High girth augmented trees are huge

Noga Alon *

Abstract

Let G be a graph consisting of a complete binary tree of depth h together with one back edge leading from each leaf to one of its ancestors, and suppose that the girth of G exceeds g. Let $h = h(g)$ be the minimum possible depth of such a graph. The existence of such graphs, for arbitrarily large g, is proved in [2], where it is shown that $h(g)$ is at most some version of the Ackermann function. Here we show that this is tight and the growth of $h(g)$ is indeed Ackermannian.

1 Introduction

An r-augmented tree is a graph consisting of a rooted tree (called the underlying tree) plus edges from each leaf to r of its ancestors (called here back edges). A complete d-ary tree of height m is a rooted tree whose internal vertices have d children and whose leaves have distance m from the root. For positive integers d, r, g let a (d, r, g)-graph be an r-augmented complete d-ary tree with girth exceeding g.

These graphs were introduced in [2], where it is proved that they exist for all integers d, r and g. As shown in [2] these graphs provide simple explicit constructions of graphs and hypergraphs of high girth and high chromatic number, and are also useful in the investigation of several natural list coloring problems. The graphs constructed in that paper are huge, their number of vertices grow like some version of the Ackermann function. It is also proved in [2] that the number of vertices must be large, specifically, the number of vertices of any $(2, 1, g)$-graph is at least a tower of height $2^{\Omega(g)}$. Although that's a fast growing function of g, the upper bound is far, far larger. Here we show that the upper bound is closer to the correct behavior, and the size of any $(2, 1, g)$ graph must indeed be Ackermannian.

For a function f, define its iterates f_i by putting $f_0(x) = x$, $f_1(x) = f(x)$ and more generally $f_i(x) = f(f_{i-1}(x))$ for all $i \geq 1$. Thus, for example, if $f(x) = 2^x$ then $f_i(1)$ is the tower of exponentials $Tower(i)$ and $Tower_i(1)$ is the iterated Tower, usually denoted by $WOW(i)$.

For positive integers x, r, g, a $TB(x, r, g)$-graph (where TB stands for Tree-Based) is a graph G whose vertices are all vertices of a complete binary tree T of some depth, called the underlying tree of

*Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel and School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540. Email: nogaa@tau.ac.il. Research supported in part by a USA-Israeli BSF grant, by an ISF grant, by the Israeli I-Core program and by the Oswald Veblen Fund.

G. Note that the graph G does not contain any of the tree edges, it is simply convenient to represent its vertices by the vertices of the underlying tree. The vertices are partitioned into levels, according to their distance from the leaves in the tree T. The leaves are at level 0, their neighbors (in T) at level 1, and so on. The edges of G must satisfy the following property. There is some integer $x' \geq x$ and a set L of leaves of the underlying tree, consisting of at least a fraction of $2^{-x'/4}$ of all leaves, so that each leaf in L has at least r back edges, all leading to levels higher than x'. Finally, the girth of G is bigger than g. Note that G is bipartite, all its edges are incident with the leaves of its underlying tree, and thus for $r = 1$ such a graph is simply a union of vertex disjoint stars.

For, say, $x \geq 100$ and integers $r \geq 2, g \geq 4$ let $m(x, r, g)$ denote the largest integer m such that any $TB(x, r, g)$-graph contains a set of at least a fraction of $2^{-m/8}$ of the leaves, so that each member of this set has a back edge leading to a level higher than m.

Note that by definition, the function $m(x, r, g)$ is monotone increasing in all its 3 variables.

It is not difficult to prove that for any $g > 101$ the minimum possible height of a $(2, 1, g)$-graph is at least $m(100, 2, g/2)$ (see subsection 3.1 for the argument). Therefore, in order to establish the Ackermannian behviour of the size of any $(2, 1, g)$-graph it suffices to prove such a lower bound for $m(100, 2, g)$. This is done in the next sections.

In Section 2 we show that $m(100, r, 4)$ grows at least like a tower function of r and that $m(100, r, 8)$ grows at least like a WOW function of r. In Section 3 we note that a similar reasoning implies that whenever g is increased by a factor of 2, the lower bound for the growth of $m(100, r, g)$ as a function of r shifts to the next level in the Ackermann hierarchy. In Section 4 we establish an upper bound for $m(x, r, g)$ (and in fact for a larger function), showing that indeed for any fixed $s \geq 1$, $m(100, r, 2^s)$ as a function of r belongs to level number s of the Ackermann hierarchy. The final Section 5 contains some concluding remarks.

2 Tower and WOW

Lemma 2.1 *Let f be the function defined by $f(y) = y + 2^{y/4}$. Then for all $r \geq 2$ and every integer $x \geq 100$*

$$m(x, r, 4) \geq f_{r-1}(x).$$

Thus the function $m(100, r, 4)$ grows at least as $Tower(r + \Theta(1))$.

Proof. Let G be a $TB(x, r, g)$-graph. By definition, there is an integer $x' \geq x$ such that there is a set L of leaves of the underlying tree, consisting of at least a fraction of $2^{-x'/4}$ of all leaves, so that each leaf in L has at least r back edges, all leading to levels higher than x'.

Split the levels above x' into groups, where each group consists of an interval of consecutive levels, as follows. Group number 1 consists of all levels in the interval $[x' + 1, f(x')]$. Group 2 consists of all levels in $[f(x') + 1, f(f(x'))] = [f(x') + 1, f_2(x)]$ and so on. Thus group number i consists of all levels in $[f_{i-1}(x') + 1, f_i(x')]$.

If there is some i so that for $z = f_{i-1}(x')$ more than a fraction of $2^{-z/2}$ of all leaves have two back edges leading to levels in group number i, then by averaging there is a subtree rooted at level $z+1$ so that more than $2^{z/2}$ of its leaves have two back edges leading to levels in $[z+1, f(z)] = [z+1, z+2^{z/4}]$. These back edges thus all lead to the unique path of $2^{z/4}$ vertices in the underlying tree leading from the root of the subtree up. Hence, by the pigeonhole principle, two distinct leaves must have their two back-edges leading to the same pair of vertices in the path, providing a cycle of length 4. This is impossible, as the girth of G is larger, hence we conclude that there are less than a fraction of $2^{-z/2}$ of all leaves with two back edges leading to group number i. Since the sum

$$\sum_{i \geq 0} \frac{1}{2^{f_i(x')/2}}$$

is much smaller than $2^{-x'/4}$, we conclude that (much) more than half the leaves in L have their r back edges leading to distinct groups, and hence the top one is in group r in which all levels are higher than $f_{r-1}(x') = t$. The fraction of the leaves with this property is close to $2^{-x'/4}$, which is much larger than $2^{-t/8}$, completing the proof of the lemma. □

The following lemma establishes the key ingredient needed for showing that when g is doubled, the lower bound for m shifts to the next level in the Ackermann hierarchy.

Lemma 2.2 *Let $x \geq 100$ be an integer. For every $TB(x, 2, 8)$-graph G there is an $x'' \geq x$ and a $z' \geq m(2^{x''/4}, 2^{x''/2}, 4)$ such that at least a fraction of $2^{-z'/8}$ $(> 2^{-z'/4})$ of the leaves have back edges to levels higher than z'. In particular, $m(x, 2, 8) \geq m(2^{x/4}, 2^{x/2}, 4)$.*

Proof. Let G be a $TB(x, 2, 8)$-graph. Thus there is an $x' \geq x$ and a fraction of at least $2^{-x'/4}$ of the leaves with 2 back edges to levels higher than x'. Since

$$\sum_{x'' > x'} \frac{1}{10 \cdot 2^{x''/4}} < \frac{1}{2^{x'/4}}$$

there is a level $x'' \geq x'$ so that a fraction of at least $0.1 \cdot 2^{-x''/4}$ of the leaves have their lower back edge (among the two) leading to level x'' (precisely). As the girth exceeds 8 (and hence exceeds 4), a fraction of at most $2^{-3x''/4}$ of the leaves among those have their higher back edge lead to level in $[x''+1, x''+2^{x''/4}]$, since otherwise, by applying the pigeonhole principle to an appropriate subtree with a root at level x'', we'll get a cycle of length 4, which is impossible. Therefore there is a set L consisting of more than, say, a fraction of

$$p = \frac{1}{11 \cdot 2^{x''/4}}$$

of all leaves so that each vertex in L has its lower back edge leading to level x'' and its higher leading to level above $x'' + 2^{x''/4}$.

Consider the following auxiliary bipartite graph H (which is simply a union of stars). Its classes of vertices are the set, call it X'', of all vertices at level x'' in the underlying tree of G, and the set of all leaves, where a leaf v is adjacent to a vertex $w \in X''$ iff v lies in L and its lower back edge is the edge vw. This graph is a union of stars, and the number of its edges is exactly $|L|$, which exceeds a fraction of p of all leaves. Therefore the average degree of a vertex in X'' is at least $d = p \cdot 2^{x''}$, and no degree can exceed $2^{x''}$. It follows that there is a subset $S \subset X''$, with $|S| \geq 0.5p|X''|$, so that the degree (in H) of each vertex of S is at least $d/2$.

Define now a new graph F as follows. Its vertices are all vertices of the underlying tree of G in levels x'' and above. Its edges are back edges from the vertices in S (which are leaves of the underlying tree of this new graph), and are defined as follows. From each $w \in S$ and each leaf $v \in L$ connected to w in H, take a back edge from w to w', where vw' is the higher back-edge of v in G. Note that w' is in level higher than $x'' + 2^{x''/4}$. In this new graph F, the number of back edges from each leaf that lies in S is at least

$$d/2 \geq \frac{2^{3x''/4}}{22} > 2^{x''/2} + 3$$

and all of them lead to levels higher than $y = 2^{x''/4}$. Moreover, the fraction of the members of S in X'' is at least

$$0.5p \geq \frac{1}{22 \cdot 2^{x''/4}}$$

which is far larger than $2^{-y/4}$. Crucially, the graph F contains no cycle of length at most 4. Indeed, each edge of F corresponds to a path of length 2 in G (consisting of the two back edges from a leaf in L), hence a cycle of length at most 4 in F would provide a cycle of length at most 8 in G, contradicting the fact there are no such cycles. Therefore, even if we ignore the top 3 back edges from each leaf of F, for $z = m(2^{x''/4}, 2^{x''/2}, 4)$ there is a fraction of at least $2^{-z/8}$ of the leaves in X'' (all being members of S) with back edges leading to levels higher than z. Most of these cannot have two edges leading to the same group, where groups of levels are defined as before, starting with level z, and since we have 3 extra back edges from each such member of S, for most of them the top back edge leads to a much higher level z', and the fraction of leaves we get with such a top back edge is much bigger than, say, $2^{-z'/16}$. This gives us a fraction of more than $2^{-z'/16}2^{-x''} > 2^{z'/8}$ leaves of the original graph G with a back edge to a level higher than z', completing the proof of the lemma. □

Lemma 2.3 *For any $r \geq 3$ and $x \geq 100$*

$$m(x, r, 8) \geq m(m(2^{x/4}, 2^{x/2}, 4), r - 1, 8).$$

Note that by iterating the above lemma, recalling that by Lemma 2.1 $m(2^{x/4}, 2^{x/2}, 4)$ is (much) larger than $Tower(x)$, it follows that in each iteration the second variable r decreases by 1 and the first, x, is replaced by something exceeding $Tower(x)$. Thus $m(100, r, 8)$ is at least a WOW function obtained by iterating the tower function $r - 2$ times.

Proof. Let G be a $TB(x, r, 8)$-graph. Then it contains a set of at least a fraction $2^{-x'/4}$ of the leaves with r back edges to levels higher than x' for some $x' \geq x$. Order the back edges from each such leaf from low to high and consider the 2 lowest back edges from each of these leaves. By Lemma 2.2 there is a $z \geq m(2^{x/4}, 2^{x/2}, 4)$ with a fraction of more than $2^{-z/4}$ of the leaves having their second lowest back edge reaching level higher than z (where we have used the monotonicity of m to replace x'' by x). Now omit the lowest back edge from each of these leaves, they still have $r-1$ back edges reaching higher than z, and this implies the assertion of the lemma. $\square$

3 Higher girth

The reasoning in the previous section applies to higher values of g as well. The following lemma can be proved by repeating the arguments of the proof of Lemma 2.2.

Lemma 3.1 *Let $s \geq 2$ be an integer. Then for any $x \geq 100$ and for any $TB(x, 2, 2^{s+1})$-graph G there is an $x'' \geq x$ and a $z' \geq m(2^{x''/4}, 2^{x''/2}, 2^s)$ such that at least a fraction of $2^{-z'/8}$ ($> 2^{-z'/4}$) of the leaves have back edges to levels higher than z'. In particular, $m(x, 2, 2^{s+1}) \geq m(2^{x/4}, 2^{x/2}, 2^s)$.*

Using this lemma, we get the following extension of Lemma 2.3.

Lemma 3.2 *For any $s \geq 2$, $r \geq 3$ and $x \geq 100$*

$$m(x, r, 2^{s+1}) \geq m(m(2^{x/4}, 2^{x/2}, 2^s), r-1, 2^s).$$

Therefore, for any fixed $s \geq 2$, $m(100, r, 2^s)$ grows, as a function of r, at least as the function in level number s of the Ackermann hierarchy, where level number 2 is the Tower function and level number 3 the WOW function.

3.1 1-augmented 2 trees

Recall that a $(2, 1, g)$-graph G, is a 1-augmented 2-tree of girth exceeding g, that is a graph consisting of a complete binary tree, with one back edge from each leaf, whose girth exceeds g. Let $h(g)$ denote the minimum possible height of the underlying tree of such a graph. In [2] it is proved that $h(g)$ is at least $Tower(2^{(1+o(1))g/4})$. Our results here supply a far higher lower bound, showing that $h(g)$ is Ackermannian and hence not primitive recursive. To see that this follows from the bounds above we prove the following simple result.

Proposition 3.3 *For any even $g > 102$, $h(g) \geq 1 + m(100, 2, g/2)$*

Proof. Let G be a $(2, 1, g)$-graph of height $h = h(g)$, and let T be its underlying tree. Using G, we construct a $TB(100, 2, g/2)$ graph H as follows. The underlying tree of H is the underlying tree T' of G without its leaves. Thus, the depth of T' is $h-1$, and its leaves are the vertices at level 1 of T. For each such vertex w whose two children in T are v, v' with their back edges in G being vu and vu' let the two back edges of w in H be wu and wu'. Clearly all back edges lead to vertices at level higher than 100, as G contains no cycle of length at most 102. In addition, any cycle C in H is of even length, as all edges are connecting leaves and non-leaves. Let $x_1y_1x_2y_2\dots x_ry_r$ be such a cycle, where x_i are leaves and y_j non-leaves (which are all at levels higher than a 100.) Let z_{i0} and z_{i1} be the two children of x_i in T, where $z_{i0}y_{i-1}$ and $z_{i1}y_i$ are their back edges (and $y_0 = y_r$). Then the cycle

$$z_{10}x_1z_{11}y_1z_{20}x_2z_{21}y_2\cdots y_{r-1}z_{r0}x_rz_{r1}y_r$$

is a cycle of length $4r$ is G. Thus $4r > g$ and hence $2r > g/2$, showing that the girth of H exceeds $g/2$. Moreover, the back edges from all leaves in H, and hence certainly more than a fraction of 2^{-25} of them, lead to levels above 100. Therefore, H is a $TB(100, 2, g/2)$-graph, and its depth $h-1$ is thus at least $m(100, 2, g/2)$, as needed. $\square$

4 Upper bounds

In this section we show that the lower bound proved in the previous sections for $m(x, r, g)$ is not too far from the correct behaviour of this function. This is done by a modification of the construction given in [2]. It is convenient to prove the upper bound for a function that grows faster than $m(x, r, g)$.

Call a graph G an $FTB(x, d, r, g)$-graph (where FTB stands for Full-Tree-Based), if its vertices are all vertices of a complete d-ary tree, its edges are all edges of the tree incident with leaves as well as r back edges from each leaf, all leading to vertices at levels higher than x, and the girth exceeds g. Note that any $FTB(x, 2, r, g)$-graph is also a $TB(x, r, g)$-graph, where we only require a fraction of the leaves to have those r back edges. In addition, an $FTB(x, 2, r, g)$-graph contains also the edges of the underlying tree incident with its leaves, unlike a $TB(x, r, g)$-graph. Let $M(x, d, r, g)$ denote the minimum possible height of the underlying tree of an $FTB(x, d, r, g)$-graph. Obviously $M(x, 2, r, g) \geq m(x, r, g)$.

To prove an upper bound for $M(x, d, r, g)$ we describe a construction of $FTB(x, d, r, g)$-graphs, by induction. This is done, following the method in [2], by proving the following three lemmas.

Lemma 4.1 *For all positive integers x, d, r,* $M(x, d, r, 2) = x + r$

Lemma 4.2 *For all positive integers x, d and g,* $M(x, d, 1, 2g) \leq 1 + M(x-1, d, d, g)$

Lemma 4.3 *For all positive integers x, d, r, g,*

$$M(x, d, r+1, g) \leq M(x, d, 1, g) + M(1, d^{M(x,d,1,g)}, r, g) - 1.$$

Lemma 4.1 is trivial: the back edges from each leaf are connected to its ancestors at levels $x+1, x+2, \dots, x+r$.

Proof of Lemma 4.2.

Let G' with underlying tree T' be an $FTB(x-1, d, d, g)$-graph with height $M(x-1, d, d, g)$. Replace each leaf v of T' with a star S_v of d leaves rooted at v. Replace the back edges from v to its ancestors by letting the d lower endpoints be the leaves of the star S_v instead of v. This produces one back edge from each leaf, and all back edges go to levels higher than x. Let G be the graph consisting of all these back edges together with the edges of the stars S_v. In order to show that this is an $FTB(x,d,1,2g)$-graph we have to prove that it contains no cycle of length at most $2g$.

Any cycle C in G must contain a back edge, and the two other edges of the cycle incident with this back edge must be another back edge and a star edge. Similarly, the two edges incident with any star edge along the cycle must be a star edge and a back edge. Thus any cycle is of length divisible by 4, consisting of two back edges followed by two star edges, followed by two back edges, etc. Contracting all the star edges of the cycle produces a cycle of length $|C|/2$ which is a cycle in G' (consisting only of back edges). Thus $|C|/2 > g$, showing that the girth of G exceeds $2g$, as needed. □

Proof of Lemma 4.3.

Fix r. Assuming that for all x, d and g $M(x,d,r,g)$ and $M(x,d,1,g)$ are finite, put $M_1 = M(x,d,1,g)$ and $M_2 = M(1,d^{M_1},r,g)$. We construct the desired graph G from two graphs G_1 and G_2.

For G_1 we use an $FTB(x,d,1,g)$-graph having height M_1. For G_2, define $d' = d^{M_1}$, and consider an $FTB(1,d',r,g)$-graph H having height M_2. Let G_2 be an induced subgraph of H formed by starting from the root of the underlying tree of H and keeping only d children of each included vertex, except that all d' children are kept at the last level. Thus G_2 has an underlying tree T' of height M_2, and deleting the $d^{M_2-1}d'$ leaves of T' yields a complete d-ary tree of height $M_2 - 1$. All ancestors in H of a leaf of T' appear in T', so each leaf of T' has r ancestors as neighbors in G_2.

Now we construct G from G_1 and G_2. In G_2, let $S(u)$ be the star consisting of a vertex u at level 1 and its d' leaf children. (Recall that levels are counted from the leaves, hence level number 1 is the one containing their neighbors in the underlying tree). Replace each $S(u)$ with a copy $G_1(u)$ of the graph G_1, so that the d' leaves in G_1 each become one of the leaves in $S(u)$, inheriting the r back edges that were incident to that leaf in G_2. We call the back edges obtained from G_2 in this way long edges; the back edges in $G_1(u)$ are short edges.

The underlying tree in our construction thus has two parts. The top part is the tree T' for G_2 without its bottom level; it has height $M_2 - 1$. The bottom part, with height M_1, consists of copies of G_1. Each leaf has one incident short edge from G_1 and r incident long edges inherited from G_2. Clearly all back edges lead to levels higher than x.

A cycle C in G that contains no long edges is a cycle in a copy of G_1 and hence has length at least g. When C contains at least one long edge, it is not difficult to see that the graph G_2 contains a cycle C' containing the same long edges, whose length is at most that of C. Indeed, C' can be obtained from C by replacing each interval of consecutive edges of C that belong to some subtree $G_1(u)$ and leads from vertex v to vertex v' (which are necessarily both leaves) by the two edges vu and uv' of G_2. Since any such interval must contain at least two edges, as no two leaves are connected, this produces a cycle C' of G_2 whose length is at most C. Thus $|C| \geq |C'| \geq g$, completing the proof. □

The above three lemmas imply that for any $s \geq 1$, $M(100,2,r,2^s)$ is bounded by a function in level number s of the Ackermann hierarchy, where level number 2 is that of the Tower function and level number 3 is that of the WOW function. We omit the simple though somewhat tedious computation. Note that in view of the lower bound for the numbers $m(100, r, 2^s)$ this determines the correct location of both functions in the hierarchy (although, of course, there is still a huge gap between the upper and lower bounds).

5 Concluding remarks

We have shown that the function $h(g)$ which is the minimum possible height of a binary tree with one back edge from each leaf and girth exceeding g grows like a version of the Ackermann function. Therefore, this is a natural combinatorially defined function which is not primitive recursive. There are several known examples of functions corresponding to natural combinatorial problems or natural questions in complexity theory which are closely related to the Ackermann or the inverse Ackermann function. Some of these can be found in [9], [10], [6], [4], [3], [5], [7], [8]. The function $h(g)$ is possibly the one corresponding to the combinatorial problem with the simplest description among all of the above.

In view of of its rapid growth, it may be interesting to try and compute or estimate $h(g)$ more accurately for small values of g. It is not difficult to check that $h(1) = 1, h(2) = 2, h(3) = 3, h(4) = 5, h(5) = 7, h(6) = 10$ and $h(7) = 13$. Without trying to optimize the computation very carefully, we can show that $h(8) > 270$. Our proofs, and to some extent also the bounds above, suggest that $h(g) - h(g-1)$ is particularly large for values of g that are powers of 2. Indeed, following the reasoning in our proofs here (again, without optimizing the estimates very carefully), we can show that

$$h(16) > Tower(2^{2^{2^{18}}}).$$

We believe that the problem of determining the precise value of $h(16)$, and certainly that of determining precisely $h(g)$ for higher values of g, is beyond reach.

References

[1] N. Alon, H. Kaplan, G. Nivasch, M. Sharir and S. Smorodinsky, Weak ϵ-nets and interval chains, Proc. of the Nineteenth Ann. ACM-SIAM SODA (2008), 1194-1203. Also: J. ACM 55 (2008), no. 6, Art. 28, 32 pp.

[2] N. Alon, A. Kostochka, B. Reiniger, D. B. West and X. Zhu, Coloring, sparseness, and girth, Israel J. Math., to appear.

[3] N. Alon and B. Schieber, Optimal preprocessing for answering on-line product queries, Tech. Report 71/87, The Moise and Frida Eskenasy Institute of Computer Science, Tel Aviv University, 1987.

[4] A. K. Chandra, S. Fortune, and R. Lipton, Unbounded fan-in circuits and associative functions, J. Comput. Syst. Sci., 30:222–234, 1985.

[5] B. Chazelle and B. Rosenberg, The complexity of computing partial sums off-line, Int. J. Comput. Geom. Appl., 1:33–45, 1991.

[6] D. Dolev, C. Dwork, N. Pippenger, and A. Wigderson, Superconcentrators, generalizers and generalized connectors with limited depth, in Proc. 15th Ann. ACM Sympos. on Theory of Comput., pp. 42–51, 1983.

[7] P. Pudlak, Communication in bounded depth circuits, Combinatorica, 14:203–216, 1994.

[8] M. Sharir and P. K. Agarwal, Davenport-Schinzel Sequences and Their Geometric Applications, Cambridge University Press, Cambridge, 1995.

[9] R. E. Tarjan, Efficiency of a good but not linear set union algorithm, J. ACM 22 (1975), 215–225. 68A10

[10] A. C. Yao, Space-time tradeoff for answering range queries, in Proc. 14th Ann. ACM Sympos. on Theory of Comput., pp. 128–136, 1982.