A Rounding by Sampling Approach to the Minimum Size k-Arc Connected Subgraph Problem

Bundit Laekhanukit[1], Shayan Oveis Gharan[2], and Mohit Singh[3]

[1] School of Computer Science, McGill University
[2] Department of Management Science and Engineering, Stanford University
[3] McGill University and Microsoft Research, Redmond

Abstract. In the k-arc connected subgraph problem, we are given a directed graph G and an integer k and the goal is the find a subgraph of minimum cost such that there are at least k-arc disjoint paths between any pair of vertices. We give a simple $(1 + 1/k)$-approximation to the unweighted variant of the problem, where all arcs of G have the same cost. This improves on the $1+2/k$ approximation of Gabow et al. [GGTW09]. Similar to the 2-approximation algorithm for this problem [FJ81], our algorithm simply takes the union of a k in-arborescence and a k out-arborescence. The main difference is in the selection of the two arborescences. Here, inspired by the recent applications of the rounding by sampling method (see e.g. [AGM$^+$10, MOS11, OSS11, AKS12]), we select the arborescences randomly by sampling from a distribution on unions of k arborescences that is defined based on an *extreme point solution* of the linear programming relaxation of the problem. In the analysis, we crucially utilize the sparsity property of the extreme point solution to upper-bound the size of the union of the sampled arborescences.

To complement the algorithm, we also show that the integrality gap of the minimum cost strongly connected subgraph problem (i.e., when $k = 1$) is at least $3/2 - \epsilon$, for any $\epsilon > 0$. Our integrality gap instance is inspired by the integrality gap example of the asymmetric traveling salesman problem [CGK06], hence providing further evidence of connections between the approximability of the two problems.

1 Introduction

In the *minimum cost k-arc connected spanning subgraph* (min-cost k-ACSS) problem, we are given a directed graph $G = (V, A)$ with cost $c : A \to R$ on the arcs and a connectivity requirement k. The goal is to find a spanning subgraph $G' = (V, A')$ of G of minimum total cost which is *k-arc connected*, i.e., every pair of vertices have at least k-arc disjoint paths between them. The special case of $k = 1$, 1-ACSS problem, is called the *minimum cost strongly connected subgraph* problem. In the unweighted variant of k-ACSS, the *minimum size k-arc connected spanning subgraph* (min-size k-ACSS) problem, where all arcs of G have the same cost, we want to minimize the number of arcs that we choose.

The min-cost k-ACSS problem has a 2-approximation algorithm [FJ81], and it has been a long standing open problem to improve this bound. Significant attention has been given to the unweighted variant of the problem. In particular, the minimum size strongly connected subgraph problem is very well studied [FJ81, KRY94, KRY96, Vet01, ZNI03], and the current best approximation ratio is $3/2$, which is due to Vetta [Vet01]. The min-size k-ACSS problem has been shown to be easier as k increases [CT00, Gab04, GGTW09], and the best approximation ratio is $1 + 2/k$ that is given in the work of Gabow et al. [GGTW09]. This approximation ratio is almost tight as the min-size k-ACSS problem does not admit $(1 + \epsilon/k)$-approximation, for some fixed $\epsilon > 0$, unless P=NP [GGTW09]. Similar to the directed case, the *minimum size k-edge connected subgraph spanning* problem, an undirected variant of the min-size k-ACSS problem, is known to be easier as k increases, and the best known approximation ratio for this problem is $1 + 1/(2k) + O(1/k^2)$ due to Gabow and Gallagher [GG08].

1.1 Our Results

In this paper, we give improved upper and lower bounds for the k-ACSS problem. We first show the following improved algorithms for the min-size k-ACSS problem.

Theorem 1. *For any $k \geq 1$, there is a $\min\{7/4, 1 + 1/k\}$-approximation algorithm for the min-size k-ACSS problem.*

Similar to the simple 2-approximation algorithm for the minimum-cost k-ACSS problem, our algorithm takes the union of a k in-arborescence and a k out-arborescence. The main difference is in the selection of the two arborescences. Here, we select the arborescences randomly by sampling from a distribution on unions of k arborescences that is defined by the linear programming relaxation of the problem. In particular, we write a convex combination of the unions of k-arborescences such that the marginal probability of each arc is bounded above by its fraction in the solution of LP relaxation.

The algorithm essentially employs the rounding by sampling method that recently has been applied to various problems in the algorithm design and online optimization literature (c.f. [AGM$^+$10, MOS11, OSS11, AKS12]), while the analysis is much simpler in our setting. Here, the main technical difference is a crucial use of the extreme point solutions of LP relaxation. In particular, because of the sparsity of the extreme point solutions, we can argue that the union of k in-arborescences and k out-arborescences is not much larger than the size of the support of the LP extreme point solution and thus the size of the optimum.

Our result improves on the $(1+\frac{2}{k})$-approximation of Gabow et al. [GGTW09] for the min-size k-ACSS problem, for any $k > 0$. Furthermore, for the minimum size strongly connected subgraph problem, while we do not improve the approximation factor of $\frac{3}{2}$ [Vet01], our algorithm is much simpler and gives a possible direction for weighted version of the problem.

To complement the positive results, we prove that the integrality gap of the natural linear programming relaxation of the strongly connected subgraph problem is bounded below by $3/2 - \epsilon$ for any $\epsilon > 0$.

Theorem 2. *For any $\epsilon > 0$, the integrality gap of the standard linear programming relaxation for the minimum cost strongly connected subgraph problem is at least $\frac{3}{2} - \epsilon$.*

To the best of our knowledge, there is no explicit construction that gives a lower bound on the integrality gap of the minimum cost strongly connected subgraph problem. Our integrality gap example builds on a similar construction for the asymmetric traveling salesman problem [CGK06] and shows stronger connections between the two problems.

1.2 Notations

Let $\delta^+_G(U) := \{(u,v) \in E : u \in U, v \in V \setminus U\}$ denote the set of arcs leaving U in a graph G; if G is clear in the context, we will skip the subscript.

A graph G is *k-arc connected* if and only if every (proper) subset of vertices $U \subset V$ have at least k leaving arcs, i.e., $|\delta^+_G(U)| \geq k$, and G is *strongly connected* if it is 1-arc connected. We may drop the subscript if G is clear in the context. We use the following Linear Programming relaxation for k-ACSS.

$$\begin{array}{llll}
\text{(LP-ACSS)} & \text{minimize} & \displaystyle\sum_{a \in A} c_a x_a & \\
& \text{subject to} & x(\delta^+(U)) \geq k & \forall U \neq \emptyset, U \subsetneq V \\
& & 0 \leq x_a \leq 1 & \forall e \in E,
\end{array}$$

where $x(\delta^+(U)) = \sum_{a \in \delta^+(U)} x_a$. Throughout the paper x will always be an optimum solution of the (LP-ACSS).

For any vector $y : A \to \mathbb{R}$, and a set $F \subset A$ of arcs, $y(F) := \sum_{a \in F} y_a$, is the sum of the values of the arcs in F, and $c(F) := \sum_{a \in F} c_a$ is the sum of the cost of the arcs in F. Also, $\chi(F)$ denotes the characteristic vector of the set F, i.e., $\chi(F)_a = 1$ if $a \in F$ and $\chi(F)_a = 0$ otherwise.

2 An Approximation Algorithm for Min-Size k-ACSS

In this section, we prove Theorem 1: given a graph G, we give a polynomial time algorithm that finds a k-arc connected subgraph of G such that it has no more than $\min\{1+1/k, 7/4\}$ of the arcs of the optimum solution. Before describing the algorithm, we need to recall some of the properties of arborescences in directed graphs.

Given a directed graph G and a (root) vertex $r \in V$, an *r-out arborescence* T of G is a directed tree rooted at r that contains a path from r to every other vertex of G. An *r-out k-arborescence* is a subgraph T of G that is the union of k arc-disjoint r-out arborescences. An *r-in arborescence* and an *r-in k-arborescence* are defined analogously. The following polyhedron plays an important role in the design and analysis of our algorithm.

$$P^{out} = \{y : y(\delta^+(U)) \geq k, \quad \forall \emptyset \neq U \subsetneq V \setminus \{r\}, 0 \leq y \leq 1\}$$

Frank [Fra79] showed that P^{out} is the up hull of the convex hull of r-out k-arborescences (see Corollary 53.6a [Sch03]), and it can be seen that every feasible solution of (LP-ACSS) is a point in P^{out}. Vempala and Carr [CV02] gave a polynomial-time algorithm that allows us to write a point $x \in P^{out}$ as a convex combination of k arc-disjoint arborescences. Their algorithm requires a polynomial-time algorithm for finding an r-out k-arborescences [Edm73, Gab91].

Lemma 1. *[Fra79, CV02, Edm73, Gab91] P^{out} is the convex hull of subsets of A containing r-out k-arborescences. Moreover, given any fractional solution $y \in P^{out}$, there is a polynomial time algorithm that finds a convex combination of r-out k-arborescences, $T_1, \ldots, T_l$, such that*

$$y \geq \sum_{i=1}^{l} \lambda_i \chi(T_i).$$

The above lemma holds analogously for the r-in arborescences. Now, since $x \in P^{out}$, we can write a distribution of r-out(in) k-arborescences such that probability of each arc $a \in A$ chosen in a random k-arborescence is bounded above by x_a:

Corollary 1. *There are distributions $\mathcal{D}_{in}(r)$ and $\mathcal{D}_{out}(r)$ of r-in k-arborescences and r-out k-arborescences, such that the marginal value of each arc $a \in A$ is bounded above by x_a, i.e., for all arcs $a \in A$,*

$$\mathbf{P}_{T \sim \mathcal{D}_{in}(r)}[a \in T] \leq x_a,$$
$$\mathbf{P}_{T \sim \mathcal{D}_{out}(r)}[a \in T] \leq x_a.$$

Moreover, these distributions can be computed in polynomial time.

Now, we are ready to describe our algorithm. We sample k-arborescences T_{in} and T_{out} independently from $\mathcal{D}_{in}$ and $\mathcal{D}_{out}$, respectively, and we then return $T_{in} \cup T_{out}$ as an output. The details are described in Algorithm 1.

Next, we show that the approximation ratio of the above algorithm is no more than $1 + 1/k$.

Theorem 3. *For any directed graph G, Algorithm 1 always produces a k-arc connected subgraph of G such that the expected size of the solution is no more than* $\min\{7/4, 1 + 1/k\}$ *of the optimum.*

Proof. First, we show that the union of any pair of r-in and r-out k-arborescences is k-arc connected. Let $T_{in}(T_{out})$ be a r-in (r-out) k-arborescence, and $H =$

Algorithm 1 Approximation Algorithm for Min-Size k-ACSS

```
1: Solve (LP-ACSS) to get an optimum *extreme point* solution $x$.
2: Find distributions $\mathcal{D}_{in}(r)$ and $\mathcal{D}_{out}(r)$ on $r$-in and $r$-out $k$-arborescences, respectively, such that the marginal value of each arc $a \in A$ is bounded above by $x_a$.
3: Sample an $r$-in $k$-arborescence $T_{in}$ from $\mathcal{D}_{in}(r)$ and an $r$-out $k$-arborescence $T_{out}$, *independently*, from $\mathcal{D}_{out}(r)$.
4: **return** $T_{in} \cup T_{out}$.
```

$T_{in} \cup T_{out}$. Since both T_{in} and T_{out} are unions of k arc-disjoint arborescences, there are k arc-disjoint paths from each of the vertices to r and k arc-disjoint paths from r to each of the vertices. Therefore, H remains strongly connected after removing any set of $k-1$ arcs. Hence, H is k-arc connected.

It remains to show that the expected size of the solution is no more than $\min\{1+1/k, 7/4\}$ of the optimum, i.e.,

$$\frac{\mathbf{E}_{T_{in}\sim\mathcal{D}_{in}(r),T_{out}\sim\mathcal{D}_{out}(r)}\left[|T_{in} \cup T_{out}|\right]}{|\mathrm{OPT}|} \le \min\left\{\frac{7}{4}, 1+\frac{1}{k}\right\}.$$

To simplify the notation, we will skip the subscript and write $\mathbf{E}\left[|T_{in} \cup T_{out}|\right]$ to mean $\mathbf{E}_{T_{in}\sim\mathcal{D}_{in}(r),T_{out}\sim\mathcal{D}_{out}(r)}\left[|T_{in} \cup T_{out}|\right]$. Similarly, we will skip the subscripts for $\mathbf{P}_{T_{in}\sim\mathcal{D}_{in}(r)}\left[a \in T_{in}\right]$ and $\mathbf{P}_{T_{out}\sim\mathcal{D}_{out}(r)}\left[a \in T_{out}\right]$.

Since T_{in} and T_{out} are chosen independently,

$$\begin{aligned}\mathbf{E}\left[|T_{in} \cup T_{out}|\right] &= \sum_{a\in A}\left\{\mathbf{P}\left[a \in T_{in}\right] + \mathbf{P}\left[a \in T_{out}\right] - \mathbf{P}\left[a \in T_{in}\right]\cdot\mathbf{P}\left[a \in T_{out}\right]\right\}\\ &\le \sum_{a\in A} 2x_a - \sum_{a\in A} x_a^2.\end{aligned}$$

The last inequality follows from Corollary 1 and the fact that $x_a \le 1$ for all $a \in A$. Let $F := \{a : 0 < x_a < 1\}$ be the set of the fractional arcs (i.e., set of arcs with non-integer values in the solution of (LP-ACSS)). Since x is an optimal solution of (LP-ACSS), $|\mathrm{OPT}| \ge \sum_{a\in A} x_a$. Therefore,

$$\begin{aligned}\frac{\mathbf{E}\left[|T_{in} \cup T_{out}|\right]}{|\mathrm{OPT}|} &\le 1 + \frac{\sum_{a\in A} x_a - \sum_{a\in A} x_a^2}{\sum_{a\in A} x_a}\\ &= 1 + \frac{x(F) - \sum_{a\in F} x_a^2}{x(A)}\\ &\le 1 + \frac{x(F) - x(F)^2/|F|}{x(A)}, \qquad (1)\end{aligned}$$

where the last inequality follows from Jenson's inequality and the fact that $f(t) = -t^2$ is a concave function.

Since x is an extreme point solution of (LP-ACSS), x is a sparse vector. It follows from the work of Melkonian and Tardos [MT04] (see also [GGTW09]), that the number of fractional arcs, $|F|$, is no more than $4n$. Hence,

$$\frac{x(F) - x(F)^2/|F|}{x(A)} \leq \frac{x(F) - x(F)^2/4n}{x(A)} \leq \frac{n}{x(A)} \leq \frac{1}{k}, \tag{2}$$

where the second inequality follows since $x(F) - x(F)^2/4n$ attains its maximum at $x(F) = 2n$, and the last inequality follows from the fact that $x(A) = \sum_{v \in V} x(\delta^+(v)) \geq nk$. On the other hand, since $x(F) \leq x(A)$, we get

$$\frac{x(F) - x(F)^2/|F|}{x(A)} \leq \frac{1}{2} + \frac{x(F) - x(F)^2/2n}{2x(A)} \leq \frac{1}{2} + \frac{n}{4x(A)} \leq \frac{3}{4}. \tag{3}$$

The theorem simply follows by putting equations (1),(2),(3) together. □

Remark 1. Since the distributions $\mathcal{D}_{in}(r)$ and $\mathcal{D}_{out}(r)$ can be constructed such that the support of each distribution has size only polynomially large in n, the algorithm can be derandomized simply by choosing a pair of k-arborescences that have the minimum number of arcs in their union.

3 A Lower Bound on the Integrality Gap

In this section, we prove Theorem 2: we show a lower-bound of $1.5 - \epsilon$, for any arbitrary small $\epsilon > 0$, on the integrality gap of (LP-ACSS) for $k = 1$. Our construction is based on the LP-gap construction of the asymmetric traveling saleman problem by Charikar, Goemans and Karloff [CGK06].

3.1 Construction

Let $r > 0$ be an integral parameter that will be defined later. We start by defining the integrality gap example, $G(d, s, t)$, by a recursive construction of depth d. In any graph $G(d, s, t)$, d is the *depth*, r is the number of *columns*, s, t are the *source, sink* vertices, respectively. We allow s and t to be the same vertex. We will construct $G(d, s, t)$ inductively such that it contains exactly r copies of $G(d - 1, ., .)$.

We start by describing $G(1, s, t)$. The graph consists of s, t and r distinct vertices $v_1, \ldots, v_r$. Let $v_0 = s$ and $v_{r+1} = t$; note that v_0 and v_{r+1} may be the same depending on the given parameters s and t. For any $1 \leq i \leq r + 1$, we include arcs (v_i, v_{i-1}) and (v_{i-1}, v_i) in $G(1, s, t)$. Therefore,

$$A(G(1, s, t)) := \{(v_{i-1}, v_i), (v_i, v_{i-1}), 1 \leq i \leq r + 1\}.$$

Next, we define $G(d, s, t)$. The graph consists of s, t and r distinct copies of $G(d - 1, ., .)$. In particular, let $v_1, \ldots, v_r, u_1, \ldots, u_r$ be $2r$ distinct vertices, and $v_0 = u_{r+1} = s$ and $v_{r+1} = u_0 = t$. For any $1 \leq i \leq r$, include a distinct copy of

$G(d-1,.,.)$ with source u_i and sink v_i. Also, for any $1 \le i \le r+1$, include the arcs (v_i, v_{i-1}) and (u_{i-1}, u_i). Therefore,

$$A(G(d,s,t)) := \{(u_{i-1},u_i),(v_i,v_{i-1}), 1 \le i \le r+1\} \cup \left\{ \bigcup_{i=1}^{r} A(G(d-1,u_i,v_i)) \right\}.$$

Figure 3.1 illustrates the graph $G(3,s,s)$ for $r=3$.

Our integrality gap example is $G_d := G(d,s,s)$, where the source and the sink are unified. The i^{th} *column* of G_d is defined to be the i^{th} copy of the $G(d-1,.,.)$, i.e., $G_d^{(i)} := G(d-1,u_i,v_i)$. The set of arcs that connect the r columns with s and t, i.e., $A(G_d) \setminus \bigcup_{i=1}^{r} A(G_d^{(i)})$, are denoted by d^{th} *level arcs*. Similarly, the l^{th} level arcs of G_d are defined to be set of arcs included at the l^{th} level of induction. For example, the $(d-1)^{th}$ level arcs of G_d are $\bigcup_{i=1}^{r}\left(A(G_d^{(i)}) \setminus \bigcup_{j=1}^{r} A(G_d^{(i;j)})\right)$, where $G_d^{(i;j)}$ is the j^{th} column of $G_d^{(i)}$.

We define the costs of the arcs of G_d such that, for any $1 \le l \le d$, the total cost of the arcs at level l is equal to 1. In other words, the cost of each arc at level l, $c_d(l)$, is the reciprocal of the number of arcs at level l. By the construction of G_d, we have

$$c_d(l) := \frac{1}{2(r+1)r^{d-l}}. \tag{4}$$

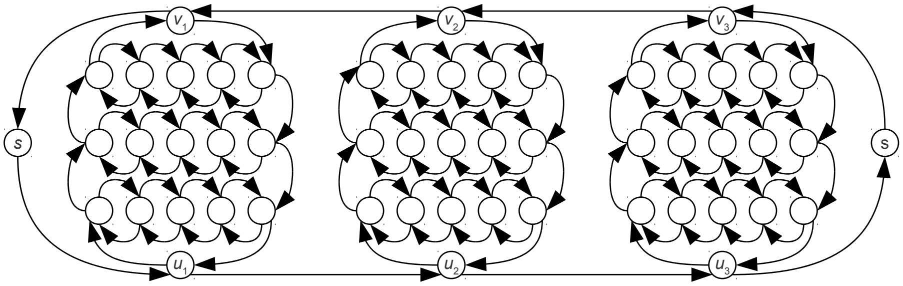

Fig. 1. An illustration of the graph $G(3,s,s)$, for $r=3$. Note that the vertices labeled "s" on the left and on the right are the same.

3.2 Lower Bounding the Integrality Gap

We show that for any $d>0$, and for a sufficiently large r, the integrality gap of the instance $G(d,s,s)$ is at least $3/2 - O(1/d)$.

Theorem 4. *For any $d>0$ and $r \ge d$, the integrality gap of the instance $G(d,s,s)$ is at least $3/2 - 8/d$.*

First, we show that the optimal value of the LP is at most $d/2$. Define $x_a^* := 1/2$ for all arcs $a \in A(G_d)$. Charikar et al. [CGK06] show that x^* belongs to the Held-Karp relaxation polytope [HK70]. Since any solution of the Held-Karp relaxation polytope is a feasible solution to (LP-ACSS) for $k = 1$, x^* is also a feasible solution to (LP-ACSS). Furthermore, since the sum of the cost of the arcs of G_d is d, i.e., $c(A(G_d)) = d$, we have $\sum_a c(a)x_a^* = d/2$. Hence, the optimal value of LP is at most $d/2$.

Lemma 2 (Charikar et al. [CGK06]). *For $k = 1$, the optimum value of* (LP-ACSS) *for the graph G_d is at most $d/2$.*

For any $d > 0$, let H_d be the minimum cost strongly connected subgraph of G_d, and $T(d) := c(A(H_d))$ be the cost of H_d. In the rest of the section, we prove the following lemma:

Lemma 3. *For all $d > 0$,*

$$T(d) \geq \frac{3d-1}{4} - \frac{3d}{r}. \tag{5}$$

Let $H_d^{(i)} := H_d \cap G_d^{(i)}$ be the i^{th} column of H_d. Observe that $H_d^{(i)}$ can be incident to (at most) four arcs of the d^{th} level arcs of H_d. Let

$$A_d(i) := \{(v_i, v_{i-1}), (v_{i+1}, v_i), (u_{i-1}, u_i), (u_i, u_{i+1})\} \cap A(H_d),$$

be the set of those arcs. We can lower-bound $c(A(H_d^{(i)}))$ based on the number of arcs that is incident to $H_d^{(i)}$ (note that since H_d is strongly connected, $|A_d(i)| \geq 2$):

Case 1: $|A_d(i)| \geq 3$
In this case, we must have

$$c(A(H_d^{(i)})) \geq T(d-1)/r. \tag{6}$$

The inequality essentially follows from the fact that $H_d^{(i)}$ is a strongly connected subgraph of G_{d-1}. This is because the remaining arcs of the graph, $H_d \setminus H_d^{(i)}$, can only connect (or unify) the source and sink of $H_d^{(i)}$, i.e., u_i and v_i. The $1/r$ factor follows from the fact that the cost of each arc of G_{d-1} is r times the corresponding arc in $G_d^{(i)}$.

Case 2: $|A_d(i)| = 2$, and each of u_i and v_i is incident to exactly one arc of $A_d(i)$
Similar to the previous case, here we have

$$c(A(H_d^{(i)})) \geq T(d-1)/r. \tag{7}$$

As we will see in Lemma 4, at most two columns of H_d may satisfy this case. Therefore, although we have the worse lower-bound on $c(H_d^{(i)})$ in this case, it has an insignificant effect on the final lower-bound.

Case 3: $|A_d(i)| = 2$, and one of u_i or v_i is incident to none of the arcs of $A_d(i)$
Here we obtain a better lower-bound. For $1 \le j \le r$, let $H_d^{(i;j)}$ be the j^{th} column of $H_d^{(i)}$ with source $u_{i,j}$ and sink $v_{i,j}$. It follows that the only u_i, v_i (or v_i, u_i) path in H_d is the one that is made by the $d-1$ level arcs connecting the columns of $H_d^{(i)}$, i.e., $u_i, u_{i,1}, u_{i,2}, \ldots, u_{i,r}, v_i$ (resp. $v_i, v_{i,r}, v_{i,r-1}, \ldots, v_{i,1}, u_i$). Therefore, $H_d^{(i)}$ must contain all of the $(d-1)^{th}$ level arcs of $G_d^{(i)}$. Since each column of $H_d^{(i)}$ is incident to 4 arcs of level $(d-1)^{th}$, by repeated application of case 1, we obtain

$$\begin{aligned} c(A(H_d^{(i)})) &\ge 2(r+1)c_d(d-1) + \sum_{j=1}^{r} c(A(H_d^{(i;j)})) \\ &= 2(r+1)c_d(d-1) + \frac{T(d-2)}{r}. \end{aligned} \quad (8)$$

Next, we show that there are at most 2 columns satisfying the second case.

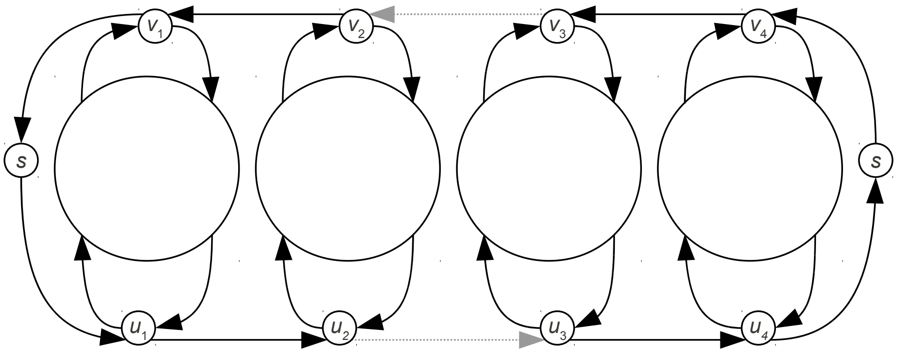

Fig. 2. An illustration of H_d where the second column satisfies Case 2. The black arcs represent the arcs of H_d, and grey arcs represent the removed arcs. Observe that every arc at level d is a min-cut of H_d.

Lemma 4. *At most two columns of H_d satisfy the second case.*

Proof. The proof is a simple case analysis argument. First, observe that there exists a column satisfying the second case in H_d if and only if $(v_i, v_{i-1}), (u_{i-1}, u_i) \notin H_d$ for some $1 \le i \le r+1$. Now, suppose this is the case. It then follows that H_d must contain all arcs at level d except these two arcs because each of the other arcs is a min-cut of H_d. See Figure 3.2. Therefore, all except (at most) two of the columns of H_d are adjacent to exactly 4 arcs at level d. □

Now we are ready to prove Lemma 3.

Proof of Lemma 3. We prove by induction. First observe that $T(0) = 0$ and $T(1) = 1/2$ satisfying (5). Let $N_1, N_2, (r - N_1 - N_2)$ be the number of columns satisfying case 1, 2, 3, respectively. We divide the cost of each arc at level d equally between the columns incident to it. This incurs a cost of $3c_d(d)/2$ to the columns satisfying case 1, $c_d(d)$ to the rest of the columns and at least $c_d(d)$ to the source vertex s (note that s is adjacent to at least two arcs at level d). Using equations (6), (7), (8) we get:

$$\begin{aligned}
T(d) \geq c_d(d) + \min_{0 \leq N_1, N_2 \leq r} \Bigg\{ & N_1 \left(\frac{3c_d(d)}{2} + \frac{T(d-1)}{r} \right) + N_2 \left(c_d(d) + \frac{T(d-1)}{r} \right) \\
& + (r - N_1 - N_2) \left(c_d(d) + 2(r+1)c_d(d-1) + \frac{T(d-2)}{r} \right) \Bigg\} \\
\geq \min_{0 \leq N \leq r} \Bigg\{ & N \left(\frac{3c_d(d)}{2} + \frac{T(d-1)}{r} \right) \\
& + (r - N) \left(c_d(d) + 2(r+1)c_d(d-1) + \frac{T(d-2)}{r} \right) \Bigg\} \\
\geq \min_{0 \leq \alpha \leq 1} \Bigg\{ & \alpha \left(\frac{3r}{4(r+1)} + T(d-1) \right) + (1-\alpha) \left(\frac{3r}{2(r+1)} + T(d-2) \right) \Bigg\} \\
\geq \min \{ & 3/4 + T(d-1), 3/2 + T(d-2) \} - 3/r.
\end{aligned}$$

The second inequality follows from the fact that $N_2 \leq 2$. The third inequality follows from equation (4), and the last one follows from a simple algebra.

Now, we may apply the induction hypothesis to $T(d-1)$ and $T(d-2)$. We get

$$\begin{aligned}
T(d) &\geq \min \left\{ \frac{3}{4} + \frac{3(d-1)-1}{4} - \frac{3(d-1)}{r}, \frac{3}{2} + \frac{3(d-2)-1}{4} - \frac{3(d-2)}{r} \right\} - \frac{3}{r} \\
&\geq \frac{3d-1}{4} - \frac{3d}{r},
\end{aligned}$$

which completes the proof. □

This completes the proof of Theorem 4.

4 Conclusion

We presented a simple $(1 + 1/k)$-approximation algorithm based on the rounding by sampling method for the minimum size k-arc connected subgraph problem. Unlike recent applications of the rounding by sampling method [AGM$^+$10, OSS11], our algorithm has a flavor of the *iterated rounding method* [Jai01] in its particular use of the extreme point solutions. The main open problem is to find a better than factor 2-approximation for the minimum cost strongly connected subgraph problem.

We also showed that the integrality gap of the minimum cost strongly connected subgraph problem is at least $1.5 - \epsilon$, for any $\epsilon > 0$. This leaves an interesting open question whether the lower bound of $1 + \Omega(1/k)$ is achievable for the minimum size k-arc connected subgraph problem as well.

Acknowledgments: We thank Joseph Cheriyan for useful discussions on the preliminary construction of the integrality-gap instance.

References

[AGM$^+$10] Arash Asadpour, Michel X. Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. An $O(\log n/\log\log n)$-approximation algorithm for the asymmetric traveling salesman problem. In *SODA*, pages 379–389, 2010.

[AKS12] Hyung-Chan An, Robert Kleinberg, and David B. Shmoys. Improving Christofides' algorithm for the s-t path tsp. In *STOC (to appear)*, 2012.

[CGK06] Moses Charikar, Michel X. Goemans, and Howard J. Karloff. On the integrality ratio for the asymmetric traveling salesman problem. *Math. Oper. Res.*, 31(2):245–252, 2006. Preliminary version in FOCS 2004.

[CT00] Joseph Cheriyan and Ramakrishna Thurimella. Approximating minimum-size k-connected spanning subgraphs via matching. *SIAM J. Comput.*, 30(2):528–560, 2000. Preliminary version in FOCS 1996.

[CV02] Robert D. Carr and Santosh Vempala. Randomized metarounding. *Random Struct. Algorithms*, 20(3):343–352, 2002.

[Edm73] Jack Edmonds. Edge-disjoint branchings. *Combinatorial algorithms (Courant Comput. Sci. Sympos. 9, New York Univ., New York, 1972)*, pages 91–96, 1973.

[FJ81] Greg N. Frederickson and Joseph JáJá. Approximation algorithms for several graph augmentation problems. *SIAM J. Comput.*, 10(2):270–283, 1981.

[Fra79] A. Frank. Covering branchings. *Acta Scientiarum Mathematicarum (Szeged)*, 41:77–81, 1979.

[Gab91] Harold N. Gabow. A matroid approach to finding edge connectivity and packing arborescences. In *STOC*, pages 112–122, 1991.

[Gab04] Harold N. Gabow. Special edges, and approximating the smallest directed k-edge connected spanning subgraph. In *SODA*, pages 234–243, 2004.

[GG08] Harold N. Gabow and Suzanne Gallagher. Iterated rounding algorithms for the smallest k-edge connected spanning subgraph. In *SODA*, pages 550–559, 2008.

[GGTW09] Harold N. Gabow, Michel X. Goemans, Éva Tardos, and David P. Williamson. Approximating the smallest k-edge connected spanning subgraph by LP-rounding. *Networks*, 53(4):345–357, 2009. Preliminary version in SODA 2005.

[HK70] M. Held and R. Karp. The traveling salesman problem and minimum spanning trees. *Operations Research*, 18:1138–1162, 1970.

[Jai01] Kamal Jain. A factor 2 approximation algorithm for the generalized steiner network problem. *Combinatorica*, 21(1):39–60, 2001. Preliminary version in FOCS 1998.

[KRY94] Samir Khuller, Balaji Raghavachari, and Neal E. Young. Approximating the minimum equivalent digraph. In *Proceedings of the fifth annual*

ACM-SIAM symposium on Discrete algorithms, SODA '94, pages 177–186, Philadelphia, PA, USA, 1994. Society for Industrial and Applied Mathematics.

[KRY96] Samir Khuller, Balaji Raghavachari, and Neal E. Young. On strongly connected digraphs with bounded cycle length. *Discrete Applied Mathematics*, 69(3):281–289, 1996.

[MOS11] Vahideh H. Manshadi, Shayan Oveis Gharan, and Amin Saberi. Online stochastic matching: Online actions based on offline statistics. In *SODA*, pages 1285–1294, 2011.

[MT04] V. Melkonian and E. Tardos. Algorithms for a network design problem with crossing supermodular demands. *Networks*, 43(4):256–265, 2004.

[OSS11] Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. A randomized rounding approach to the traveling salesman problem. In *FOCS*, pages 550–559, 2011.

[Sch03] A. Schrijver. *Combinatorial Optimization*. Springer, 2003.

[Vet01] Adrian Vetta. Approximating the minimum strongly connected subgraph via a matching lower bound. In *SODA*, pages 417–426, 2001.

[ZNI03] Liang Zhao, Hiroshi Nagamochi, and Toshihide Ibaraki. A linear time 5/3-approximation for the minimum strongly-connected spanning subgraph problem. *Inf. Process. Lett.*, 86:63–70, April 2003.