcal{O}_{(H,\omega)}$ of optimal subsets of X (where subsets can have any size between 0 and $|X|$) even forms a *strong greedoid*, i.e. it satisfies the following conditions:

(S1) For every $Y \in \mathcal{O}_{(H,\omega)}$, $Y \neq \emptyset$, there exists at least one $y \in Y$ such that $(Y \setminus \{y\}) \in \mathcal{O}_{(H,\omega)}$.
(S2) For every $Y_1, Y_2 \in \mathcal{O}_{(H,\omega)}$ such that $|Y_1| + 1 = |Y_2|$ there exists at least one $y \in Y_2 \setminus Y_1$ with the property that $(Y_1 \cup \{y\}) \in \mathcal{O}_{(H,\omega)}$ and $(Y_2 \setminus \{y\}) \in \mathcal{O}_{(H,\omega)}$.

Note that strong greedoids were introduced to provide a framework for optimization problems where the greedy algorithm is compatible with the structure of the optimal sets [11–13].

Intriguingly, it is not hard to show using [10, Theorem 3.2] that hierarchies H are in fact characterized by the property that $\mathcal{O}_{(H,\omega)}$ is a strong greedoid for *any* weighting ω of H. Motivated by this fact as well as recent extensions of phylogenetic diversity to non-hierarchical structures [14,15], in this note we study weighted hypergraphs (H, ω) for which $\mathcal{O}_{(H,\omega)}$ is a strong greedoid, which we call *strongly greedy* hypergraphs for short. In particular, although it appears to be an interesting but difficult problem to characterize strongly greedy hypergraphs in general, restricting our attention to r-uniform hypergraphs H (hypergraphs in which every edge has cardinality $r \in \mathbb{N}$ [1, p. 3]) with *unit* edge weights, and noting that a *clique* in H is a subset $C \subseteq X$ such that every r-element subset of C is an edge of H, we present a proof for the following result.

Theorem 1.1. *Let $H = (X, E)$ be an r-uniform hypergraph. If every connected component of H is a clique, then, denoting by* $\mathbf{1}$ *the weight function that assigns weight 1 to every edge in H, the hypergraph $(H, \mathbf{1})$ is strongly greedy. Moreover, if r is prime and $(H, \mathbf{1})$ is strongly greedy, then every connected component of H is a clique.*

Before presenting the proof of this result, we close this section by noting that for r not prime, there exist r-uniform hypergraphs that are strongly greedy in which not every connected component is a clique. Indeed, it is straight-forward to check that the 4-uniform hypergraph $H = (X, E)$ with $X = \{a_1, \ldots, a_4, b_1, \ldots, b_4\}$ and E consisting of the following edges:

$$\{a_1, a_2, b_1, b_2\}\ \{a_2, a_3, b_2, b_3\}\ \{a_1, a_2, b_3, b_4\}\ \{a_2, a_3, b_1, b_4\}\ \{a_1, a_2, a_3, a_4\}$$
$$\{a_1, a_3, b_1, b_3\}\ \{a_2, a_4, b_2, b_4\}\ \{a_1, a_3, b_2, b_4\}\ \{a_2, a_4, b_1, b_3\}\ \{b_1, b_2, b_3, b_4\}$$
$$\{a_1, a_4, b_1, b_4\}\ \{a_3, a_4, b_3, b_4\}\ \{a_1, a_4, b_2, b_3\}\ \{a_3, a_4, b_1, b_2\}$$

is strongly greedy, although it has only one connected component, namely X, which is not a clique since H has less then $\binom{8}{4} = 70$ edges. Even so, this hypergraph is still highly symmetric, and so it could be of interest to understand which highly symmetric r-uniform hypergraphs yield strong greedoids in the case where r is not a prime.

2. Proof of Theorem 1.1

First note that, for any hypergraph H, $\mathcal{O}_{(H,\mathbf{1})}$ contains at least one set of size k for every $k \in \{0, 1, \ldots, |X|\}$. Hence, if $\mathcal{O}_{(H,\mathbf{1})}$ satisfies (S2), then it also satisfies (S1), and so $(H, \mathbf{1})$ is strongly greedy if and only if $\mathcal{O}_{(H,\mathbf{1})}$ satisfies (S2).

Now suppose that H is a hypergraph in which every connected component of H is a clique. We show that $(H, \mathbf{1})$ is strongly greedy. Let $C_1, \ldots, C_l$ be the connected components of H. Fix an arbitrary $k \in \{0, 1, \ldots, |X| - 1\}$. Let $A, B \in \mathcal{O} = \mathcal{O}_{(H,\mathbf{1})}$, $|A| = |B| - 1 = k$. For $1 \leq i \leq l$ define $A_i = C_i \cap A$ and $B_i = C_i \cap B$. Fix an arbitrary $x \in B \setminus A$ and let j be such that $x \in C_j$. Define $A' = A \cup \{x\}$ and $B' = B \setminus \{x\}$.

Since $A, B \in \mathcal{O}$, we must have $\sigma(B') \leq \sigma(A)$ and $\sigma(A') \leq \sigma(B)$. Moreover, since

$$\sigma(B') = \sigma(B) - \binom{|C_j \setminus B_j|}{r-1} \quad \text{and} \quad \sigma(A') = \sigma(A) + \binom{|C_j \setminus A_j| - 1}{r-1}$$

both clearly hold, it follows that

$$\binom{|C_j \setminus A_j| - 1}{r-1} \leq \sigma(B) - \sigma(A) \leq \binom{|C_j \setminus B_j|}{r-1}.$$

Download English Version:

https://daneshyari.com/en/article/1710030

Download Persian Version:

https://daneshyari.com/article/1710030

Daneshyari.com