Simple labeling schemes for graph connectivity

Rani Izsak and Zeev Nutov
The Open University of Israel. ranizsak@gmail.com, nutov@openu.ac.il

Abstract. Let $G = (V, E)$ be an undirected graph and let $S \subseteq V$. The $S$-connectivity $\lambda^S_G(u, v)$ of a pair of nodes $u, v$ in $G$ is the maximum number of $uv$-paths that no two of them have an edge or a node in $S \setminus \{u, v\}$ in common. Edge-connectivity is the case $S = \emptyset$ and node-connectivity is the case $S = V$. Given a graph $G = (V, E)$, an integer $k$, a subset $T \subseteq V$ of terminals, and $S \subseteq V$, we consider the problem of assigning small “labels” (binary strings) to the terminals, so that given the labels of two terminals $u, v \in T$, one can decide whether $\lambda^S_G(u, v) \geq k$ ($k$-partial labeling scheme) or to return $\min\{\lambda^S_G(u, v), k\}$ ($k$-full labeling scheme).

For edge-connectivity, there are known labeling schemes with max-label size $O(\log |T|)$ in the $k$-partial case and $O(\log |T| \cdot \min\{k, \log |T|\})$ in the $k$-full case [3]. We observe that this result extends to $S$-connectivity when $S \subset V \setminus T$ — so called “element-connectivity”, and combine it with the recently discovered decomposition of Chuzhoy and Khanna [1] of node-connectivity problems into element-connectivity problems to obtain simple labeling schemes for node-connectivity. If $q$ distinct queries are expected, our labeling schemes have max-label size $O(k \log |T| \log q)$ in the $k$-partial case and $O(k \log |T| \log q \cdot \min\{k, \log |T|\})$ in the $k$-full case, with success probability $1 - \frac{1}{q}$ for all queries. For a constant number of queries, this matches the lower bound $\Omega(k \log n)$ for the $k$-partial case of [3]. Consequently, we obtain deterministic labeling schemes with max-label size $O(k \log^2 |T|)$ in the $k$-partial case and $O(k \log^2 |T| \cdot \min\{k, \log |T|\})$ in the $k$-full case. This improves the bounds of Korman [4] $O(k^2 \log n)$ and $O(k^3 \log n)$, respectively, for $k = \Omega(\log |T|)$.

1 Introduction

1.1 Problem definition

Labeling schemes represent information of a graph by distributing it to the nodes. This is in contrast with the traditional “global” representations such as adjacency list/matrix. This distributed representation should enable answering predefined type of queries only by accessing the information (labels) at the relevant nodes. One can label any information by using enough bits at each node, but our aim is to find labels as small as possible that enable answering relevant queries in a time polynomial in the label sizes. In this paper we consider answering queries on connectivity between pairs of nodes. Formally, we are interested in the following type of labeling schemes, c.f. [2–5].
Definition 1 (Labeling scheme). Let $f$ be a function defined on pairs of a groundset $T$. An $f$-labeling scheme is a pair $< M_f, D_f >$ of algorithms, where the marker $M_f$ assigns a label (binary string) $L_f(v)$ to each $v \in T$, and the decoder $D_f$ is a polynomial time algorithm that given labels $L_f(u), L_f(v)$ of $u, v \in T$ returns $f(u, v)$.

We now define some graph-connectivity concepts. Let $G = (V, E)$ be an undirected graph and let $S \subseteq V$. The $S$-connectivity $\lambda^S_G(u, v)$ of a node pair $u,v$ in $G$ is the maximum number of $uv$-paths that no two of them have an edge or a node in $S \setminus \{u,v\}$ in common. Node-connectivity is the case $S = V$ and edge-connectivity is the case $S = \emptyset$. We use $\kappa_G(u, v) = \lambda^\emptyset_G(u, v)$ for the node-connectivity between $u,v$ in $G$ (the maximum number of internally-disjoint $uv$-paths in $G$), and $\lambda_G(u, v) = \lambda^\emptyset_G(u, v)$ for the edge-connectivity between $u,v$ in $G$ (the maximum number of edge-disjoint $uv$-paths in $G$). If we are also given a set $T \subseteq V$ of terminals so that $T \subseteq V \setminus S$, then for $u,v \in T$ we say that $\lambda^S_G(u, v)$ is the element-connectivity between $u,v$ in $G$. We will often omit the subscript “$G$” when it is clear from the context.

Given a graph $G = (V, E)$, $S \subseteq V$, an integer $k$, and a subset $T \subseteq V$ of terminals, we consider the problem of labeling the terminals while minimizing the max-label size, so that the given labels of two terminals $u,v \in T$, one can decide whether $\lambda^S_G(u, v) \geq k$ ($k$-partial labeling scheme) or to return $\min\{\lambda^S_G(u, v), k\}$ ($k$-full labeling scheme). Namely, in terms of Definition 1, $f(u, v)$ is defined for $u,v \in T$ as follows. In the $k$-partial case, $f(u, v) = TRUE$ if $\lambda^S_G(u, v) \geq k$ and $f(u, v) = FALSE$ otherwise. In the $k$-full case $f(u, v) = \min\{\lambda^S_G(u, v), k\}$. Clearly, by invoking a factor of $k$ in the max-label size one can obtain a $k$-full labeling scheme from a $k$-partial one.

1.2 Our results

Our results are summarized in the following two theorems.

For the edge-connectivity case, there are known labeling schemes with max-label size $O(\log |T|)$ in the $k$-partial case and $O(\log |T| \cdot \min\{k, \log |T|\})$ in the $k$-full case [3]. We observe that this result extends to element-connectivity (namely, to $S$-connectivity when $S \subseteq V \setminus T$).

Theorem 1. For element-connectivity, there exist labeling schemes with max-label size $O(\log |T|)$ in the $k$-partial case and $O(\log |T| \cdot \min\{k, \log |T|\})$ in the $k$-full case.

This result is asymptotically optimal, since it is optimal even for edge-connectivity [3], which is a special case of element-connectivity.

Chuzhoy and Khanna [1] gave an $O(k^3 \log |T|)$-approximation algorithm for the node-connectivity Steiner Network problem (also called the Survivable Network Design Problem), by showing that it can be decomposed into $O(k^3 \log |T|)$ element-connectivity Steiner Network problems. We combine a modification of this idea with Theorem 1 to obtain simple labeling schemes for node-connectivity.
Theorem 2. For node-connectivity, there exist randomized labeling schemes with max-label size $O(k \log |T| \log q)$ in the $k$-partial case and $O(k \log |T| \log q \cdot \min\{k, \log |T|\})$ in the $k$-full case, with success probability $1 - 1/q$ for $q$ queries. There also exist deterministic labeling schemes with max-label size $O(k \log^2 |T|)$ in the $k$-partial case and $O(k \log^2 |T| \cdot \min\{k, \log |T|\})$ in the $k$-full case.

There is a lower bound of $\Omega(k \log n)$ on the $k$-partial case [3]. Using the first part of Theorem 2, we may become close to this bound, as we wish, in the trade off of reducing the number of supported queries. In particular, for a constant number of supported queries, we achieve the bound $O(k \log n)$. We emphasize that the queries themselves are not known ahead; just their number determines our max-label size. The second part of Theorem 2 improves the bounds of Korman [4] $O(k^2 \log n)$ and $O(k^3 \log n)$, respectively, for $k = \Omega(\log |T|)$.

2 Labeling schemes for element-connectivity (Proof of Theorem 1)

We prove a generalization of Theorem 1. This is done by extending the labeling scheme of [3] for edge-connectivity to the following type of functions, that capture also element-connectivity.

Definition 2 (Equivalence function). A function $f$ that maps ordered pairs of a ground-set $T$ to non-negative integers or to $\infty$ is said to be an equivalence function if the following holds:
- $f(v, v) = \infty$ for all $v \in T$ (reflexivity).
- $f(u, v) = f(v, u)$ for all $u, v \in T$ (symmetry).
- $f(u, w) \geq \min\{f(u, v), f(v, w)\}$ for all $u, v, w \in T$ (transitivity).

The following statement is straightforward.

Claim. Let $f$ be an equivalence function on $T$. The relation $R_i = \{(u, v) \mid f(u, v) \geq i, u, v \in T\}$ is an equivalence for any integer $i \geq 0$. Moreover, $R_i$ is a refinement of $R_j$ if $i > j$, namely, for each equivalence class $C_i$ of $R_i$ there exists an equivalence class $C_j$ of $R_j$ such that $C_i \subseteq C_j$.

The following known statement can easily be deduced from Menger’s Theorem for $S$-connectivity.

Claim. Let $G = (V, E)$ be a graph, let $T \subseteq V$ and let $S \subseteq V \setminus T$ (namely, $S, T$ is a subpartition of $V$). Then the function $f(u, v) = \lambda^S_T(u, v)$ for all $u, v \in T$ is an equivalence function. Thus, element-connectivity is an equivalence function.

Thus, the following Theorem is a generalization of Theorem 1:

Theorem 3. Any equivalence function $f$ admits labeling schemes with max-label size $O(\log |T|)$ in the $k$-partial case and $O(\log |T| \cdot \min\{k, \log |T|\})$ in the $k$-full case.

In the rest of this section we prove Theorem 3, by adapting the edge-connectivity scheme presented in [3].
The \( k \)-partial case: For the marker \( M \), we give a unique label for each equivalence class of \( R_k \). The label \( L(v) \) of a node \( v \in T \) is the label of the unique equivalence class it belongs to. The label size is \( O(\log |T|) \), since there are no more equivalence classes than terminals. For the decoder \( D \), given two nodes \( u, v \in T \), we return \( TRUE \) if and only if \( L(u) = L(v) \).

The \( k \)-full case: We describe the marker \( M \). Recall that \( R_i \) is an equivalence relation for every \( i \). Using the idea from [3], we label each node using a least common ancestor labeling scheme in the following tree. The root is \( R_0 = V \). The \( k^{th} \) level relates to \( R_k \) and is a refinement of the \((k-1)^{th}\). The leaves are the singleton sets. The decoder is of the least common ancestor labeling scheme. The correctness is proved at [3] for edge-connectivity and it relies on the fact that it induces a family of equivalence relations with the refinement property we have mentioned. Therefore, the adaptation for equivalence function is immediate. The complexity of this scheme (as proved at [3] for \( n \) nodes) is \( O(\log^2 |T|) \). If we want to achieve a bound of \( O(k \log |T|) \), we can just use \( k \) times the scheme of the \( k \)-partial case, in a straightforward manner.

This completes the proof of Theorem 1.

3 Labeling schemes for node-connectivity (Proof of Theorem 2)

As was mentioned, the proof of Theorem 2 uses an idea of Chuzhoy and Khanna [1], modified and adjusted to our problem. We decompose the terminal set \( T \) into \( p = \left(\frac{4k}{2} \right) \) terminals sets \( T_i \). Setting \( S_i = V \setminus T_i \) we obtain \( p \) instances of element-connectivity. We show that these \( p \) instances together encode the node-connectivity in \( G \) of pairs from \( T \), with high probability. The \( p \) terminal sets \( T_i \) are defined as follows. Let each node in \( T \) select uniformly independently at random the unique subset it belongs to among \( 4k \frac{n}{n-k} \) subsets. \( T_1, \ldots, T_p \) are all the unions of any two such subsets. We will refer to the collection of sets \( T_1, \ldots, T_p \), as the basic decomposition.

Let \( S_i = V \setminus T_i \), \( i = 1, \ldots, p \). For \( u, v \in T_i \) let \( \lambda^i(u, v) = \lambda^S_{G}(u, v) \) denote the \( S_i \)-connectivity between \( u, v \) in \( G \). From the definition of \( S \)-connectivity we have:

**Fact 4** \( \lambda^S_{G}(u, v) \geq \kappa_G(u, v) \) for any \( S \subseteq V \) and \( u, v \in V \).

Using Fact 4 and some probabilistic arguments, we show the following:

**Lemma 1.** \( \Pr[\lambda^i(u, v) = \kappa(u, v)] \geq \frac{1}{16} \) for any \( u, v \in T_i \) with \( \kappa(u, v) \leq k \).

**Proof.** Let \( C \subseteq (V \setminus \{u, v\}) \cup E \) be a \( uv \)-cut of nodes and edges of size \( |C| = \kappa(u, v) \) so that there is no \( uv \)-path in \( G \setminus C \). Such \( C \) exists by Menger’s Theorem for node-connectivity; in fact, \( C \) consists of nodes only if \( u, v \) are not adjacent, or contains the unique edge \( uv \) otherwise. By Fact 4 \( \lambda^i(u, v) \geq \kappa(u, v) \). Thus \( \lambda^i(u, v) = \kappa(u, v) \) if all the nodes of \( C \) are in \( S^i = V \setminus T_i \), since then \( C \) is also a
$uv$-cut for $S_i$-connectivity of size $|C| = \kappa(u, v)$. Consequently, it is sufficient to show that the probability of the event $C \cap T_i = \emptyset$ is at least $1/(4e)$.

Note that the expected size of $T_i$ is no more than $\frac{n}{2k\ln n} = \frac{n-k}{2k}$. Therefore, by Markov Inequality:

$$\Pr \left[ |T_i| \leq \frac{n-k}{k} \right] \geq \frac{1}{2}$$

By considering the nodes of $C$ sequentially, and assuming the first $j$ nodes are in $S_i$, the probability that node $j+1$ belongs to $S_i$ is $\frac{(\kappa - |T_i|)-j}{\kappa - j}$. If $|T_i| \leq \frac{n-k}{4k}$, then this probability is at least $1-1/k$, and since $|C| = \kappa(u, v) \leq k$, the probability that all the nodes of $C$ are not in $T_i$ is at least $(1-1/k)^k \geq 1/(2e)$. Since $|T_i| \leq \frac{n-k}{k}$ with probability at least $1/2$, the probability that $C \cap T_i = \emptyset$ is at least $1/(4e)$, as claimed.

By the definition of the sets $T_1, \ldots, T_p$, for any $u, v \in T$ there exists $i$ so that $u, v \in T_i$. Thus combining Lemma 1 with Fact 4 we obtain:

**Corollary 1.** $\Pr[\min_i\{\lambda^i(u, v) \mid u, v \in T_i\} = \kappa(u, v)] \geq \frac{1}{4e}$ for any $u, v \in T$ with $\kappa(u, v) \leq k$.

A better lower bound on the probability of the event “$\min_i\{\lambda^i(u, v) \mid u, v \in T_i\} = \kappa(u, v)$” is achieved by repeating the basic decomposition uniformly and independently several times. We may also consider a subset $Q$ of pairs from $T$ rather than all pairs. Then our bounds are summarized in the following lemma:

**Lemma 2.** Let $Q$ be a set of pairs of nodes from $T$. If we define a new decomposition by repeating the basic decomposition, uniformly and independently, $s = \lceil 8e \ln |Q| \rceil$ times, then:

$$\Pr \left[ \min\{\lambda^i(u, v) \mid u, v \in T_i\} = \kappa(u, v) \text{ for all } \{u, v\} \in Q \text{ with } \kappa(u, v) \leq k \right] \geq 1-1/|Q|.$$  

In particular, for $Q$ being the set of all pairs from $T$ we obtain:

$$\Pr \left[ \min\{\lambda^i(u, v) \mid u, v \in T_i\} = \kappa(u, v) \text{ for all } u, v \in T \text{ with } \kappa(u, v) \leq k \right] \geq 1-2/|T|^2.$$  

**Proof.** Let $\varepsilon^{m}_{u,v}$ be the failure probability for $u, v \in Q$ and $m$ uniform independent repetitions of the basic decomposition (i.e. $\kappa(u, v) \leq k$, but $\min_i\{\lambda^i(u, v) \mid u, v \in T_i\} \neq \kappa(u, v)$). By Corollary 1, $\varepsilon^1_{u,v} \leq 1 - \frac{1}{4e}$. Therefore, $\varepsilon^m_{u,v} \leq (1 - 1/(4e))^m$, and then:

$$\varepsilon^{\lceil 8e \ln |Q| \rceil}_{u,v} \leq \left( (1 - 1/(4e))^{\ln |Q|} \right)^2 \leq \left( (1/e)^{\ln |Q|} \right)^2 = 1/|Q|^2.$$  

The result now follows from the union bound.

For the proof of theorem 2, we may assume that $k \leq n/2$, as otherwise we can trivially meet our max-label size bounds by storing at each node the data it needs regarding any other node. Under this assumption $p = O(k^2)$, but the key point is that every node $v \in T$ appears in only $O(k)$ sets from $T_1, \ldots, T_p$.

We now describe the marker and the decoder of our node-connectivity labeling schemes.
We repeat the basic decomposition \([8e] \ln |Q|\) times. This gives a (multi-)collection \(T\) of \(O(k^2 \log |Q|)\) subsets of \(T\) (\(T\) may contain the same set more than once). For \(v \in T\) let \(J(v)\) be the set of indices of sets in \(T\) that contain \(v\). Note that \(|J(v)| = O(k \log |Q|)\) for every \(v \in T\). Each \(T^j \in T\) defines an instance of element-connectivity, and for every \(v \in T^j\) so that \(j \in J(v)\) we set \(L^j(v)\) to be the corresponding labeling of \(v\) as in Theorem 1. The label \(L(v)\) of \(v\) will be a list of labels \(L^j(v), j \in J(v)\), where for each label we also specify the appropriate index \(j\). Consequently, \(|L(v)|\) is bounded by \(|T(v)|\) times the bounds for element-connectivity given in Theorem 1, plus the size of the index \(j\). Recall that \(|T(v)| = O(k \log |Q|)\), that our bounds for element-connectivity are \(O(\log n)\) in the \(k\)-partial case and \(O(\log n \cdot \min\{k, \log n\})\) in the \(k\)-full case, and note that the size of the index \(j\) is \(O(\log(k^2 \log |Q|)) = O(\log k + \log \log |Q|)\); the latter is dominated by the other parts. This gives the bounds in Theorem 2.

Let \(D\) be the decoder of the corresponding element-connectivity scheme. Given \(u, v \in T\) we act as follows. For every \(j \in J(u) \cap J(v)\) (i.e. for every index that appears in both \(L(u)\) and \(L(v)\)) we use \(D\) to compute the element-connectivity value \(f^j(u, v)\) for \(S^j = V \setminus T^j\) (may be boolean or integer, depending on the scheme). In the \(k\)-partial case we return \(TRUE\) if \(f^j(u, v) = TRUE\) for all \(j \in J(u) \cap J(v)\), and \(FALSE\) otherwise (i.e., we return “AND” of the labels).

In the \(k\)-full case we return the minimum among \(\min\{f^j(u, v) \mid j \in J(u) \cap J(v)\}\) and \(k\). Clearly, the running time is polynomial in the labels’ sizes. Correctness follows from Lemma 2.

**Deterministic labeling schemes:** We show existence of deterministic labeling schemes as in Theorem 2. The randomized decomposition we described, with \((\text{high})\) non zero probability, satisfies the following for all pairs from \(T\):

\[
- \min_i \{\lambda^i(u, v) \mid u, v \in T_i\} = \kappa(u, v) \text{ if } \kappa(u, v) \leq k;
- \lambda^i(u, v) \geq \kappa(u, v) \text{ for all } i.
\]

Since our analysis is valid for any input (it is independent of it), there exists a decomposition obeying both conditions, for any input graph. The result follows by using this decomposition at the marker.

This completes the proof of Theorem 2.

**References**