On Spanning Trees and Cuts

Arash Asadpour
Management Science and Engineering Department, Stanford University,
Stanford, CA 94305, USA
a.asadpour@gmail.com,
Tomás Feder
268 Waverley St., Palo Alto, CA 94301, USA
tomas@theory.stanford.edu,
Amin Saberi
Management Science and Engineering Department, Stanford University,
Stanford, CA 94305, USA
saberi@stanford.edu

May 2, 2009

1 Introduction

Let G be a graph that is the union of r edge-disjoint spanning trees $T_1, \ldots, T_r$ of G. The graph G may have parallel edges. We say that r is the *joint parameter* of G. A cut (A, B, K) in G is a partition of the vertices $V(G)$ into two subsets A and B, such that K is the set of edges $e = (x, y) \in E(G)$ with $x \in A$ and $y \in B$. We say that a spanning tree T of G *achieves fraction* ϵ, if for every cut (A, B, K) in G, we have $|K \cap E(T)| \leq \epsilon |K|$.

2 Achieving fraction $O(\epsilon)$ with joint parameter $r = \log n/\epsilon$

Lemma 2.1 *For any value r of the joint parameter, if we choose $cn \log n$ edges at random, we obtain a subgraph of G that contains a spanning tree of G, with probability at least $1 - 1/c$.*

Proof. Suppose we have obtained a spanning forest F with $t < n - 1$ edges. For each T_i in the representation of G, there exists a spanning forest F_i with $n - 1 - t$ edges such that $E(F_i) \subseteq E(T_i)$ and $F \cup F_i$ is a spanning tree of G. We thus have a choice of at least $(n - 1 - t)r$ edges e that we can select such that $F \cup \{e\}$ is a forest. Call such edges e correct. If we choose e at random, the probability that we choose a correct edge e to increase t to $t + 1$ is thus at least $(n - 1 - t)/(n - 1)$. The expected time number of edges to be chosen until a correct edge e is chosen is thus at most $(n - 1)/(n - 1 - t)$. The total expected time until t reaches n is thus at most

$$\sum_{0 \leq t < n-1} (n - 1)/(n - 1 - t) \leq n \log n.$$

The claim then follows from Markov's inequality. □

Lemma 2.2 *The total number of cuts with $|K| = k$ is at most $rn^{k/r}$.*

Proof. Choose T_i out of $T_1, \ldots, T_r$ such that $|E(T_i) \cap K| \leq k/r$. Specifying $E(T_i) \cap K$ gives the two sets A, B in the cut (up to exchanging A and B) and thus K. The number of ways of choosing at most k/r edges out of $n-1$ is at most $n^{k/r}$. The claim then follows by multiplying $n^{k/r}$ by the number r of choices of T_i. □

Theorem 2.3 *We can achieve fraction $\mu = O(\epsilon)$ with joint parameter $r = \log n/\epsilon$ in randomized expected polynomial time.*

Proof. We can choose $cn \log n$ random edges as in Lemma 2.1 to obtain a spanning tree T with probability at least $1 - 1/c$. We may assume these edges are chosen by selecting each edge independently at random with probability ϵc. We shall choose $\mu = e^2 c\epsilon$. Consider all cuts K with $|K| = k$. Note that $k \geq r$, as cuts must contain at least one edge in each T_i. The number of subsets S of such K of size μk is at most $rn^{k/r}\binom{k}{\mu k} \leq rn^{k/r}(e/\mu)^{\mu k}$ by Lemma 2.2. The probability that T will contain such an S is at most $\epsilon c^{\mu k}$. The result then follows by observing that for each k, the probability that some S is covered by T is at most

$$\begin{aligned}
&\epsilon c^{\mu k} r n^{k/r} (e/\mu)^{\mu k} \\
&\leq e^{\mu \log(c\epsilon) k} k n^{k\epsilon/\log n} e^{k\mu(1-\log\mu)} \\
&\leq e^{k(\epsilon + \mu(1+\log(c\epsilon/\mu)) + \log k} \\
&\leq e^{k(\epsilon-\mu)+\log k} \\
&\leq e^{r\epsilon(1-e^2c)+\log k} \\
&\leq e^{(2-e^2c)\log n} \\
&\leq 1/n^2
\end{aligned}$$

for $\mu = O(\epsilon)$. □

3 Achieving fraction $O(\epsilon)$ with joint parameter $r = \log n/(\epsilon \log\log n)$

The main difficulty with choosing spanning trees T uniformly at random in G is that the probability $p(e)$ of $e \in T$ may not be uniform. For this purpose we shall replace each edge e with some number w of parallel edges $e_1, \ldots, e_w$, and consider that $e \in T$ if $e_i \in T$ for some $1 \leq i \leq w$. We call $w = w(e)$ the *weight* of e, and shall allow $w(e) \geq 0$ to be any nonnegative real number. Thus the probability of a given spanning tree T is proportional to the product of the $w(e)$ over $e \in E(T)$.

Lemma 3.1 *There exists an assignment of weights $w(e)$ to each $e \in E(G)$, the probability that e occurs in T is $p(e) = 1/r$. We can compute in polynomial time weights w_e so that $(1/r)(1-\delta) \leq p(e) \leq (1/r)(1+\delta)$ for $\delta = o(1)$.*

Proof. We consider the objective function $F = \sum_{e \in E(G)} (p(e))^2$, which we shall reduce to its minimum at $p(e) = 1/r$, $F^* = (n-1)r/r^2 = (n-1)/r$. Spanning trees are balanced matroids, and thus satisfy the negative correlation property $p(ef) \leq p(e)p(f)$ for two distinct edges $e, f \in E(G)$, and more generally $p(e_1 \cdots e_s) \leq p(e_1) \cdots p(e_s)$ for distinct edges $e_1, \ldots, e_s \in E(G)$ by induction on the minors of G [1].

Let $p_M > 1/r$ be the largest current $p(e)$ for $e \in G$, and let $1/r < p_m \le p_M$ be some $p(e') > 1/r$ for $e \in G$. Let $e_1, \ldots, e_s$ be the edges with $p(e_i) < p_m$, and let $f_1, \ldots, f_t$ be the edges with $p(f_j) \ge p_m$, with $s + t = (n-1)r$. If we replace each $w(e_i)$ by $Lw(e_i)$ for some $L > 1$, then the new probabilities will satisfy $p'(e_i) \ge p(e_i)$ and $p'(f_i) \le p(f_i)$ by the negative correlation property. The reason is that increasing $w(e_i)$ to $Lw(e_i)$ makes L copies of the spanning trees containing e_i, and $p(f_j|e_i) \le p(f_j|\overline{e_i})$, so $p'(f_j) \le p(f_j)$; similarly, we make first make L copies of all edges, leaving all probabilities the same, and decreasing $Lw(f_j)$ to $w(f_j)$ replaces L copies of the spanning trees containing f_j with a single copy, and $p(e_i|f_j) \le p(e_i|\overline{f_j})$, so $p'(e_i) \ge p(e_i)$.

The probability of a tree containing q of the edges e_i is proportional to L^q, so if L is chosen large enough, with hight probability we will choose T so as to maximize q, and thus so as to minimize $v = n - 1 - q$, the number of edges f_j in T. The minimum v satisfies $v \le t/r$, as one of $T_1, \ldots, T_r$ contains at most t/r of the t edges f_j. Therefore the new probabilities for this large L satisfy $p'(f_1) + \cdots + p'(f_t) \le t/r$, so $p'(f_j) \le 1/r$ for some f_j. Similarly, the corresponding maximum q satisfies $q \ge s/r$, since $q = n - 1 - v \ge n - 1 - t/r \ge s/r$, so the new probabilities for this large L satisfy $p'(e_1) + \cdots + p'(e_s) \ge s/r$, so $p'(e_i) \ge 1/r$ for some e_i.

We may thus increase L starting at $L = 1$ until $p'(e_i) = p'(f_j)$, thus reducing F, for as long as $p_M > 1/r$ can be chosen, so in the end all $p'(e) = 1/r$. Computationally, it suffices to choose $p_m = p(f_j)$ so that $p_m - p(e_i) \ge (p_M - 1/r)/((n-1)r - 1)$ so that the number of iterations of choosing L will be polynomial in n. The computation of probabilities for a given choice of $w(e), e \in E(G)$ can be done by determinants. □

Theorem 3.2 *We can achieve fraction $\mu = O(\epsilon)$ with joint parameter $r = \log n/(\epsilon \log\log n)$ in randomized expected polynomial time.*

Proof. By the negative correlation property, see [1], we have after assigning weights as in Lemma 3.1 that the probability that a set S with μk edges is covered by T is $p(e_1 \cdots e_{k\mu}) \le p(e_1) \cdots p(e_{k\mu}) \le (e/r)^{\mu k}$. As in the proof of Theorem 2.3, the result then follows by observing that for each k, the probability that some S is covered by T is at most

$$
\begin{aligned}
&(e/r)^{\mu k} r n^{k/r} (e/\mu)^{\mu k} \\
&\le (e\epsilon \log\log n/\log n)^{\mu k} k e^{k\epsilon \log\log n + \mu(1-\log\mu)} \\
&\le e^{k(\mu(2+\log\epsilon-\log\mu)+\log\log n(\epsilon-\mu/2))+\log k} \\
&\le e^{-2k\epsilon\log\log n} \\
&\le e^{-2r\epsilon\log\log n} \\
&\le 1/n^2
\end{aligned}
$$

for $\mu = O(\epsilon)$. □

4 No-sink digraphs

A *no-sink digraph* is a digraph where each vertex has out-degree at least 1. In this section, we proceed as in the introduction, except that G is now a graph that is the union of r edge-disjoint spanning connected no-sink digraphs $T_1, \ldots, T_r$ of G. We say that r is the *joint parameter* of G. A cut (A, B, K) in G is a partition of the vertices $V(G)$ into two subsets A and B, such that K is the set of edges $e = (x, y) \in E(G)$ with $x \in A$ and $y \in B$, or with $x \in B$ and $y \in A$. We say that spanning connected no-sink digraph T of G *achieves fraction* ϵ, if for every cut (A, B, K) in G, we have $|K \cap E(T)| \leq \epsilon|K|$.

Theorem 4.1 *For the no-sink digraph problem, we can achieve fraction $\mu = O(\epsilon)$ with joint parameter $r = \log n/(\epsilon \log\log n)$ in randomized expected polynomial time.*

Proof. Select for each vertex v in each T_i a single edge e_i coming out of v, and then select at random one of $e_1, \ldots, e_r$ for v to obtain the spanning connected no-sink digraph T of G. Proceeding as in the proof of Theorem 3.2, consider a cut K with $|K| = k$ and a set S of μk edges. If a vertex v has at least two outgoing edges in S, then at least one of these edges will not be covered by T. The probability that S is covered by T is $p(e_1 \cdots e_{k\mu}) = p(e_1) \cdots p(e_{k\mu}) = (1/r)^{\mu k}$. As in the proof of Theorem 3.2, the result then follows by observing that for each k, the probability that some S is covered by T is at most

$$(1/r)^{\mu k} r n^{k/r} (e/\mu)^{\mu k}$$
$$\leq 1/n^2$$

for $\mu = O(\epsilon)$. To make T connected, consider spanning subtrees T'_i of T_i, find a spanning tree achieving μ from Theorem 3.2, and return the union of T and T' achieving $2\mu = O(\epsilon)$. □

An interesting open problem is the case where the connected spanning no-sink digraphs T_i are each a directed cycle (which traverses all vertices of G in some order), and the spanning no-sink digraph T is required to be connected with each vertex having out-degree 1.

5 *st*-paths

An *arborescence* is a rooted directed tree with all edges oriented away from the root. In this section G is the union of r strongly connected spanning graphs that contain spanning arborescences, $T_1, \ldots, T_r$ of G, where all spanning arborescences of the T_i have the same root vertex s. A strict cut (A, B, K) in G is a partition of the vertices $V(G)$ into two subsets A and B, such that K is the set of edges $e = (x, y) \in E(G)$ with $x \in A$ and $y \in B$. Let t be a specified vertex in G, and let P be a directed path from s to t in G. We say that P *achieves strict fraction* ϵ, if for every strict cut (A, B, K) in G, we have $|K \cap E(P)| \leq \epsilon|K|$.

Theorem 5.1 *For the st-path problem, we can achieve strict fraction $\mu = 2\epsilon$ with joint parameter $r = n^{2/3}/\epsilon$ in deterministic polynomial time.*

Proof. An edge e of G is *thin* if G has at most $p = 1/\epsilon$ edges e' such that e and e' are parallel and in the same direction. Given a vertex v, let $\ell = \ell(v)$ be the minimum number of thin edges in a directed path from s to v. Let $Q(\ell)$ be the set of vertices v such that $\ell(v) = \ell$, and let $q(\ell) = |Q(\ell)|$.

Suppose $Q(\ell)$ is nonempty for some $\ell \geq 1$. The edges from $Q(\ell - 1)$ to $Q(\ell)$ are all thin, otherwise some vertex of $Q(\ell)$ would be in $Q(\ell - 1)$. The number $f(\ell)$ of edges from $Q(\ell - 1)$ to $Q(\ell)$ satisfies $f(\ell) \geq r$, since each T_i contains at least one such edge. Since these $f(\ell)$ edges are thin, we have that $q(\ell - 1)q(\ell) \geq r/p = \epsilon r$, so $q(\ell - 1) + q(\ell) \geq 2\sqrt{\epsilon r}$. If ℓ is chosen largest such that $Q(\ell)$ is nonempty, we have that $n = \sum_{0 \leq i \leq \ell} q(\ell) \geq \ell\sqrt{\epsilon r}$. Therefore $\ell \leq n/\sqrt{\epsilon r} = n^{2/3}$.

It follows that there is a directed path P from s to t that has at most $n^{2/3}$ thin edges. A strict cut (A, B, K) has $k = |K| \geq r = n^{2/3}/\epsilon$ edges across the cut, at least one per T_i. Of these k edges, at most $n^{2/3} \leq k\epsilon$ are thin edges in P, and at most $k/p = k\epsilon$ are edges in P that are not thin, for a total of at most $2k\epsilon$ edges in P across the cut K, so we achieve strict fraction $\mu = 2\epsilon$. □

An interesting special case of the st-path problem is the case where the spanning arborescences of the T_i themselves are spanning st-paths of G.

References

[1] T. Feder and M. Mihail, Balanced Matroids. Proc. 24th Ann. ACM Symp. on Theory of Computing (1992) 26-38.