On the Approximability of the Steiner Tree Problem

Martin Thimm [1]

Humboldt Universität zu Berlin
Institut für Informatik
Lehrstuhl für Algorithmen und Komplexität
D-10099 Berlin
thimm@informatik.hu-berlin.de

Abstract

We show that it is not possible to approximate the minimum Steiner tree problem within $1 + \frac{1}{162}$ unless $RP = NP$. The currently best known lower bound is $1 + \frac{1}{400}$. The reduction is from Håstad's nonapproximability result for maximum satisfiability of linear equation modulo 2. The improvement on the nonapproximability ratio is mainly based on the fact that our reduction does not use variable gadgets. This idea was introduced by Papadimitriou and Vempala.

Key words: Minimum Steiner tree, Approximability, Gadget reduction, Lower bounds.

1 Introduction

Suppose that we are given a graph $G = (V, E)$, a metric given by edge weights $c : E \to \mathbb{R}$, and a set of required vertices $T \subset V$, the *terminals*. The minimum Steiner tree problem consists of finding a subtree of G of minimum weight that spans all vertices in T.

The Steiner tree problem in graphs has obvious applications in the design of various communications, distributions and transportation systems. For example the wire routing phase in physical VLSI–design can be formulated as a Steiner tree problem [12]. Another type of problem that can be solved with the help of Steiner trees is the computation of phylogenetic trees in biology. For more examples see [10].

[1] Graduate School “Algorithmische Diskrete Mathematik”, supported by Deutsche Forschungsgemeinschaft, grant 219/3.

The Steiner tree problem is well known to be NP-complete even in the very special cases of Euclidian or rectilinear metric. Arora [1] has shown that Euclidian and rectilinear Steiner tree problems admit a polynomial time approximation scheme, i.e. they can be approximated in polynomial time up to a factor of $1+\epsilon$ for any constant $\epsilon > 0$. In contrast to these two special cases the Steiner tree problems is known to be APX-complete [2,5] which means that unless $P = NP$ there does not exist a polynomial time approximation scheme for this problem.

During the last ten years a lot of work has been done on designing approximation algorithms for the Steiner tree problem [18,4,16,11,9]. The currently best approximation ratio is 1.550 and is due to Robins and Zelikovsky [17]. For more details on approximation algorithms for the Steiner tree problem see [7].

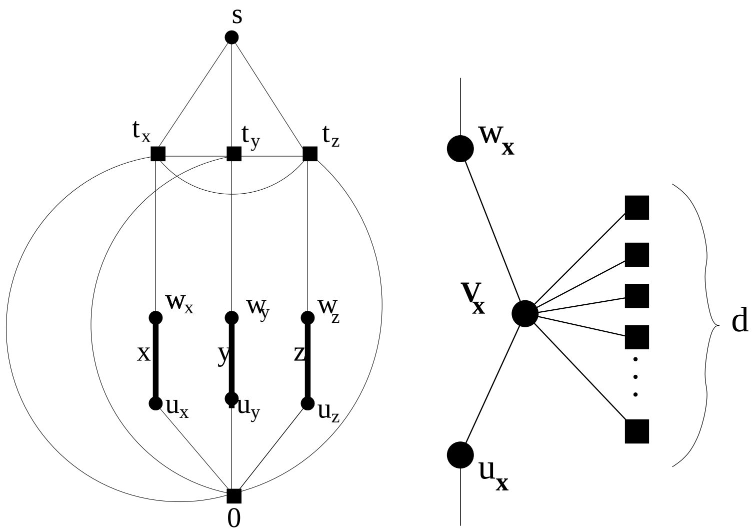
But very little is known about lower bounds. The presently best known lower bound is 1.0025 and follows from a nonapproximability result for $VERTEX$-$COVER$ in graphs of bounded degree [3]. We improve this bound to 1.0062.

The improvement on the nonapproximability ratio is mainly based on the fact that our reduction does not use variable gadgets. This idea was introduced by Papadimitriou and Vempala [15]. They prove that the (symmetric) traveling salesman problem cannot be approximated within $\frac{129}{128}$, unless $P = NP$.

We reduce from Håstad's nonapproximability result for maximum satisfiability of linear equations modulo 2 with three variables per equation, MAX-$E3$-LIN-2, [8]. Our construction uses two types of gadgets. There is one "equation gadgets" for each equation in the MAX-$E3$-LIN-2 instance. We also have "edge gadgets" which connect the nodes of the equation gadgets corresponding to literals x and $\bar{x}$ in a special way that is induced by some d-regular bipartite graph. We prove that if this graph is an expander then every optimal Steiner tree for this instance has a special structure which allows us to derive a legal truth assignment for the MAX-$E3$-LIN-2 instance (without relying on variable gadgets).

Our nonapproximability ratio only depends on the parameters of the above mentioned expanders, namely the degree d and the expansion coefficient c. In [19] the existence of such graphs is proved by counting arguments. We adapt the proof of Sarnak's result [19] for our purpose. Stronger expanders would automatically lead to a better nonapproximability ratio.

The rest of the paper is organized as follows: In Section 2 we describe the reduction in detail and in Section 3 we prove our main theorem. The existence of the expander graphs needed in the reduction is proved in Section 4. In the last Section we shortly discuss some ideas how to possibly improve our main result.



(a) equation gadget

(b) edge gadget for the edge labeled x

Fig. 1. The gadgets

2 The Reduction

2.1 The Graph

We reduce from Håstad's nonapproximability result for maximum satisfiability of linear equations modulo 2 with three variables per equation, MAX-$E3$-LIN-2, [8]: Given a set of n linear equations over $\mathbb{Z}_2$, with exactly k variables in each equation, find the maximal number of equations that can be satisfied by any assignment.

As already mentioned in [15] we can state Håstad's result [8] as follows:

Theorem 1 *[8] For every $\epsilon > 0$ there is an integer k such that it is NP-hard to tell whether a set of n linear equations modulo 2 with three variables per equation and with $2k$ occurrences of each variable has an assignment that satisfies $n(1-\epsilon)$ equations, or has no assignment that satisfies more than $n(\frac{1}{2}+\epsilon)$ equations.*

We start with such an instance of MAX-$E3$-LIN-2, namely a set of n linear equations modulo 2, where each equation has exactly three literals and with exactly $2k$ occurrences of each variable. We may also assume that each variable appears exactly k times negated and k times unnegated and also that all equations are of the form $x+y+z=1$. The latter condition can be enforced by flipping some literals, the former by adding three copies of each equation with all pairs of literals negated. (e.g. $x+y+z=0$ may be transformed to $x+y+\bar{z}=1$ and we add $\bar{x}+\bar{y}+\bar{z}=1$, $\bar{x}+y+z=1$ and $x+\bar{y}+z=1$.)

We construct an instance for the Steiner tree problem as follows:

The graph consists of several gadgets. For each equation there is an *equation*

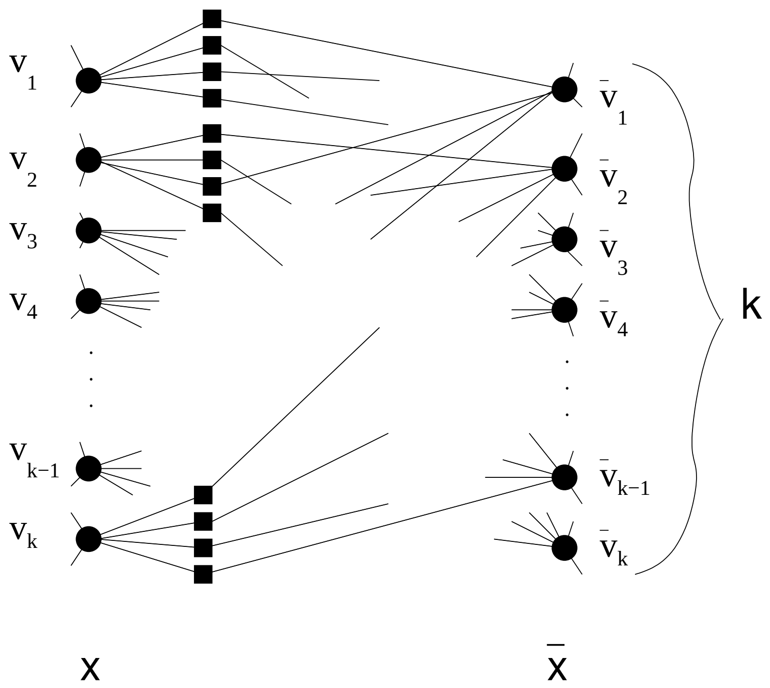


Fig. 2. edge gadgets of variable x after identification of the terminals

gadget (see Figure 1(a)). The nodes drawn as boxes are terminals. All equation gadgets share the node 0. The bold edges labeled with x, y and z correspond to the variables in an equation of the form $x + y + z = 1$ and are not really edges of the graph but represent whole substructures as shown in Figure 1(b). (The edge weights and the value of d in the edge gadget will be specified later.)

As in [15] we do not use "variable gadgets" to assure a correct assignment of the literals and their opposite. The construction will enforce this implicitly. By assumption each variable appears exactly k times negated and k times unnegated. Consider one variable, say x. In our graph we then have k edge gadgets that correspond to the occurrences of x and k edge gadgets for the occurrences of $\bar{x}$. We now connect these edge gadgets in the following way: Suppose we are given a bipartite d-regular graph $H = (A \cup B, E)$, $|A| = |B| = k$. Now identify a terminal in the i-th edge gadget corresponding to an occurrence of x with a terminal in the j-th edge gadget corresponding to a occurrence of $\bar{x}$ iff the i-th vertex in A is connected to the j-th vertex in B $(1 \leq i, j \leq k)$ (see Figure 2).

We will see in what follows that the only thing we need to know about H is that it is an expander, i.e. with $V(H) = A \cup B$, $|A| = |B| = k$ we have that for all $S \subset A$ with $|S| \leq \frac{k}{2}$: $|\Gamma(S)| \geq c|S|$ for some $c > 1$ (where $\Gamma(S)$ denotes the set of neighbours of S).

If we now define the edge weights appropriately, our graph will have some useful properties. First, a truth assignment for our MAX-$E3$-LIN-2 instance directly yields a solution for the Steiner tree instance which has a "nice" structure: All subtrees of the resulting Steiner tree which correspond to satisfied equations have the same length l_s and a special structure. Also the subtrees which correspond to the equations that are not satisfied have some special structure and the same length l_{ns}. Additionally we have that $l_{ns} = l_s + const$. By this the length of the Steiner tree reflects the number of equations that are not satisfied.

Second, we will see that we always may assume that an optimal Steiner tree has a simple special structure which makes it easy to define the corresponding truth assignment.

2.2 The Edge Weights

We have already defined the graph for the Steiner tree instance. Let us now define the edge weights. First, we want to guarantee that in any optimal Steiner tree all the terminals in the edge gadgets have degree 1. We define the weight of any edge in any edge gadget which is incident to a terminal to be b. Then it suffices to have that the weight of a path $(0, u_\bullet, v_\bullet)$ and $(v_\bullet, w_\bullet, t_\bullet)$ is at most b. (Whenever we write $u_\bullet$, $w_\bullet$ etc. this always means u_x, w_x etc. for all variables x. To indicate that the variables should be distinct we write $c(t_\bullet, t_\star)$ etc. for $c(t_x, t_y)$ for all variables x and y with $x \neq y$.)

Suppose now we are given a truth assignment to our MAX-$E3$-LIN-2 instance. We define the corresponding Steiner tree as follows: Consider the nodes $v_\bullet$ in the edge gadgets corresponding to variables with assigned truth value 1 and connect them to all their adjacent terminals. Now connect all these vertices $v_\bullet$ to the remaining nodes in the equation gadgets to get a Steiner tree (using optimal subtrees in the equation gadgets). If we consider an equation gadget there are four possible cases: None, one, two or three of the nodes $v_\bullet$ have to be connected to the tree. The cases with one or three of these nodes correspond to equations that are satisfied. We want them to have partial solutions (subtrees of the equation gadgets) of the same weight l_s. If none or two nodes have to be connected to the tree, the corresponding equations were not satisfied. The subtrees in these equation gadgets should also have the same weight, l_{ns}.

For technical reasons which will become clear during the proof of the main theorem we add up the edge weights of the edge gadgets and those of the equation gadgets separately. If we now define $c(\mathbf{0}, u_\bullet) = a$, $c(u_\bullet, v_\bullet) = b - a$, $c(v_\bullet, w_\bullet) = 0$ and $c(w_\bullet, t_\bullet) = b$ we have fulfilled the above mentioned condition. Furthermore, connecting a vertex $v_\bullet$ to the tree always costs b and without loss of generality we may assume that connecting $v_\bullet$ to the tree is the same as connecting $u_\bullet$ to the tree (that is to pay a in the equation gadget and $b-a$ in the edge gadget). Let us now define the remaining edge weights: $c(\mathbf{0}, t_\bullet) = f$, where $f = 2a + b$, $c(s, t_\bullet) = e$ where $e = \frac{2}{3}(a + b)$ and $c(t_\bullet, t_\star) = f$ (see Figure 3).

It is easy to find the optimal partial solution in all the four cases. They are displayed in Figure 4. The subtrees corresponding to satisfied equations (one or three variables have truth value 1) have weight $3a + 3b$ (Figure 4, right

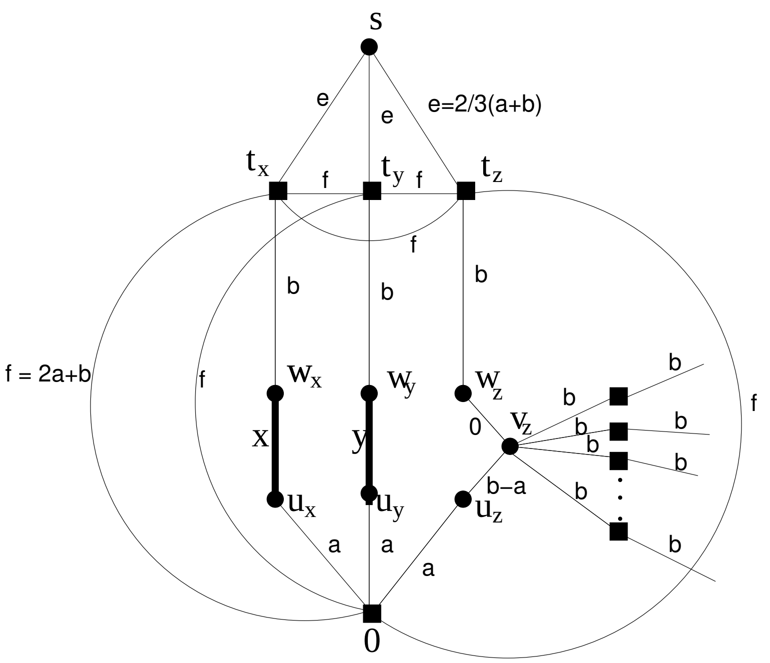


Fig. 3. edge weights (edge gadget for z inserted)

hand side), those corresponding to equations that are not satisfied have weight $4a + 3b$ (Figure 4, left hand side). Remember that we count the weights in the edge gadgets separately.

The way these numbers a and b are related will become clear at the end of the proof of the main theorem. We will also see how d, the number of terminals in the edge gadgets, has to be chosen.

We have now fully described our Steiner tree instance and prove our main theorem.

3 Proof of the Main Theorem

Theorem 2 *No polynomial time approximation algorithm for the minimum Steiner tree problem can have a peformance ratio below* 1.0062*, unless* $RP = NP$*.*

PROOF. Given a truth assignment to our MAX-$E3$-LIN-2 instance. We have already seen in the last Section how to define the corresponding Steiner tree. The weight of this Steiner tree consists of the weight of the edge gadgets and the weight of the equation gadgets.

There are $3n$ edge gadgets, half of them correspond to literals with truth value 1 (since each variable appears the same number of times and also the same number of times negated and unnegated). The weight of an edge gadget is $b-a+db$ so that the total weight of all edge gadgets sums up to $\frac{3}{2}n(b-a+db)$.

The subtrees in the equation gadgets corresponding to satisfied equations have weight $3a + 3b$, those corresponding to unsatisfied equations have weight

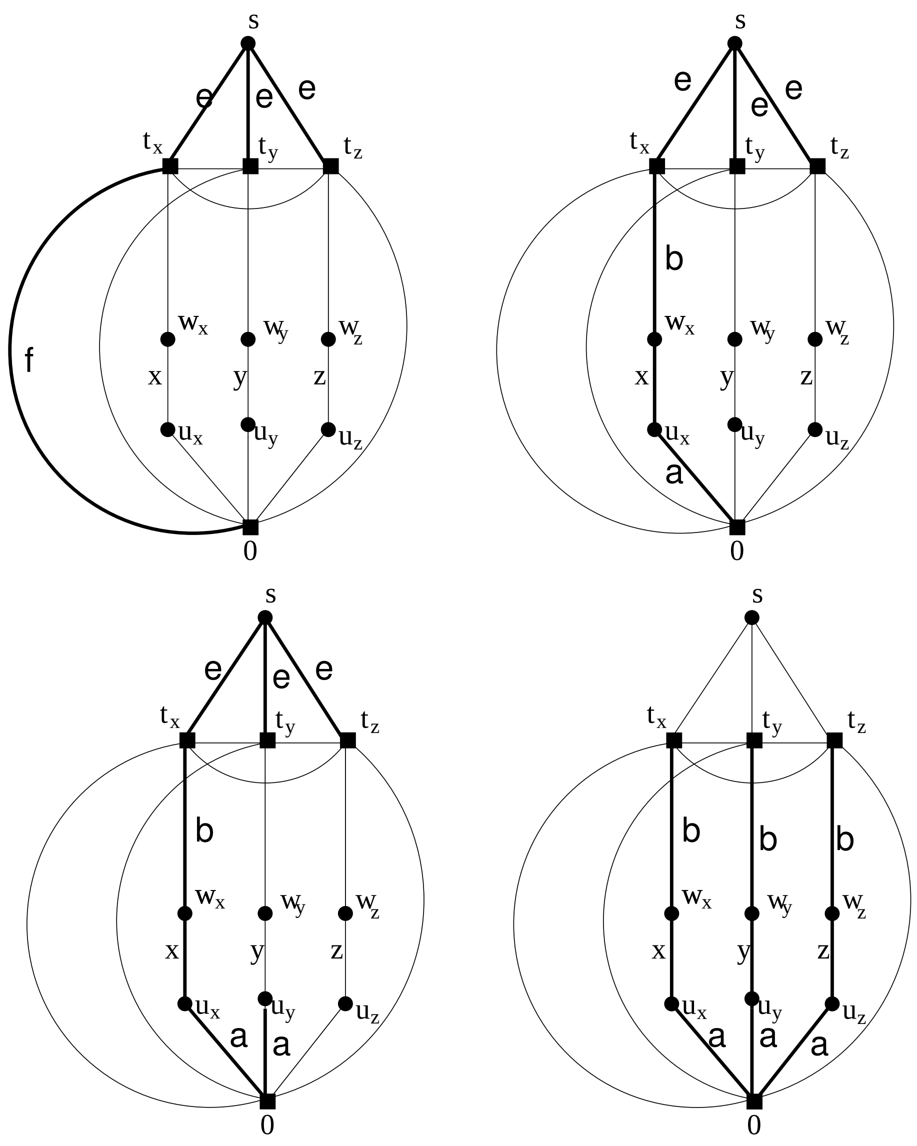


Fig. 4. The four cases of optimal subtrees

$4a + 3b$.

Suppose the truth assignment satisfies all but M equations. Then the Steiner tree constructed above has weight $\frac{3}{2}n(b - a + db) + n(3a + 3b) + Ma$.

We call such a tree *standard* (for this assignment), i.e. for the collection of edge gadget of any variable, say x, it is true that either all the nodes v_x corresponing to the occurrences of the literal x are Steiner nodes or all the nodes $v_{\bar{x}}$ corresponing to the occurrences of the literal $\bar{x}$ are of that kind.

To prove the theorem we have to show that an optimal Steiner tree is standard (for some assignment). This is done in Lemma 3.

Now we use Håstad's result: for every $\epsilon > 0$ it is NP-hard to decide whether a set of n linear equations modulo 2 with three variables per equation has an assignment that satisfies $n(1-\epsilon)$ equations, or has no assignment that satisfies more than $n(\frac{1}{2} + \epsilon)$ equations. By our reduction the same is true for standard trees of length $n(\frac{3}{2}(db+(b-a))+(3a+3b)-\epsilon)$ and $n(\frac{3}{2}(db+(b-a))+(3a+3b)+\frac{1}{2}a+\epsilon)$.

We get an nonapproximability ratio of

$$r = \frac{\frac{3}{2}n(db+(b-a))+n(3a+3b)+\frac{1}{2}na}{\frac{3}{2}n(db+(b-a))+n(3a+3b)} \ .$$

If we set $a = tb$ and cancel n and b, this reads as

$$r = \frac{\frac{3}{2}(d+1-t)+3(1+t)+\frac{1}{2}t}{\frac{3}{2}(d+1-t)+3(1+t)} \ .$$

With $a = tb$ and $c = \frac{b}{b-2a}$ (see Lemma 3 below), we also get $c = \frac{1}{1-2t}$ ($t \leq \frac{1}{3}$, since $t = \frac{c-1}{2c}$ and $c \leq 2$). Plugging this in, we finally get

$$\begin{aligned} r &= \frac{\frac{3}{2}(d+1-\frac{c-1}{2c})+3(1+\frac{c-1}{2c})+\frac{1}{2}\frac{c-1}{2c}}{\frac{3}{2}(d+1-\frac{c-1}{2c})+3(1+\frac{c-1}{2c})} \\ &= 1+\frac{\frac{c-1}{2c}}{3(d+3+\frac{c-1}{2c})} \ . \end{aligned} \tag{1}$$

With $d = 6$ and $c = 1.5144$ the theorem follows (see Lemma 4). $\square$

Lemma 3 *Every optimal Steiner tree can be transformed into a standard tree without increasing the weight.*

PROOF. Given an optimal Steiner tree. Remember that we may assume that all terminals in the edge gadgets have degree 1.

Consider all edge gadgets corresponding to one variable, say x and $\bar{x}$ and the graph H_x which shows the identification of the terminals of the edge gadgets. We partition the node set $V(H_x)$ into three classes A_x, T_x and $B_{\bar{x}}$ (see Figure 5). We then partition A_x into $C1$, which are the nodes in A_x that are Steiner nodes of the tree, and $U1$, all other nodes in A_x. In the same way we partition $B_{\bar{x}}$ into $C2$ and $U2$ (see Figure 5).

Without loss of generality let $|U1| \leq |U2|$. Consider the following modification of the Steiner tree: (see Figure 5) All nodes in A_x and none in $B_{\bar{x}}$ are Steiner nodes after this step, all terminals in the edge gadgets are linked to these Steiner nodes. The subtrees of those equation gadgets which contain the nodes lying in $U1$ and $C2$ are changed. We construct them according to the new conditions (number of vertices $v_\bullet$ to be connected to the tree) completely new (the result is on of the subtrees in Figure 4).

We claim that if we modify the Steiner tree in this way , the weight of the resulting new tree does not increase. It is easy to see that if we have done this for all variables, one after another, the result is a tree in standard form.

It remains to show that this modification does not increase the weight of the tree. Consider H_x. In order to connect the Steiner nodes $v_\bullet$ in H_x to the

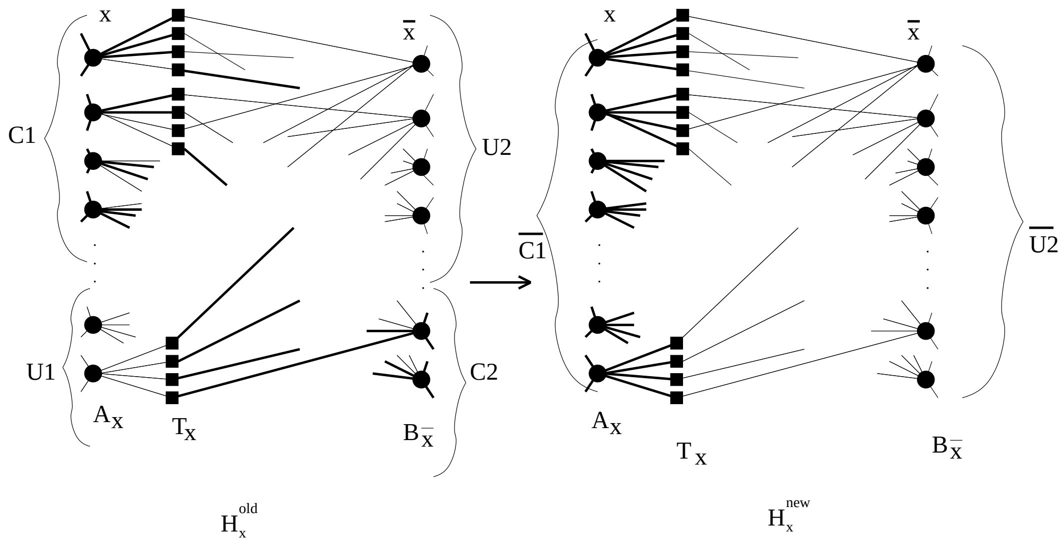


Fig. 5. Modification of the edges of the Steiner tree in the edge gadgets

tree we have to pay a weight of $b - a$ for each of them. So the cost to do this in the old tree is $(b - a)(|C1| + |C2|)$. In the new tree we only need $k(b - a)$ with $k = |U1| + |C1|$, hence we gain $(b - a)(|C2| - |U1|)$.

On the other hand we have to look at the newly constructed subtrees of the equation gadgets, which contain nodes out of $U1$ and $C2$. Since the number of nodes $v_\bullet$ which have to be connected to the subtrees has changed, in each such case we may have to pay an extra cost of a (see Figure 4). So the total extra cost can be bounded by $a(|U1| + |C2|)$.

We are done if we can show that

$$(|C2| - |U1|)(b - a) \geq (|U1| + |C2|)a \ .$$

This inequality can be rearranged to stand as

$$|C2| \geq \frac{b}{b - 2a}|U1| \ . \tag{2}$$

Let $\Gamma(U1)$ be the set of nodes in $B_{\bar{x}}$ which are reachable from $U1$ by paths of length 2. We get that $\Gamma(U1) \subset C2$. If we can show that

$$\Gamma(U1) \geq \frac{b}{b - 2a}|U1|, \tag{3}$$

we have proven (2).

But if we consider H_x as a bipartite graph with node sets A_x and $B_{\bar{x}}$ and recall that $|U1| \leq \frac{k}{2}$, (3) is just a typical expander condition with expansion constant $c = \frac{b}{b-2a}$.

It is known that such bipartite regular expanders exist, as well as how to construct them probabilistically. The following Section will deal with this question in detail. To get our nonapproximability result we use 6-regular expanders with $c = 1.5144$ (see Lemma 4). □

4 Expanders

Let $G = (I \cup O, E)$ a bipartite d-regular graph, $d \geq 3$, $|I| = |O| = n$. We call G a (n, d, c)-*expander*, if for all $A \subset I$, $|A| \leq \frac{n}{2}$ and some $2 \geq c > 1$

$$\Gamma(A) \geq c|A|. \tag{4}$$

Explicit constructions of such expanders seem to be very hard. This was first done by Margulis in [14]. He could prove that the constant c arising in his construction is bounded away from 1. In [6] such a constant was explicitly calculated by Gabber and Galil, before Lubotzky, Philips and Sarnak [13] came up with a different construction and a better constant c.

Compared to this it is quite easy to prove the existence of linear expanders by counting arguments.

Following the ideas of [19] we will do this in greater detail. The main result of this Section is

Lemma 4 *For sufficiently large n there exist (n, d, c)-expanders for*

$$d > max\left\{c + \frac{3}{2}, \frac{2}{2-c}, \frac{\frac{c}{2}\ln(\frac{c}{2}) + (1-\frac{c}{2})\ln(1-\frac{c}{2}) - \ln(2)}{\frac{c}{2}\ln(c) - \frac{c-1}{2}\ln(c-1) - \ln(2)}\right\}. \tag{5}$$

PROOF. Let $I = O = \{1, 2, \ldots, n\}$ and X be a bipartite graph with $V(X) = I \cup O$. Consider d permutations $\pi_1, \pi_2, \ldots, \pi_d$ of I and connect each $j \in I$ with all $\pi_r(j) \in O, 1 \leq r \leq d$. We get a bipartite d-regular (multi)graph. Let us call $\pi = (\pi_1, \pi_2, \ldots, \pi_d)$ *bad*, if there is a set $A \subset I$, $|A| \leq \frac{n}{2}$ and a set $B \subset O$, $|B| = c|A|$ such that $\pi_j(A) \subset B$ for all $j = 1, 2, \ldots, d$.

We will bound the number of bad $\pi's$ from above. For some given A and B, $|A| = t \leq \frac{n}{2}, |B| = ct$, the number of bad $\pi's$ is

$$(ct(ct-1)(ct-2)\ldots(ct-t+1)(n-t)!)^d = \left(\frac{(ct)!(n-t)!}{(ct-t)!}\right)^d.$$

So the total number of bad $\pi's$, BAD, over all A and B is given by

$$BAD = \sum_{t \leq \frac{n}{2}} \binom{n}{t}\binom{n}{ct}\left(\frac{(ct)!(n-t)!}{(ct-t)!}\right)^d. \tag{6}$$

Since there are $(n!)^d$ $\pi's$ in total, we are interested in the function

$$R(t) = \binom{n}{t}\binom{n}{ct}\left(\frac{(ct)!(n-t)!}{(ct-t)!n!}\right)^d, 1 \leq t \leq \frac{n}{2}, \tag{7}$$

especially for $n \to \infty$.

$R(t) \to 0$ for $n \to \infty$ means that if we pick a π randomly to construct our graph X, we will get a (n, d, c)-expander with very high probability.

We proceed as follows: To analyse the function $R(t)$ we consider its continuous version and its first derivative. We show that $R'(t) \approx R(t)\ln(f(t))$

for some function f. By looking at f we can prove that $R(t)$ decreases in the interval $1 \leq t \leq \alpha$ for some α that only depends on c and d and inreases for all other values of t up to $\frac{n}{2}$. We bound the values of $R(1)$, $R(\sqrt{n})$ and $R(\frac{n}{2})$ to give upper bounds on (6) by $\sqrt{n}R(1) + \alpha n R(\sqrt{n}) + \frac{n}{2}R(\frac{n}{2})$.

Let once again

$$BAD = \sum_{t \leq \frac{n}{2}} \binom{n}{t}\binom{n}{ct}\Big(\frac{(ct)!(n-t)!}{(ct-t)!}\Big)^d$$

and

$$R(t) = \binom{n}{t}\binom{n}{ct}\Big(\frac{(ct)!(n-t)!}{(ct-t)!n!}\Big)^d, 1 \leq t \leq \frac{n}{2}.$$

We use that $n! = \Gamma(n+1)$ for $n \in \mathbb{N}$ and $\Gamma'(t)/\Gamma(t) = \gamma(t)$ with $\gamma(n) = \gamma + \sum_{k=1}^{n-1} \frac{1}{k}$ for $n \in \mathbb{N}$ and $\gamma = 0.577216\ldots$. Furthermore we have $\ln(n) = \sum_{k=1}^{n} \frac{1}{k} + \gamma + o(1)$.

To shorten the notation we set

$$\begin{aligned}
f_1(t) &:= \binom{n}{t} = \frac{n!}{t!(n-t)!} \\
f_2(t) &:= \binom{n}{ct} = \frac{n!}{(ct)!(n-(ct))!} \\
f_3(t) &:= f_1(t)f_2(t) \\
g(t) &:= \Big(\frac{(ct)!(n-t)!}{(ct-t)!n!}\Big), \text{and} \\
R(t) &= f_3(t)g(t)^d = f_1(t)f_2(t)g(t)^d.
\end{aligned}$$

Hence we have

$$R'(t) = (f_1'(t)f_2(t) + f_1(t)f_2'(t))g(t)^d + df_3(t)g^{d-1}(t)g'(t).$$

We consider these functions one after another:

Since

$$\begin{aligned}
f_1(t) &= \frac{\Gamma(n+1)}{\Gamma(t+1)\Gamma(n-t+1)} \text{ we get} \\
f_1'(t) &= \Gamma(n+1)\Big(\frac{-\Gamma(t+1)\gamma(t+1)\Gamma(n-t+1) - \Gamma(t+1)\Gamma(n-t+1)\gamma(n-t+1)}{(\Gamma(t+1)\Gamma(n-t+1))^2}\Big) \\
&= \frac{\Gamma(n+1)}{\Gamma(t+1)\Gamma(n-t+1)}(\gamma(n-t+1) - \gamma(t+1)) \\
&= f_1(t)\tilde{f}_1(t) \text{ ,where} \\
\tilde{f}_1(t) &:= \gamma(n-t+1) - \gamma(t+1).
\end{aligned}$$

Similarly we get

$$\begin{aligned}
f_2'(t) &= f_2(t)\tilde{f}_2(t) \text{ ,where} \\
\tilde{f}_2(t) &:= c(\gamma(n-ct+1)-\gamma(ct+1), \\
g'(t) &= g(t)\tilde{g}(t) \text{ ,where} \\
\tilde{g}(t) &:= c\gamma(ct+1)-\gamma(n-t+1)-(c-1)\gamma(t(c-1)+1).
\end{aligned}$$

Putting things together yields

$$\begin{aligned}
R'(t) &= R(t)(\tilde{f}_1(t)+\tilde{f}_2(t)+d\tilde{g}(t)) = R(t)\tilde{f}(t) \text{ ,where} \\
\tilde{f}(t) &:= \tilde{f}_1(t)+\tilde{f}_2(t)+d\tilde{g}(t)
\end{aligned}$$

Using the facts above we can write $\tilde{f}(t)$ as

$$\begin{aligned}
\tilde{f}(t) &= \ln(n-t)-\ln(t)+c(\ln(n-ct)-\ln(ct)) \\
&\qquad + d(c\ln(ct)-\ln(n-t)-(c-1)\ln(t(c-1)))+o(1) \\
&= \ln((n-t)^{1-d})-\ln(t)+\ln((ct)^{(c(d-1))}) \\
&\qquad + c\ln((n-ct)^c)-\ln((t(c-1))^{d(c-1)})+o(1) \\
&= \ln\left(\frac{(ct)^{c(d-1)}(n-ct)^c}{(n-t)^{d-1}t(t(c-1))^{d(c-1)}}\right)+o(1) \\
&= \ln\left(t^{d-c-1}\frac{(n-ct)^c}{(n-t)^{d-1}}\frac{c^{c(d-1)}}{(c-1)^{d(c-1)}}\right)+o(1) \\
&= \ln(f(t))+o(1) \qquad \text{where} \\
f(t) &:= t^{d-c-1}\frac{(n-ct)^c}{(n-t)^{d-1}}\frac{c^{c(d-1)}}{(c-1)^{d(c-1)}} \\
&= t^{d-c-1}\frac{(n-ct)^c}{(n-t)^{d-1}}\text{const}(c,d) \qquad \text{where } \text{const}(c,d) := \frac{c^{c(d-1)}}{(c-1)^{d(c-1)}}
\end{aligned}$$

If we set $t = xn$, $0 < x \leq \frac{1}{2}$ we can write f as a function of x

$$f(x) = x^{d-c-1}\frac{(1-cx)^c}{(1-x)^{d-1}}\text{const}(c,d)$$

and

$$\begin{aligned}
\frac{f'(x)}{\text{const}(c,d)} &= (d-c-1)x^{d-c-2}\frac{(1-cx)^c}{(1-x)^{d-1}} \\
&\qquad + x^{d-c-1}\left(\frac{c(1-cx)^{c-1}(-c)(1-x)^{d-1}+(1-cx)^c(d-1)(1-x)^{d-2}}{(1-x)^{2(d-1)}}\right) \\
&= \frac{x^{d-c-2}(1-cx)^{c-1}}{(1-x)^d}\Big((d-c-1)(1-x)(1-cx) \\
&\qquad - c^2x(1-x)+(d-1)x(1-cx)\Big) \\
&= \frac{x^{d-c-2}(1-cx)^{c-1}}{(1-x)^d}\Big((d-c-1)+xc(2-d)\Big).
\end{aligned}$$

From this we conclude that f has its extremal points at $x = \frac{d-c-1}{c(d-2)}$ and $x = \frac{1}{c}$. If we restrict our attention to the case where $d \geq \frac{2}{2-c}$ (we will later on see that this is not really a restriction) it follows that both extremal points are outside the interval $0 < x \leq \frac{1}{2}$. This means that f increases for all x in this interval.

Since $f(t) < 1$ for $t = 1$ we conclude that $R(t)$ decreases at the beginning of the interval $0 < t \leq \frac{n}{2}$. To be able to prove Lemma 4 for small values of d, such as $d = 3$ or $d = 4$, we have to show that $R(t)$ decreases not only for small values of t, but in the whole interval $0 < t < \alpha n$ for some α that only depends on c and d. We claim that there is some $m \in \mathbb{N}$ such that with $\hat{x} = \frac{1}{m}$ we have $f(\hat{x}) < 1$. Indeed,

$$\begin{aligned}
f\Big(\frac{1}{m}\Big) &= \Big(\frac{1}{m}\Big)^{d-c-1} \frac{(1-\frac{c}{m})^c}{(1-\frac{1}{m})^{d-1}} const(c,d) \\
&= \frac{(\frac{1}{m})^{d-c-1}}{(\frac{m-1}{m})^{d-c-1}} \frac{(\frac{m-c}{m})^c}{(\frac{m-1}{m})^c} const(c,d) \\
&= \Big(\frac{1}{m-1}\Big)^{d-c-1} \Big(\frac{m-c}{m}\Big)^c const(c,d) \\
&< \Big(\frac{1}{m-1}\Big)^{d-c-1} const(c,d)
\end{aligned}$$

Hence $f(\frac{1}{m}) < 1$ if $\ln(m-1) > \frac{\ln(const(c,d))}{d-c-1}$.

We bound $R(\sqrt{n})$ by

$$\begin{aligned}
R(\sqrt{n}) &= \binom{n}{\sqrt{n}}\binom{n}{c\sqrt{n}}\Big(\frac{(c\sqrt{n})!(n-\sqrt{n})!}{n!((c-1)\sqrt{n})!}\Big)^d \\
&\leq (e\sqrt{n})^{\sqrt{n}}(\frac{e}{c}\sqrt{n})^{c\sqrt{n}}\Big(\frac{(c\sqrt{n})^{\sqrt{n}}}{(n-\sqrt{n})^{\sqrt{n}}}\Big)^d \\
&\leq e^{(1+c)\sqrt{n}}c^{(d-c)\sqrt{n}}\Big(\frac{1}{1-\epsilon}\Big)^{\sqrt{n}}\sqrt{n}^{(1+c-d)\sqrt{n}} \\
&\text{since for any fixed } \epsilon > 0 \\
&\text{we know that} \quad n - \sqrt{n} > n(1-\epsilon) \text{ for sufficiently large } n
\end{aligned}$$

$$\text{Hence } nR(\sqrt{n}) \to 0 \text{ for } n \to \infty \text{ for all } d > 1+c+\epsilon', \epsilon' > 0. \tag{8}$$

To bound $R(\frac{n}{2})$ we write

$$
\begin{aligned}
R\Big(\frac{n}{2}\Big) &= \binom{n}{\frac{n}{2}}\binom{n}{\frac{cn}{2}}\Big(\frac{(\frac{cn}{2})!(\frac{n}{2})!}{((c-1)\frac{n}{2})!n!}\Big)^d \\
&\approx \Big(\Big(\frac{1}{2}\Big)^{-\frac{1}{2}}\Big(\frac{1}{2}\Big)^{-\frac{1}{2}}\Big)^n\Big(\Big(\frac{c}{2}\Big)^{-\frac{c}{2}}\Big(1-\frac{c}{2}\Big)^{-(1-\frac{c}{2})}\Big)^n\Big(\frac{c^{\frac{c}{2}}2^{\frac{c-1}{2}}}{2^{\frac{c}{2}}2^{\frac{1}{2}}(c-1)^{\frac{c-1}{2}}}\Big)^{dn} \\
&= \Big(2\Big(\frac{c}{2}\Big)^{-\frac{c}{2}}\Big(1-\frac{c}{2}\Big)^{-(1-\frac{c}{2})}\Big)^n\Big(\frac{c^{\frac{c}{2}}}{2(c-1)^{\frac{c-1}{2}}}\Big)^{dn},
\end{aligned}
$$

from which we conclude that $nR(\frac{n}{2}) \to 0$ for $n \to \infty$, if

$$
d > \frac{\frac{c}{2}\ln(\frac{c}{2}) + (1-\frac{c}{2})\ln(1-\frac{c}{2}) - \ln(2)}{\frac{c}{2}\ln(c) - \frac{c-1}{2}\ln(c-1) - \ln(2)}. \tag{9}
$$

Finally we know that

$$
R(1) \le n\frac{n^c}{c!}\Big(\frac{c!(n-1)!}{(c-1)!n!}\Big)^d \le \frac{c^d}{n^{d-c-1}},
$$

so together with (8) and (9) this implies

$$
\begin{aligned}
BAD &\le \sqrt{n}R(1) + \frac{n}{2}(R(\sqrt{n}) + R(\frac{n}{2})) \\
&\le \frac{c^d}{n^{d-c-\frac{3}{2}}} + \frac{n}{2}\Big(R(\sqrt{n}) + R(\frac{n}{2})\Big) \to 0 \text{ for } n \to \infty, \\
&\text{if } d > max\Big\{c+\frac{3}{2}, \frac{2}{2-c}, \frac{\frac{c}{2}\ln(\frac{c}{2}) + (1-\frac{c}{2})\ln(1-\frac{c}{2}) - \ln(2)}{\frac{c}{2}\ln(c) - \frac{c-1}{2}\ln(c-1) - \ln(2)}\Big\}.
\end{aligned}
$$

□

5 Discussion

d	3	4	5	6	7	8	10	15	50	100
c	1.162	1.310	1.425	1.514	1.583	1.637	1.716	1.821	1.954	1.978

Table 1
Some values for c and d in (5)

Table 1 may help to see the numbers hidden in (5). The general form (1) of our main result – the nonapproximability ratio is just a function of c and d – immediately gives better results if better expanders are found. For example, an $(n, 5, \frac{7}{4})$-expander would give a ratio of 1.011.

We believe that using this method a nonapproximability ratio of about 1.01 is within reach.

References

[1] S. Arora, Polynomial time approximation schemes for the Euclidian TSP and other geometric problems, Proceedings of the 37th Annual Symposium on Foundations of Computer Science (1996), pp.2–11.

[2] S. Arora, C. Lund, R. Motwani, M. Sudan, M. Szegedy, Proof verification an hardness of approximation problems, Proceedings of the 33rd Annual Symposium on Foundations of Computer Science (1992), pp. 14–23.

[3] P. Berman, M. Karpinski , On Some Tighter Inapproximability Results, Further Improvements , Electronic Colloquium on Computational Complexity, Report No.65 (1998)

[4] P. Berman, V. Ramaiyer, Improved approximations for the Steiner tree problem, Journal of Algorithms 17 (1994), pp.381–408.

[5] M. Bern, P. Plassmann, The Steiner Problem with edge lengths 1 an 2, Information Processing Letters 32 (1989), pp. 171-176.

[6] O. Gabber, Z. Galil, Explicit construction of linear-sized superconcentrators, J. Comput. System Sci. 22 (1981) pp.407-420

[7] C. Gröpl, S. Hougardy, T. Nierhoff, H. J. Prömel, Approximation algorithms for the Steiner tree problem in graphs, technical report, Humboldt-Universität zu Berlin, 2000.

[8] J. Håstad, Some optimal inapproximability results, Proceedings of the 28rd Annual Symposium on Theory of Computing, ACM, 1997.

[9] S. Hougardy, H. J. Prömel, A 1.598 approximation algorithm for the Steiner Problem in graphs, In: Proceeding of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms 1999, pp. 448–453.

[10] F.K. Hwang, D.S. Richards, P. Winter, The Steiner tree problem, Elsevier, Amsterdan (1992)

[11] M. Karpinski, A. Zelikovsky, New approximation algorithms for the Steiner tree problems, Journal of Combinatorial Optimization 1 (1997), pp.47–65.

[12] T. Lengauer, Combinatorial Algorithms for Integrated Circuit Layout, John Wiley & Sons, Chichester, England (1990)

[13] A. Lubotzy, P. Philips, P. Sarnak, Ramanujan graphs, Combinatorica, 8, (1988) pp.261–277.

[14] G.A. Margulis, Explicit construction of concentrators, Problemy Inf. Trans. 9 (1973), pp.325-332

[15] C.H. Papadimitriou, S. Vempala, On the Approximability of the Traveling Salesman Problem, Proceedings of the 32nd ACM Symposium on the theory of computing, Portland, 2000.

[16] H. J. Prömel, A. Steger, A new approximation algorithm for the Steiner tree problem with perfomance ratio 5/3, Journal of Algorithms 36 (2000), pp.89-101.

[17] G. Robins, A. Zelikovsky, Improved Steiner tree approximation in graphs, In: Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms 2000, pp.770-779.

[18] A. Zelikovsky, An 11/6-approximation algorithm for the network Steiner problem, Algorithmica 9 (1993), pp. 463–470.

[19] P. Sarnak, Some Applications of Modular Forms, Cambridge Tracts in Mathematics 99, Cambridge University Press, 1990