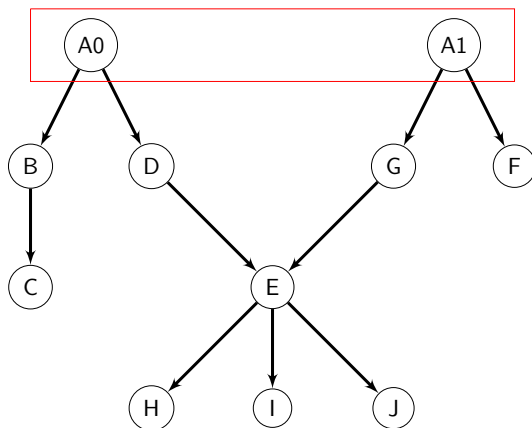
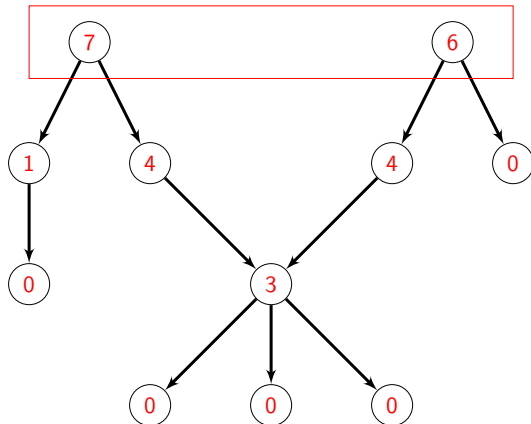


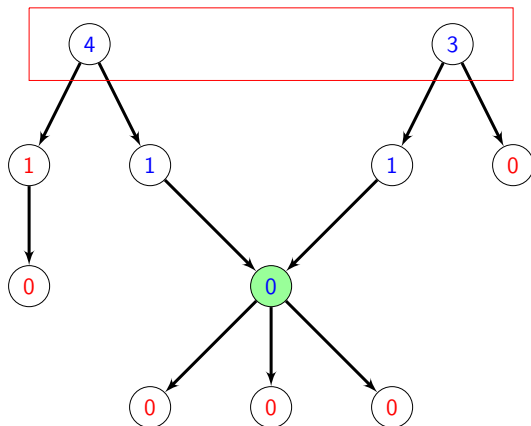
Spark Job Tree – Labels



Spark Job Tree – Operations per Node



Spark Job Tree – Caching E



Problem Formulation

$$\min \left\{ \sum_{i \in (S \subseteq V)} (c_i) : \mathbf{x} \in \mathcal{R} \cap \mathbb{B}^{|V|} \right\} \quad (1)$$

$$c_i = |N_+(i)| + \sum_{j \in N_+(i)} (1 - x_j) \cdot c_j, \quad \forall i \in V \quad (2)$$

$$\sum_{i \in V} x_i = k \quad (3)$$

$$0 \leq x_i \leq 1, \quad \forall i \in V \quad (4)$$