THE MINIMUM CONFLICT-FREE ROW SPLIT PROBLEM

A model for reconstructing perfect phylogenies from mixed tumor samples.

A. Hujdurović, Edin Husić, M. Milanič (University of Primorska); R. Rizzi (University of Verona); A. Tomescu (University of Helsinki)

Motivation

- cancer research,
- reconstruct perfect phylogeny tree from a given matrix.

Def 1. *Given $M \in \{0,1\}^{m\times n}$, columns i and j are in* conflict *if $\exists$ rows r_1, r_2, r_3 s.t.,*

$$\begin{array}{c} \\ r_1 \\ r_2 \\ r_3 \end{array} \begin{array}{c} \begin{array}{cc} i & j \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \end{array}. \qquad M \text{ is conflict-free} \Leftrightarrow \nexists \text{ conflicts in } M.$$

- M is conflict-free $\Leftrightarrow$ admits perfect phylogeny tree.

Def 2. *Let $M \in \{0,1\}^{m\times n}$ with the rows $r_1, r_2, \ldots, r_m$. $M' \in \{0,1\}^{m'\times n}$ is a* row split *of M if $\exists$ a partition of rows of M' into $R_1, R_2, \ldots R_m$ s.t. $\forall i \in \{1,\ldots,m\}$, r_i is the bitwise OR of the vectors in R_i.*

$\gamma(M)$ = the minimum number of rows in a conflict-free row split of M.

MINIMUM CONFLICT-FREE ROW SPLIT
Input: A binary matrix M.
Task: Compute $\gamma(M)$.

Def 3. *A* branching *of an acyclic digraph $D = (V, A)$ is a subset of arcs B such that (V, B) is a digraph in which $\forall v \in V$ there is at most one arc leaving v.*

For $M \in \{0,1\}^{m\times n}$, the *containment digraph* of M is $D_M = (V, A)$ with

$$\begin{aligned} V &= \{\mathrm{supp}(c) : c \in C_M\}, \\ A &= \{(v, v') : v, v' \in V \wedge v \subset v'\}. \end{aligned}$$

For $X \subseteq A$, and $v \in V$, we say that $r \in v$ is *uncovered* in v with respect to X if $\forall (v', v) \in X, r \notin v'$. For row r_i let

$$U(r_i) = \{(r_i, v) : r_i \text{ is uncovered in } v \in V\}$$

$$U(X) = \cup_{i=1}^{m} U(r_i)$$

$\beta(M)$ = the minimum size of $U(B)$ over all branchings B of D_M.

MINIMUM UNCOVERING BRANCHING
Input: A binary matrix M.
Task: Compute $\beta(M)$.

Equivalence

For a branching B of D_M, the *B-split* of M is M^B with rows indexed by the elements of $U(B)$, and columns $c'_1, \ldots, c'_n$, as follows. We set:

$$M^B_{(r,v),j} = \begin{cases} 1, & \text{if there exists a } v - v_j \text{ directed path in } (V, B); \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 1. *B-split of M is a conflict-free row split of M, the number of rows in M^B is exactly $|U(B)|$.*

For a conflict-free row split $\exists$ a corresponding branching of its containment digraph.

- integer program
- APX-hardness
- exact algorithm

Supermodular function + partition matroid

Let A be a finite set. We say that a set function $f : \mathcal{P}(A) \to \mathbb{R}$ is

supermodular: if $\forall S, T \subseteq A$ it holds $f(S \cup T) + f(S \cap T) \geq f(S) + f(T)$,
monotone decreasing: if $\forall S \subseteq T$ it holds $f(S) \geq f(T)$.

A pair $(A, \mathcal{I})$ is called a *matroid* if a collection $\mathcal{I} \subseteq \mathcal{P}(A)$ satisfies

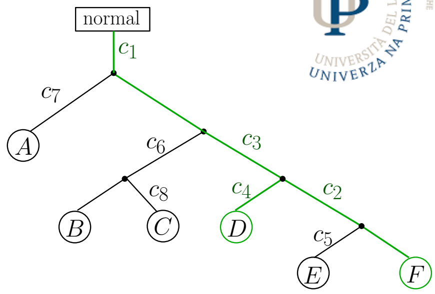
- $\emptyset \in \mathcal{I}, Y \subseteq X \in \mathcal{I} \Rightarrow Y \in \mathcal{I}$
- $X, Y \in \mathcal{I}$, $|X| < |Y| \Rightarrow \exists s \in Y \setminus X$ such that $X \cup \{s\} \in \mathcal{I}$.

Let $A_1, A_2, \ldots, A_n$ a partition of A and $k_1, k_2, \ldots, k_n \in \mathbb{N}$. Then S together with $\mathcal{I} = \{X \subseteq A : |X \cap A_i| \leq k_i$ for all $i \in \{1, 2, \ldots, n\}\}$ is a partition matroid.

Proposition 2. *The* MINIMUM UNCOVERING BRANCHING *problem can be formulated as a special case of the problem of minimizing a monotone decreasing supermodular function subject to a partition matroid constraint.*

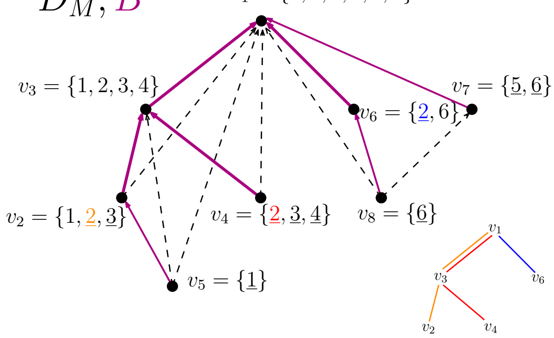
- $f(X) = |U(X)|$,
- A_i = set of edges leaving v_i and $k_i = 1$ for all i.

If we are interested in maximizing the number of "covered" pairs our task is to maximize a monotone increasing submodular function under a matroid constraint. For this problem, several $(1 - 1/e)$-approximation algorithms are known.



$$\begin{array}{r} M \\ \{E,F\} \sim r_1 \\ \{B,D,F\} \sim r_2 \\ \{D,F\} \sim r_3 \\ \{D\} \sim r_4 \\ \{A\} \sim r_5 \\ \{A,C,B\} \sim r_6 \end{array} \begin{array}{c} \begin{array}{cccccccc} c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

D_M, B



$$\begin{array}{r} M^B \\ (r_1,v_5) \\ (r_2,v_2) \\ (r_2,v_4) \\ (r_2,v_6) \\ (r_3,v_2) \\ (r_3,v_4) \\ (r_4,v_4) \\ (r_5,v_7) \\ (r_6,v_7) \\ (r_6,v_8) \end{array} \begin{array}{c} \begin{array}{cccccccc} c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ \hline 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \end{array}$$