Maximum Leaf Spanning Tree Problem for Grid Graphs

(Ben) P. C. Li *
Department of Computer Science
University of Manitoba
Winnipeg, Manitoba
Canada R3T 2N2

M. Toulouse
CIRRELT
Université de Montréal
Montréal, Québec
Canada H3T 1J4

Abstract

The complexity of the maximum leaf spanning tree problem for grid graphs is currently unknown. We determine the maximum number of leaves in a grid graph with up to 4 rows and with 6 rows. Furthermore, we give some constructions of spanning trees of grid graphs with a large number of leaves.

1 Introduction

Given a connected graph $G = (V, E)$ and an integer $k > 0$, the *maximum leaf spanning tree problem* (MLST problem) asks if G contains a spanning tree with at least k leaf nodes. This problem has been widely studied and is known to be NP-complete [4]. This problem has applications in areas of networking (see [5]) and circuit layout (see [8]). The optimization version of the MLST problem is to find the largest number of leaves in any spanning tree of G. From this point onward, we will only be concerned with the optimization version of the MLST problem. The MLST problem was shown to be MAX SNP-complete by Galbiati et. al in [3]. This implies that there exists an $\epsilon > 0$ such that there is no $(1+\epsilon)$-approximation algorithm for this problem. In [6], Lu and Ravi provided a linear-time 3-approximation algorithm for the MLST problem. Solis-Oba gave a linear-time 2-approximation algorithm in [7].

In this paper, we study the MLST problem for grid graphs and show that for an $n \times m$ grid graph, where $n \leq 4$, the maximum number of leaves in any spanning tree is known. We also show how these "optimal" spanning trees can be constructed. Furthermore, for arbitrary grid graphs, we give some constructions of spanning trees with a large number of leaves and show that the construction for grid graphs with 6 rows is optimal. The complexity of the MLST problem

*corresponding author: lipakc@cs.umanitoba.ca

for grid graphs is currently unknown. However, the problem has been shown to be NP-complete for subgraphs of grid graphs (see [1]). Therefore, we restrict our attention to grid graphs.

It should be noted that there is a close relationship between the maximum leaf spanning tree problem and the connected dominating set problem. Given a graph G, G has a spanning tree with k leaves if and only if G has a connected dominating set with $|V| - k$ vertices.

2 Preliminary Results

Let n and m be positive integers such that $n \leq m$. A $n \times m$ *grid graph* $G(n, m) = (V, E)$ is a graph where

$$\begin{aligned} V &= \{(i, j) | 1 \leq i \leq n, 1 \leq j \leq m\}, \\ E &= \{\{(i, j), (i, j+1)\} | 1 \leq i \leq n, 1 \leq j \leq m-1\} \\ &\cup \{\{(i, j), (i+1, j)\} | 1 \leq i \leq n-1, 1 \leq j \leq m\}. \end{aligned}$$

For a grid graph $G(n, m)$, let $R(n, m)$ denote the maximum number of leaf nodes in any spanning tree of $G(n, m)$. Unless otherwise stated, we always assume that $n \leq m$. We begin with a simple but useful observation.

Lemma 2.1 *Let G be a connected graph with at least 3 vertices. If T is a spanning tree of G, and x, y are leaf nodes of T, then $\{x, y\}$ is not an edge in T.*

A column of vertices in a grid graph is called an *interior column* if it is not the first or last column of a grid graph. Otherwise, it is called a *boundary column*. The terms *interior row* and *boundary row* are similarly defined. The following result restricts the number of leaf nodes an interior column can contain in any spanning tree of a grid graph.

Lemma 2.2 *Suppose T is a spanning tree of $G(n, m)$, where $n \geq 3$. Then T cannot contain an interior column of $G(n, m)$ consisting entirely of leaf nodes.*

Proof. Suppose T contains an interior column of leaf nodes. Then by Lemma 2.1, no two of these leaf nodes are adjacent in T. Since T is a spanning tree, each node in this column of leaf nodes must be adjacent to a node of $G(n, m)$ to the left or to the right of it. There are two possible cases.

Case 1. Suppose that in T, all the leaf nodes in the column of leaf nodes are adjacent to nodes to the right of it. Then select any node to the left of the column of leaf nodes and it is clear that there is no path in T from it to any of the nodes in the column of leaves, which is a contradiction. An symmetric argument can be given if all the leaf nodes in question are only adjacent to nodes to the left of it.

Case 2. Suppose that in T, there exists two leaf nodes x, y in the column of leaf nodes such that x is adjacent to a node to the left of it, and y is adjacent to a node to the right of it. It is clear that there is no path in T from x to y, since any path must cross this column of leaf nodes. Therefore this case is not possible.

Thus, no spanning tree of $G(m, n)$ can contain an interior column entirely of leaf nodes. □

Theorem 2.3 states that if we have a spanning tree T of $G(n, m)$ where an interior column i contains $n-1$ leaf nodes of T, then we get a forest of two subtrees of T when column i is deleted from T.

Theorem 2.3 *Let T be a spanning tree of $G(n, m)$ where $n \geq 3$. Suppose there is an interior column, say column i, of $G(n, m)$ such that it contains $n-1$ leaf nodes of T. Then the induced subgraph of T restricted to columns 1 to $i-1$ and the induced subgraph of T restricted to columns $i+1$ to m are both connected components of T.*

Proof. Consider any two nodes x, y of T in columns 1 to $i-1$. In T, there exists a path from x to y. If the path goes through any vertex of column i, it must be the lone non-leaf vertex of column i. But this is impossible, since both x and y are in the first $i-1$ columns and the other $n-1$ vertices of column i are leaf nodes in T. Therefore, a path between x and y in T contains only vertices in columns 1 through $i-1$. Since x, y are arbitrarily chosen, then T restricted to columns 1 to $i-i$ is a connected component of T. A similar argument applies to T restricted to columns $i+1$ to m. □

A consequence of Theorem 2.3 is that the number of leaf nodes in column $i-1$ $(i+1)$ of the subtree T restricted to columns 1 $(i+1)$ to $i-1$ (m) is at least as great as the number of leaf nodes in column $i-1$ $(i+1)$ of the tree T. We will make use of the in determining $R(4, m)$.

It is trivial to see that $R(1, 1) = 1$, and if $m \geq 2$, then $R(1, m) = 2$. Lemmas 2.4 and 2.5 state the values of $R(2, m)$ and $R(3, m)$ respectively.

Lemma 2.4 $R(2, 2) = 2$, $R(2, 3) = 4$ *and* $R(2, m) = m$ *for* $m \geq 4$.

Proof. It is clear $R(2, 2) = 2$. For $G(2, 3)$ it is easy to see that the spanning tree $\{\{(1, 1), (1, 2)\}, \{(1, 2), (1, 3)\}\ , \{(2, 1), (2, 2)\}, \{(2, 2), (2, 3)\}, \{(1, 2), (2, 2)\}\}$ yields the maximum number of leaf nodes possible.
For $m \geq 4$, the spanning tree T given by

$$\begin{aligned} T &= \{\{(1, i), (1, i+1)\} | 2 \leq i \leq m-1\} \\ &\cup \{\{(1, i), (2, i)\} | 1 \leq i \leq m\} \end{aligned}$$

has m leaf nodes. To see this is an upper bound, suppose T is a spanning tree of $G(2, m)$ with (at least) $m+1$ leaf nodes. By Lemma 2.2, there is an

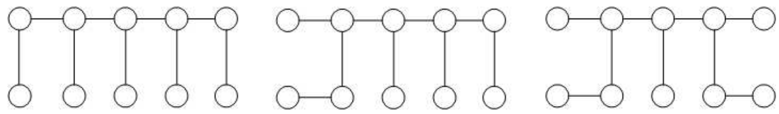

Figure 1: The three non-isomorphic maximum leaf spanning trees of $G(2,m)$

boundary column of leaf nodes in T . Assume the first column of $G(2,m)$ is a column of leaf nodes in T. Then the edges $\{(1,1),(1,2)\}$ and $\{(2,1),(2,2)\}$ are in T. In addition, the nodes $(1,2)$ and $(2,2)$ cannot be leaf nodes, as T is a spanning tree. Now consider the remaining $m-2$ columns $3,4,...,m$. If the last column is not a column of leaf nodes in T, then each of the remaining $m-2$ columns can contain at most 1 leaf node of T, in which case T contains at most m leaf nodes. Otherwise, if the last column of $G(2,m)$ is also a column of leaf nodes in T, then using a similar argument as for the first column, the nodes $(1,m-1)$ and $(2,m-1)$ are not leaf nodes in T. In this case, there are at most $2+(m-4)+2=m$ leaf nodes in T. □

It can be easily verified that, for $m \geq 4$, there are exactly three non-isomorphic spanning trees of $G(2,m)$ with m leaves. An example is given in Figure 1, with $m=5$.

Lemma 2.5 *If $m \geq 3$, then $R(3,m)=2m$.*

Proof. The following set of edges form a spanning tree T of $G(3,m)$ with $2m$ leaf nodes:

$$T=\{\{(1,i),(2,i)\},\{(2,i),(3,i)\}|i=1...m\} \cup \{\{(2,i),(2,i+1)\}|i=1..m-1\}.$$

Therefore $R(3,m) \geq 2m$. Let T be a spanning tree of $G(3,m)$. By Lemma 2.2, each interior column of $G(3,m)$ contains at most two leaf nodes of T. If each of the boundary columns of $G(3,m)$ contains at most two leaf nodes of T, then we are done. Otherwise, suppose column 1 contains three leaf nodes of T. Then column 2 contains no leaf nodes of T. Therefore, the maximum number of leaf nodes possible is $3+2(m-2)=2m-1$. □

It is easy to see from the above argument that there is exactly one non-isomorphic spanning tree of $G(3,m)$ with $2m$ leaves. An example is given in Figure 2 with $m=5$.

In [2], Fujie showed that $R(n,m) \leq \lfloor 2mn/3 \rfloor$ using a mathematical programming approach. We give an alternative proof of this result using a counting argument.

Lemma 2.6 *([2]) For the grid graph $G(n,m)$, where $n \geq 2$, $R(n,m) \leq \lfloor 2mn/3 \rfloor$.*

Proof. By Lemmas 2.4 and 2.5, the result is true for $n=2,3$. So, assume $n \geq 4$, and let T be a spanning tree of $G(n,m)$ and for $i=1,2,3,4$, let d_i denote the number of nodes of degree i in T. Then

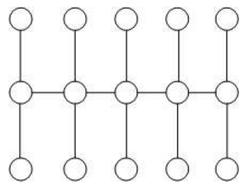

Figure 2: The unique non-isomorphic maximum leaf spanning trees of $G(3, m)$

$$d_1 + d_2 + d_3 + d_4 = mn \tag{1}$$
$$d_1 + 2d_2 + 3d_3 + 4d_4 = 2(mn - 1). \tag{2}$$

Therefore,

$$3d_1 + 2d_2 + d_3 = 2mn + 2, \tag{3}$$

or

$$d_1 = (2mn + 2 - 2d_2 - d_3)/3. \tag{4}$$

It is easy to see that for $n \geq 4$, at least one of the three nodes $\{(1,1), (1,2), (2,1)\}$, which we call a *corner* of $G(n, m)$, must be a non-leaf node of degree 2 or 3 in T. Note that none of these vertices can have degree 4. Similarly, each of the other sets of vertices $\{(1, m-1), (1, m), (2, m)\}$, $\{(m, 1), (m, 2), (m-1, 1)\}$, and $\{(m, m-1), (m, m), (m-1, m)\}$ must contain at least one non-leaf node into T. Therefore, there must be at least 4 nodes of degree 2 or 3 giving $d_1 \leq \lfloor (2mn + 2 - 4)/3 \rfloor = \lfloor (2mn - 2)/3 \rfloor \leq \lfloor 2mn/3 \rfloor$. □

In Section 4, we will use Theorem 2.6 to show that the construction of the spanning tree given for $G(6, m)$ is optimal. We now proceed to determine the value of $R(4, m)$.

3 The Main Result

In this section, we determine the value of $R(4, m)$. To start, we give a lower bound on $R(4, m)$.

Lemma 3.1 *: Suppose $m \geq 4$. Then $R(4, m) \geq 2m + \lfloor m/3 \rfloor$.*

Proof. It suffices to construct a spanning tree of $G(4, m)$ with $2m + \lfloor m/3 \rfloor$ leaf nodes. To do this, we begin with the base given by Figure 3 and will recursively build the spanning tree. Suppose a spanning tree T for a $G(4, 3k)$ has been constructed. Then, to construct a spanning tree for $G(4, 3k + 1)$, $G(4, 3k + 2)$, $G(4, 3k + 3)$, adjoin the sub-tree in Figure 4 a, b, c, respectively. Figure 5 shows

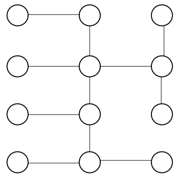

Figure 3: The Starting 4 by 3 spanning tree of $G(4,3)$

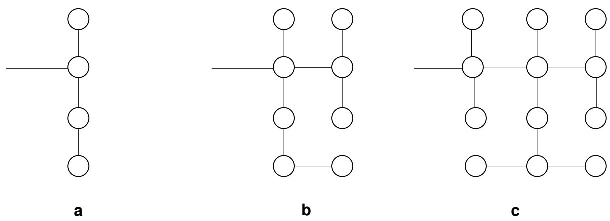

Figure 4: The subtrees to be adjoined.

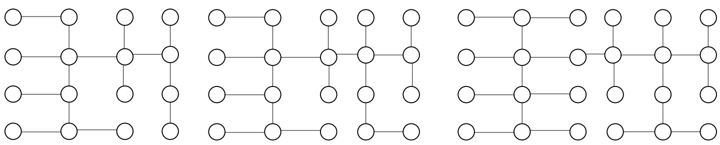

Figure 5: Spanning trees for $G(4,m)$, $m = 4,5,6$.

the spanning tree constructed for $m = 4, 5, 6$. It is easy to see that the constructed spanning tree of $G(4, m)$ has $2m + \lfloor m/3 \rfloor$ leaf nodes. □

We now show the construction outline above gives a maximum leaf spanning tree for $G(4, m)$, when $m \geq 4$. In order to do this we require the following lemmas.

Lemma 3.2 *Suppose $m \geq 4$. Then for any spanning tree T of $G(4, m)$, any 3 consecutive columns contain at most 8 leaf nodes of T and if the three columns are the rightmost (or leftmost) 3 columns of $G(4, m)$, they contain at most 7 leaf nodes of the spanning tree T.*

Proof. Let T be a spanning tree of $G(4, m)$. By Lemma 2.2, no column, except possibly a boundary column, can contain 4 leaf nodes of T.

Suppose the three columns in question are the leftmost 3 columns of the graph. If T contains 4 leaf nodes from the boundary column, then it must contain no leaf nodes from the next column. Therefore, T contains at most 7 leaf nodes from these three columns. If T contains 3 leaf nodes from the boundary column, it is easy to check that T contains $x \leq 2$ leaf nodes from the next column. If $x = 2$, then the third column can have at most 2 leaf nodes also, thereby having at most 7 leaf nodes from these three columns. If $x < 2$, then T contains at most 7 leaf nodes from these three columns, as column 3 contains at most 3 leaf nodes in T. Now, if T contains 2 leaf nodes from the boundary column, then clearly, if column 3 contains at most two leaf nodes of T, then the result is true. Therefore, assume that column 3 contains 3 leaf nodes of T. Then by Theorem 2.3, T restricted to columns 1 and 2 must be connected and hence, by Lemma 2.4, has at most 4 leaf nodes. This implies T has at most 7 leaf nodes, when the boundary column contains two leaf nodes. Finally, if the boundary column contains one leaf node, the result is clear, since interior columns may contain at most 3 leaf nodes.

Suppose the three columns are interior columns of the graph. Each column can contain at most 3 leaf nodes of T. Therefore, there are at most 9 leaf nodes in T from these 3 columns. To show that it is not possible for these three columns to have 3 leaf nodes each, suppose that it did. Applying Theorem 2.3 to the leftmost and rightmost of these 3 columns gives a spanning tree of $R(1, 4)$ with at least 3 leaves, which is impossible. □

The following lemma was confirmed using a computer program to generate all the spanning trees of $G(4, 3)$.

Lemma 3.3 *Suppose T is a spanning tree of $G(4, 3)$ such that at least one of columns 1 or 3 contains exactly two leaf nodes of T. Then T has at most 6 leaf nodes.*

We are now ready to show that $R(4, m) = 2m + \lfloor m/3 \rfloor$.

Theorem 3.4 *Suppose $m \geq 4$. Then $R(4, m) = 2m + \lfloor m/3 \rfloor$.*

Proof. It suffices to show that $R(4,m) \le 2m + \lfloor m/3 \rfloor$. We proceed to prove the result by induction on m. The result is known to be true for $m = 4$ to 9 (see [2]). Suppose $m > 9$ and let T be a spanning tree of $G(4,m)$. We begin by dealing with some special cases. If T contains 3 leaf nodes from column 3, then, by Theorem 2.3, T restricted to the first two columns form a spanning tree and contains at most 4 leaf nodes, as $R(2,4) = 4$. Also by Theorem 2.3, T minus the first three columns is also a spanning tree. By induction, the number of leaves in T is at most

$$4 + 3 + 2(m-3) + \lfloor (m-3)/3 \rfloor,$$

which is at most $2m + \lfloor m/3 \rfloor$ and therefore we have the desired result. The same argument applies if T contains 3 leaves from column $m-2$.

Now, if T contains 3 leaves from column 4, then, by the previous scenario, we can assume T contains at most 2 leaves from column 3. Since T restricted to the first three columns form a spanning tree, then by the Lemma 3.3 and Theorem 2.3, the first three columns contain at most 6 leaf nodes. By induction, the number of leaves in T is at most

$$6 + 3 + 2(m-4) + \lfloor (m-4)/3 \rfloor,$$

which is at most $2m + \lfloor m/3 \rfloor$ and therefore we have the desired result. The same argument applies if T contains 3 leaves from column $m-3$.

Therefore, from this point forward, we may assume that the spanning tree T does not contain 3 leaves from columns $3, 4, m-2, m-3$. Then it is easy to see that at most one of columns 1 and 2 contain 3 leaf nodes in T, since if column 2 contains 3 leaf nodes of T, then T restricted to column 1 must be a spanning tree of column 1, thereby having at most two leaf nodes. A similar argument holds for columns $m-1$ and m. Therefore if the only columns that contain 3 leaf nodes of T are $1, 2, m-1, m$, then T contains at most $3+2+2(m-4)+3+2 = 2m+2 \le 2m + \lfloor m/3 \rfloor$ as $m > 9$ and we are done. So we can assume that there is at least one column between columns 5 and $m-4$ which contains 3 leaves. We now consider the following 3 cases based on the value of $m \pmod 3$.

Case 1. Suppose $m \equiv 0 \pmod 3$. As T contains a column, say k, with 3 leaf nodes at some column between 5 and $m-4$. Let $m_1 = k-1$ and $m_2 = m-k$. Then, $m_1 + m_2 + 1 = m$ and by Theorem 2.3 and the induction hypothesis, the number of leaves in T is at most

$$3 + 2(m_1) + \lfloor m_1/3 \rfloor + 2(m_2) + \lfloor m_2/3 \rfloor \le 2m + 1 + \lfloor (m-1)/3 \rfloor,$$

which is at most $2m + \lfloor m/3 \rfloor$,as $m \equiv 0 \pmod 3$.

Case 2. Suppose $m \equiv 1 \pmod 3$. Denote $m = 3k+1$ for some integer $k \ge 3$. If there is a column i of $G(4,m)$ containing three leaf nodes of T, which when removed forms, by Theorem 2.3, two grid graphs with 4 rows and $m_1, m_2 \geq 4$ columns such that $m_1 + m_2 + 1 = m$ and at least one of m_1, m_2 is not divisible by 3, then a similar argument to that given in case 1 may be used to show the number of leaves in T is at most $2m + \lfloor m/3 \rfloor$, as required. Otherwise, the columns with 3 leaf nodes in T are those which m_1 and m_2 are divisible by 3 (there are $k-3$ such columns), columns 1 and 2 together contains at most 5 leaf nodes, and columns $3k$ and $3k+1$ together contains at most 5 leaf nodes. In this scenario, there are at most $3(k-3)+2(5)+2(3k+1-(k-3)-4) = 7k-1$ leaves in T which is less than $2(3k+1) + \lfloor (3k+1)/3 \rfloor = 7k+2$, as required.

Case 3. Suppose $m \equiv 2 \pmod 3$. Denote $m = 3k+2$ for some integer $k \geq 3$. If there is a column i of $G(4, m)$ containing three leaf nodes of T, which when removed forms two grid graphs with 4 rows and $m_1, m_2 \geq 4$ columns such that $m_1 + m_2 + 1 = m$ and both of m_1, m_2 is not divisible by 3, then a similar argument to that given in case 1 may be used to show the number of leaves in T is at most $2m + \lfloor m/3 \rfloor$, as required. Otherwise, each column that contains 3 leaf nodes of T will split $G(4, m)$ into two parts so that one of the two parts have $3q$ columns, for some q. These columns are exactly $4, 5, 7, 8, ..., 3k-2, 3k-1$. By assumption, column 4 does not contain 3 leaf nodes of T. Also, the first three columns contains at most 6 leaf nodes, as column 3 was assumed to contain at most 2 leaf nodes. Now the grouping of columns $4, 5, 6$ (the first grouping) can contain at most 7 leaf nodes (in which case column 5 must have three leaf nodes). Now consider the second grouping of columns $7, 8, 9$. If one of columns 7 and 8 has at most two leaf nodes, then these 3 columns have a combined total of at most 7 leaf nodes. Otherwise, if columns 7 and 8 each has 3 leaf nodes and columns $4, 5, 6$ contain 7 leaf nodes. Then, by Theorem 2.3, T restricted to columns 6 and 7 would contain a spanning tree of these two columns with 5 leaf nodes, which is impossible. Therefore the average of the first two groupings is at most 7 leaf nodes per group.

For each of the remaining $k-3$ groupings of columns $3x+1, 3x+2, 3x+3$, where $x = 3, ..., k-1$, if a grouping F contains 8 leaf nodes (which must be distributed $(3, 3, 2)$ in the columns), then note that it cannot be the last group, since we assume that column $m-3$ does not contain 3 leaf nodes of T. So we consider the next grouping G immediately following grouping F. The first column of G cannot contain 3 leaf nodes of T. For if it does, then as column 1 of grouping F has three leaf nodes of T, by Lemma 2.3, T restricted to the last two columns of the grouping F is a spanning tree of these two columns with at least 5 leaf nodes. This is not possible. Therefore, column 1 of grouping G contains at most 2 leaf nodes of T. If column 2 of grouping G contains at most 2 leaf nodes of G, then the groupings F and G contain an average of at most 7 leaf nodes each. Otherwise, if column 2 of G contains 3 leaf nodes of T and either columns 1 or 3 of G contains at most 1 leaf node of T, then again, the groupings F and G contain an average of at most 7 leaf nodes in each grouping. The only other scenario is if the columns of G contains exactly 7 leaf nodes of T where columns $1, 2, 3$ of grouping G contains $2, 3, 2$ leafs nodes of T respectively. Then, by Lemma 2.3, T restricted to columns $2, 3$ of F and column 1 of G form a

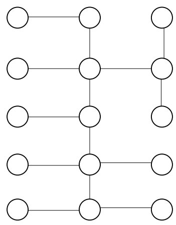

Figure 6: The Starting 5 by 3 spanning tree of $G(5,3)$

spanning tree of these columns with at least 7 leaf nodes. But it cannot contain exactly 7 leaf nodes by Lemma 3.3. Therefore, it must contain 8 leaf nodes. But this is also impossible, since there is only one non-isomorphic spanning tree on $G(4,3)$ with 8 leaf nodes, and the tree in question does not have this form (see Figure 2). So we conclude that if F has 8 leaf nodes of T, then G must have at most 6 leaf nodes of T.

Using the knowledge that the groupings contain an average of at most 7 leaf nodes of T in each grouping, we conclude that there are at most $7(k-1)$ leaf nodes of T in columns 4 to $3k$. Finally, the first two and last two columns each contain most 5 leaf nodes. Therefore the tree T contains at most $5+7(k-1)+5 = 7k+3$ leaf nodes. But note that $2m + \lfloor m/3 \rfloor = 2(k+2) + \lfloor (3k+2)/3 \rfloor = 7k+4$, and this concludes case 3 and the result is true. □

4 Other Results

Based on the construction of a maximum leaf spanning tree given in the previous section, we will extend it to give a spanning tree for $G(5,m)$ with lots of leafs.

Lemma 4.1 *For $m \geq 5$,*

$$R(5,m) \geq \begin{cases} 3m & \text{if } m \equiv 0 \pmod 3 \\ 3m-1 & \text{otherwise} \end{cases}$$

Proof. We begin with the base given by Figure 6. We will recursively build a spanning tree. Suppose a spanning tree T for a $G(5,3k)$ has been constructed. Then to construct a spanning tree for a $G(5,3k+1)$, $G(5,3k+2)$, $G(5,3k+3)$, adjoin the sub-tree in Figure 7 a, b, c, respectively. Figure 8 shows the spanning tree constructed for $m = 4,5,6$. It is easy to see that the constructed spanning tree of $G(5,m)$ has the required number of leaves. □

Lemmas 4.2, 4.3 and 4.4 provide constructions of spanning trees for $G(3q,m)$, $G(3q+1,m)$ and $G(3q+2,m)$ respectively.

Lemma 4.2 $R(3q,m) \geq q(2m-2)+2$*, for $q \geq 2$.*

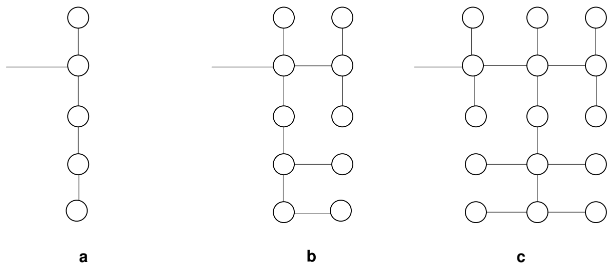

Figure 7: The subtrees to be adjoined.

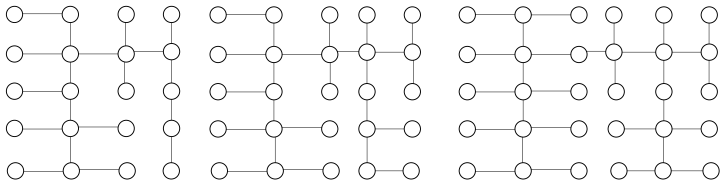

Figure 8: Spanning trees for $G(5, m)$, $m = 4, 5, 6$.

Proof. Take q copies of the spanning tree T of $G(3, m)$ given in Figure 2 and align them vertically to get $3q$ rows. Between copies i and $i + 1$ ($1 \leq i \leq q - 1$) of T arbitrarily select a column j_i and join the node in row 3, column j_i of the copy i of T to the node in row 1, column j_i of copy $i + 1$ of T. This gives a spanning tree of $G(3q, m)$ with $q(2m) - 1 - 2(q - 2) - 1 = q(2m - 2) + 2$ leaf nodes. □

Lemma 4.3 provides a lower bound for $R(3q + 1, m)$ and $R(3q + 2, m)$.

Lemma 4.3 *For* $q \geq 2$, $R(3q + 1, m) \geq (q - 1)(2m - 2) + 2m + \lfloor m/3 \rfloor + 1$

Proof. Take an optimal spanning tree S of $G(4, m)$ as given by Theorem 3.4. Take $q - 1$ copies of the spanning tree T given in Figure 2 of $G(3, m)$. Align all these trees vertically to get $3q + 1$ rows by m columns, with the tree S at the bottom. Between copies i and $i + 1$ ($1 \leq i \leq q - 2$) of T arbitrarily select a column j_i and join the node in row 3, column j_i of the copy i of T to the node in row 1, column j_i of copy $i + 1$ of T. Finally pick any vertex in the first row of S that is not a leaf and join it to the corresponding vertex in the the third row of the $(q - 1)^{st}$ copy of T. This gives a spanning tree of $G(3q + 1, m)$ with the stated number of leaf nodes. □

Lemma 4.4 *For* $q \geq 2$, $R(3q + 2, m) \geq (q - 1)(2m - 2) + 3m + 1 - i$ *where* $i = 0$ *if* $m \equiv 0 \pmod 3$, *otherwise* $i = 1$.

Proof. Use a similar construction to that of Lemma 4.3. Use a spanning tree of $G(5, m)$ given in Lemma 4.1 instead of a spanning tree of $G(4, m)$. □

For a fixed value of $n = 3q$ the construction of a spanning tree of $G(n, m)$, as stated in Lemma 4.2, is within $\lfloor \frac{2}{3} 3qm \rfloor - (q(2m - 2) + 2) = 2(q - 1)$ of an optimal spanning tree, in terms of the number of leaves.

We now show that this construction gives the maximum number of leaf nodes of any spanning tree of $G(6, m)$ and therefore $R(6, m) = 4m - 2$.

Theorem 4.5 $R(6, m) = 4m - 2$.

Proof. Suppose there exists at least $4m - 1$ leaf nodes in some spanning tree of $G(6, m)$. By Eq. 4 of Lemma 2.6, it is easy to see that there are at most 5 nodes in the spanning tree of degree 2 or 3. In addition, there are only two combinations of these $d_2 = 0, d_3 = 5$ and $d_2 = 1, d_3 = 3$. As argued in Lemma 2.6, at least 4 non-leaf nodes occur in the 4 corners (each corner consisting of three nodes). We will show that either combinations of d_2, d_3 stated leads to a contradiction.
Case 1. Suppose $d_2 = 1, d_3 = 3$. Then without loss of generality, suppose the node $b = (1, 2)$ has degree 2. It is easy to see that the nodes $a = (1, 1)$ and $c = (1, 3)$ cannot be non-leaf nodes, since the other three corners must each contain a non-leaf node of degree 3. Consider a path in the spanning tree from a to $z = (6, m)$. Each node along the path $a \to b \to \cdots \to y \to z$ must have degree 4, except for node a and node y, which must be one of the nodes of degree 3 in the spanning tree. This leads to a contradiction, since there must be three interior nodes (i, j) and $(i + 1, j)$ and $(i + 1, j + 1)$ (or (i, j) and $(i, j + 1)$ and $(i + 1, j + 1)$) in the path which have degree 4, giving a cycle in the spanning tree.
Case 2. Suppose $d_2 = 0, d_3 = 5$. Then a similar argument to the one given for Case 1 will lead to a contradiction.

Therefore, we have $R(6, m) = 4m - 2$. □

5 Conclusions

In this paper, we have determine $R(n, m)$ for all $n = 2, 3, 4, 6$. In addition, we have provided a simple construction of spanning trees for $G(n, m)$ with many leaves.

We believe that the lower bound provided by Lemma 4.1 for $R(5, m)$ is optimal, but were not able to show it.

References

[1] B. N. Clark, C. J. Colbourn, and D. S. Johnson, *Unit disk graphs*, Discrete Mathematics **86** (1990), 165–177.

[2] T. Fujie, *An exact algorithm for the maximum leaf spanning tree problem*, Computers and Operations Research **30** (2003), 1931–1944.

[3] G. Galbiati, F. Maffiol, and A. Morzenti, *A short note on the approximability of the maximum leaves spanning tree problem*, Information Processing Letters **52** (1994), 45–49.

[4] M. R. Garey and D. S. Johnson, *Computers and intractability*, Freeman, 1979.

[5] S. Guha and S. Khuller, *Approximation algorithms for connected dominating sets*, Proceedings of the Fourth Annual European Symposium on Algorithms, 1996, pp. 179–193.

[6] H. Lu and R. Ravi, *Approximating maximum leaf spanning trees in almost linear time*, Journal of Algorithms **29** (1998), 132–141.

[7] S. Solis-Oba, *2-approximation algorithm for finding a spanning tree with maximum number of leaves*, Lecture Notes in Computer Science **1461** (1998), 441–452.

[8] J. A. Storer, *Constructing full spanning trees for cubic graphs*, Information Processing Letters **13** (1981), 8–11.