Minimum k-labeled triangle forest: complexity and greedy approximations

Santiago Valdés Ravelo[1], **Cristina G. Fernandes**[2]

[1]Instituto de Computação – Universidade Estadual de Campinas
Campinas, Brasil ,

[2]Instituto de Matemática, Estatística e Ciência da Computação – Universidade de São Paulo
São Paulo, Brasil.

ravelo@unicamp.br, cris@ime.usp.br

Abstract. *We study the k-LABELED TRIANGLE FOREST problem, where, given a triangle-forest graph with edges colored from a set C, the goal is to select k colors to minimize the number of components. We prove the problem is NP-hard, provide a tight $\frac{5}{3}$-approximation via a greedy algorithm that selects the most frequent colors, and give a $(1+2/e)$-approximation using a second greedy strategy that iteratively chooses the color maximizing the reduction in components.*

1. Introduction

Many systems involve labeled connections, where only a limited subset can be active at once. Examples might include network design, resource allocation, and logistics [4].

These scenarios can be modeled as a graph $G = (V, E)$ with edges colored by $\ell \colon E \to C$, where the goal is to select at most k colors so that removing all other edges minimizes the number of connected components. A particular case is when G is a triangle forest.

The k-LABELED TRIANGLE FOREST problem is defined as follows: given a triangle-forest graph $G = (V, E)$ (i.e., each connected component is a triangle), a set of colors C, a color assignment $\ell \colon E \to C$, and an integer k, find a subset $S \subseteq C$ with $|S| = k$ such that $G[S]$ has the minimum number of components. Here, for $S \subseteq C$, $E(S) \subseteq E$ denotes the edges with colors in S, $G[S]$ is the spanning subgraph on $E(S)$, and $g(S)$ denotes the number of useful edges (edges whose addition decreases the number of components). If $n = |V|$, the number of components in $G[S]$ is

$$\mathrm{COMP}(S) = n - g(S).$$

In this work, we prove that k-LABELED TRIANGLE FOREST is NP-hard and present two greedy approximation algorithms for the problem.

2. Complexity

Theorem 1 *The decision version of k-LABELED TRIANGLE FOREST is NP-complete.*

Proof: Let $I = \langle X, \mathcal{T} \rangle$ be an instance of EXACT COVER BY 3-SETS (X3C) [1]. We construct an instance $I' = \langle G, C, \ell, k \rangle$ of k-LABELED TRIANGLE FOREST as follows:

- Let $C' = \{c_\beta\} \cup \{c_t \mid t \in \mathcal{T}\}$, where c_β is a special color.

- For each element $x \in X$, let $\mathcal{T}_x = \{t \in \mathcal{T} \mid x \in t\}$. For every pair (t, t') with $t, t' \in \mathcal{T}_x$ and $t \neq t'$, we add a triangle to G whose edges are labeled c_t, $c_{t'}$, and c_β.
- For each $t \in \mathcal{T}$, let $m_t = |E(\{c_t\})|$, and define $m = \max_{t \in \mathcal{T}} m_t$. For each t, we add $m - m_t$ triangles to G, each containing one edge labeled c_t, one edge labeled c_β, and one edge with a fresh auxiliary color. This ensures that exactly m edges in G are labeled c_t.
  Finally, we add one additional triangle containing one edge labeled c_β and two edges with fresh auxiliary colors.
- Let C be the union of C' with all auxiliary colors introduced above, and let ℓ denote the resulting labeling function.
- Set $k = \frac{|X|}{3} + 1$.

  The construction clearly runs in polynomial time. We now prove its correctness.

  We show that I has an exact cover if and only if I' admits a solution S such that

$$\mathrm{COMP}(S) \leq n - \left(\frac{|X|}{3} \cdot m + |E(\{c_\beta\})| \right).$$

Equivalently,

$$g(S) \geq \frac{|X|}{3} \cdot m + |E(\{c_\beta\})|.$$

($\Rightarrow$) Let $T \subseteq \mathcal{T}$ be an exact cover for I. Define

$$S = \{c_t \mid t \in T\} \cup \{c_\beta\}.$$

Since T is an exact cover, no two triples in T share an element. Hence, for every triangle corresponding to a pair (t, t') of triples sharing an element, at most one of the labels $c_t, c_{t'}$ belongs to S. Moreover, since S contains no auxiliary colors, no triangle induces a cycle in $G[S]$. Therefore, $G[S]$ is a forest, and every edge in $E(G[S])$ is useful, implying that $g(S) = |E(G[S])|$.
Furthermore, $|S| = \frac{|X|}{3} + 1 = k$, so S is feasible. By construction, each selected color c_t contributes exactly m edges, and all edges labeled c_β are included in $E(G[S])$. Hence,

$$g(S) = |E(G[S])| = \frac{|X|}{3} \cdot m + |E(\{c_\beta\})|.$$

($\Leftarrow$) Let $S \subseteq C$ with $|S| \leq k$ and satisfying

$$g(S) \geq \frac{|X|}{3} \cdot m + |E(\{c_\beta\})|.$$

First, we show that $c_\beta \in S$. Otherwise, since each color other than c_β contributes at most m edges,

$$g(S) \leq k \cdot m = \left(\frac{|X|}{3} + 1 \right) m.$$

Since $|E(\{c_\beta\})| > m$, this contradicts the assumed bound. Hence, $c_\beta \in S$.
Next, to achieve the required number of edges, S must include exactly $\frac{|X|}{3}$ colors of the form c_t, each contributing all its m edges; otherwise, $g(S)$ would be strictly smaller than $\frac{|X|}{3} \cdot m + |E(\{c_\beta\})|$, contradicting the bound. Let $T = \{t \in \mathcal{T} \mid c_t \in S\}$, then,

$$|T| = \frac{|X|}{3}.$$

Finally, observe that $G[S]$ must be acyclic. If $G[S]$ contained a cycle, then at least one edge would not be useful, implying $g(S) < |E(G[S])|$, and thus $g(S) < \frac{|X|}{3} \cdot m + |E(\{c_\beta\})|$, a contradiction. Therefore, no triangle can have all three edges selected. In particular, for any triangle corresponding to a pair (t, t'), both c_t and $c_{t'}$ cannot belong to S, since together with c_β they would form a cycle. Hence, no two triples in T share an element.
Thus, T is a family of $\frac{|X|}{3}$ pairwise disjoint triples, i.e., an exact cover of X.

Therefore, the reduction is correct. Since X3C is NP-complete [1], k-LABELED TRIANGLE FOREST is NP-hard. □

3. Top k-colors algorithm

In this section, we consider the greedy algorithm that selects the k most frequent colors in G, and prove that it yields a $\frac{5}{3}$-approximation. We also show that this approximation factor is tight.

Theorem 2 *Let $\langle G, C, \ell, k\rangle$ be an instance of the k-*LABELED TRIANGLE FOREST *problem. Selecting the k most frequent colors yields a $\frac{5}{3}$-approximation.*

Proof: Let $S \subseteq C$ be the set of colors selected by the algorithm (i.e., the k most frequent colors), and let $O \subseteq C$ be the set of colors in an optimal solution. Then,

$$\mathrm{COMP}(S) = n - g(S) = n - g(S) - g(O) + g(O) = \mathrm{COMP}(O) + g(O) - g(S).$$

Since each triangle contains at three edges, and at most one of them can be non-useful, $g(S) \geq \frac{2}{3}|E(S)|$. Hence,

$$\mathrm{COMP}(S) \leq \mathrm{COMP}(O) + g(O) - \frac{2}{3}|E(S)|.$$

By the greedy choice, $|E(S)| \geq |E(O)| \geq g(O)$ and, therefore,

$$\mathrm{COMP}(S) \leq \mathrm{COMP}(O) + \frac{1}{3}g(O).$$

Since the triangles are disjoint, each component can contain at most two useful edges in any solution. Thus, $\mathrm{COMP}(O) \geq \frac{1}{2}g(O)$, which implies

$$\mathrm{COMP}(S) \leq \mathrm{COMP}(O) + \frac{2}{3}\mathrm{COMP}(O) = \frac{5}{3}\mathrm{COMP}(O).$$

□

3.1. Tightness of the approximation factor

We now show that the $\frac{5}{3}$-approximation factor is tight for this algorithm.

For a positive integer m, consider an instance of the problem with five colors $C = \{a, b, c, d, e\}$, $k = 3$, and two types of triangles, as in Figure 1, with G being a collection of triangles as follows:

- $2m + 1$ triangles of type 1: each with one edge of color a, one of color b, and one of color c;

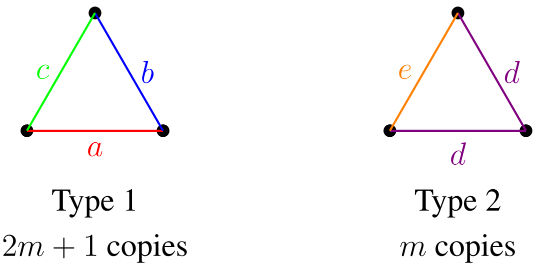

Figure 1. Triangles used for the tightness instance.

- m triangles of type 2: each with two edges of color d and one edge of color e.

The greedy algorithm selects colors $\{a, b, c\}$, producing a solution S with $2m + 1$ components from triangles of type 1 and $3m$ components from triangles of type 2. In contrast, an optimal solution selects colors $\{a, b, d\}$, producing $2m + 1$ components from triangles of type 1 and m components from triangles of type 2. Thus,

$$\lim_{m\to\infty} \frac{\mathrm{COMP}(S)}{\mathrm{COMP}(O)} = \lim_{m\to\infty} \frac{5m+1}{3m+1} = \frac{5}{3}.$$

Therefore, there exists a family of instances for which the greedy algorithm produces a solution whose approximation ratio is arbitrarily close to $\frac{5}{3}$, proving that the asymptotically analysis is tight.

Theorem 3 *The $\frac{5}{3}$-approximation factor for selecting the k most frequent colors is tight.*

4. Incremental greedy algorithm

In this section, we analyze a second greedy algorithm. Instead of selecting the k most frequent colors, this algorithm incrementally selects the color that maximizes the reduction in the number of components. That is, the algorithm runs for k iterations, and at each iteration selects the color that produces the largest number of useful edges.

Theorem 4 *The incremental greedy algorithm yields a $\left(1 + \frac{2}{e}\right)$-approximation.*

Proof: Let $S \subseteq C$ be the set of colors selected by the algorithm, and let $O \subseteq C$ be the set of colors in an optimal solution. Since g is a submodular monotonic function [2, 3, 5]:

$$g(S) \geq \left(1 - \frac{1}{e}\right) g(O).$$

As seen before, $\mathrm{COMP}(O) \geq \frac{1}{2} g(O)$ and,

$$\mathrm{COMP}(S) = \mathrm{COMP}(O) + g(O) - g(S) \leq \mathrm{COMP}(O) + \frac{1}{e} g(O) = \left(1 + \frac{2}{e}\right) \mathrm{COMP}(O).$$

$\square$

5. Conclusion

We show that k-LABELED TRIANGLE FOREST is NP-hard, provide a tight $\frac{5}{3}$-approximation via the k most frequent colors, and a $(1 + 2/e)$-approximation exploiting the submodularity of the usable-edge function. In future works we will study generalizations of triangle forests and classes of instances for which the problem is tractable.

References

[1] R. Karp. Reducibility among combinatorial problems. volume 40, pages 85–103, 01 1972.

[2] G. L. Nemhauser and L. A. Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of Operations Research*, 3(3):177–188, 1978.

[3] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, 14:265–294, 1978.

[4] T. F. Pinheiro, S. V. Ravelo, and L. S. Buriol. A fix-and-optimize matheuristic for the k-labelled spanning forest problem. In *2022 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8, 2022.

[5] D. P. Williamson and D. B. Shmoys. *The Design of Approximation Algorithms.* Cambridge University Press, 2011.