Theoretical Computer Science 648 (2016) 72–83

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

The label cut problem with respect to path length and label frequency ☆

Peng Zhang [a], Bin Fu [b,*]

[a] School of Computer Science and Technology, Shandong University, Jinan 250101, China
[b] Department of Computer Science, University of Texas-Pan American, Edinburg, TX 78539, USA

ARTICLE INFO

Article history:
Received 16 January 2016
Received in revised form 17 July 2016
Accepted 2 August 2016
Available online 9 August 2016
Communicated by G. Ausiello

Keywords:
Label *s-t* Cut
Global Label Cut
Parameterized algorithm
Approximation algorithm
Computational complexity

ABSTRACT

Given a graph with labels defined on edges and a source-sink pair (s, t), the Label *s-t* Cut problem asks for a minimum number of labels such that the removal of edges with these labels disconnects s and t. Similarly, the Global Label Cut problem asks for a minimum number of labels to disconnect G itself. For these two problems, we identify two useful parameters, i.e., $l_{\max}$, the maximum length of any *s-t* path (only applies to Label *s-t* Cut), and $f_{\max}$, the maximum number of appearances of any label in the graph (applies to the two problems). We show that $l_{\max} = 2$ and $f_{\max} = 2$ are two complexity thresholds for Label *s-t* Cut. Furthermore, we give (i) an $O^*(c^k)$ time parameterized algorithm for Label *s-t* Cut with $l_{\max}$ bounded from above, where parameter k is the number of labels in a solution, and c is a constant with $l_{\max} - 1 < c < l_{\max}$, (ii) a combinatorial $l_{\max}$-approximation algorithm for Label *s-t* Cut, and (iii) a polynomial time exact algorithm for Global Label Cut with $f_{\max}$ bounded from above.

1. Introduction

The Label *s-t* Cut problem is an edge-classification Min *s-t* Cut problem defined as follows.

Label *s-t* Cut

Instance: We are given a (directed or undirected) graph $G = (V, E)$, a source-sink pair (s, t), and a label set $L = \{\ell_1, \ell_2, \ldots, \ell_q\}$. Each edge $e \in E$ has a label $\ell(e) \in L$.

Goal: A *label s-t cut* $L' \subseteq L$ is a subset of labels such that the removal of all edges with these labels from G disconnects s and t. The goal of the problem is to find a label *s-t* cut of the minimum size.

Recently, the Label *s-t* Cut problem attracted a good deal of attention from researchers (see [7,8,15,16,25,26,30]). As introduced in [30], this problem comes from system security, in particular from intrusion detection and from the generation and analysis of attack graphs [16,25]. The reader who is interested in the origin of the problem is referred to [30] for a brief introduction. Since Label *s-t* Cut is NP-hard [16], people usually seek approximation algorithms [7,26,30] and parameterized algorithms for this problem. The crux is that in Label *s-t* Cut many edges may share the same label. Otherwise (each edge has a distinct label), Label *s-t* Cut degenerates to the well-known Min *s-t* Cut problem (to find a cut of minimum capacity that disconnects s and t in the input graph).

☆ A preliminary version of partial results in this paper appeared in the Proceedings of the 11th Annual Conference on Theory and Applications of Models of Computation (TAMC 2014) [29].

* Corresponding author.
E-mail addresses: algzhang@sdu.edu.cn (P. Zhang), bfu@utpa.edu (B. Fu).

http://dx.doi.org/10.1016/j.tcs.2016.08.006

Besides Label s-t Cut, there are still many network optimization problems which have been studied in edge-labeled graphs, such as the Minimum Spanning Tree problem [4,20,6], the Minimum Steiner Tree problem [5], the Max Flow problem [12], the Shortest Path problem [13], the Perfect Matching problem [21], and the Traveling Salesman problem [28].

Moreover, it is natural to define the Global Label Cut problem.

Global Label Cut

Instance: In the problem we are given an undirected graph $G = (V, E)$ and a label set $L = \{\ell_1, \ell_2, \dots, \ell_q\}$. Each edge $e \in E$ has a label $\ell(e) \in L$.

Goal: A *global label cut* $L' \subseteq L$ is a subset of labels such that G is disconnected after removing all edges with labels in L'. The goal of the problem is to find a global label cut of the minimum size.

Global Min Cut is the global version of the Label s-t Cut problem. This is just like that the Global Min Cut problem (to find a cut of minimum capacity that disconnects the input graph) is the global version of the Min s-t Cut problem.

Notations. Throughout the paper, the symbol ℓ denotes a label in L, and also denotes a mapping from edges to labels when we write, e.g., $\ell(e)$. We do not introduce more symbols to distinguish between these two cases. A similar thing happens to the symbol L: It denotes the input label set, and it also denotes a mapping from 2^E to 2^L. In the latter case, we use $L(E')$ to denote the set $\{\ell(e) : e \in E'\}$ for an edge subset E'.

1.1. Our results

In this paper, we identify two parameters of the Label s-t Cut problem and the Global Label Cut problem, and show how they affect the complexity of the problems. In several cases, we get complexity thresholds and efficient algorithms.

First consider the parameter $l_{\max}$. This parameter only applies to the Label s-t Cut problem.

Definition 1.1 *($l_{\max}$).* Given a graph $G = (V, E)$ with source $s \in V$ and sink $t \in V$, the *maximum s-t length* $l_{\max}$ is the length of a longest simple s-t path in terms of the number of edges.

We prove that Label s-t Cut is NP-hard even when restricted to instances with $l_{\max} = 2$ (Theorem 2.1). Since the problem with $l_{\max} = 1$ is trivial, this shows that $l_{\max} = 2$ is a complexity threshold of the Label s-t Cut problem. (When $l_{\max} = 1$, there is only one path between s and t, whose length is 1. This is a trivial case.) Then, we show that when $l_{\max}$ is upper bounded by a constant and the Label s-t Cut problem is parameterized by k, i.e., the number of labels in a solution, the problem is fixed parameter tractable and is solvable in time $O^*(c^k)$ (Theorem 3.1), where $c \in (l_{\max} - 1, l_{\max})$ is a constant. (Recall that in the $O^*()$ notation we omit any factor polynomial in the input length.) This is compared with the $O^*(l_{\max}^k)$-time parameterized algorithm for the problem using the standard depth-bounded search tree technique [23]. We point out that the proof of Theorem 3.1 is considerably involved and this theorem is one of our main results. Moreover, we give an $l_{\max}$-approximation algorithm for the Label s-t Cut problem (Theorem 4.1). The algorithm is purely combinatorial, simpler and faster than the LP-rounding algorithm in [26], and has a primal-dual explanation.

Then consider the parameter $f_{\max}$. This parameter applies to the Label s-t Cut problem and the Global Label Cut problem.

Definition 1.2 *($f_{\max}$).* Given a graph G with labels in L defined on edges, the *label frequency* $f(\ell)$ of label $\ell \in L$ is the number of edges in G whose label is ℓ. The *maximum label frequency* $f_{\max}$ is the maximum of $f(\ell)$ over all labels.

We prove that Label s-t Cut is NP-hard when $f_{\max} = 2$ (Theorem 2.2). Since when $f_{\max} = 1$ Label s-t Cut degenerates to the Min s-t Cut problem, which is polynomial time solvable [1], this shows that $f_{\max} = 2$ is a complexity threshold of the Label s-t Cut problem. Furthermore, we prove that when $f_{\max}$ is upper bounded by a constant, the Global Label Cut problem is polynomial time solvable (Theorem 3.2).

Of course, besides the two parameters considered above, for the Label s-t Cut problem there are still several parameters which are potential candidates for the investigation, such as the treewidth of the input graph, the number of edges in the optimal label s-t cut, etc. The reason why we study Label s-t Cut with respect to path length and label frequency is actually already shown by our results. First, these two parameters dramatically affect the computational complexity of the problem. Second, in several cases we obtain good algorithms for the problem using these two parameters.

A preliminary version of partial results in this paper appeared in the Proceedings of the 11th Annual Conference on Theory and Applications of Models of Computation (TAMC 2014) [29]. This paper is the full and extended version of [29]. More importantly, two algorithmic results in this paper, i.e., Theorem 3.1 and Theorem 3.3, are thoroughly new; we obtained them after [29] was published. In particular, Theorem 3.1 is the most important algorithmic result in this paper. However, according to the "first complexity and then algorithms" order, we present it in Section 3, after the hardness results in Section 2.

1.2. Related work

It is well-known that Set Cover has a greedy polynomial time $(1 + \ln n)$-approximation algorithm, where n is the size of the underlying set. The same algorithm can be translated to Hitting Set by duality. In [16], Jha et al. expressed the Label