Compacted binary trees admit stretched exponentials

Andrew Elvey Price* Wenjie Fang† Michael Wallner‡

Compacted binary trees are a special class of directed acyclic graphs (DAGs) that are used as models for data structures. They appear in computer algebra systems such as Maple and in the compression of XML documents resulting in up to 90% gain in memory [1]. Given a rooted binary tree of size n (i.e., n internal nodes), its compacted form can be computed in expected and worst-case time $\mathcal{O}(n)$ with expected compacted size $\Theta(n/\sqrt{\log n})$ [5]; see Figure 1. Recently, Genitrini, Gittenberger, Kauers, and Wallner solved the reversed question on the asymptotic number of compacted trees of size n under certain height restrictions [7]; however the asymptotic number in the unrestricted case, whose counting sequence is given by OEIS A254789[1], remained out of reach. We solved this problem in [4] and showed that this difficult asymptotic enumeration problem admits a *stretched exponential* μ^{n^σ}, where $\mu > 0$ and $\sigma \in (0, 1)$.

Theorem 1 ([4]). *The number c_n of unrestricted compacted binary trees of size n satisfies*

$$c_n = \Theta\left(n!\, 4^n e^{3a_1 n^{1/3}} n^{3/4}\right)$$

for $n \to \infty$, where $a_1 \approx -2.338$ is the largest root of the Airy function $\mathrm{Ai}(x)$ *defined as the unique function satisfying* $\mathrm{Ai}''(x) = x\mathrm{Ai}(x)$ *and* $\lim_{n\to\infty} \mathrm{Ai}(x) = 0$.

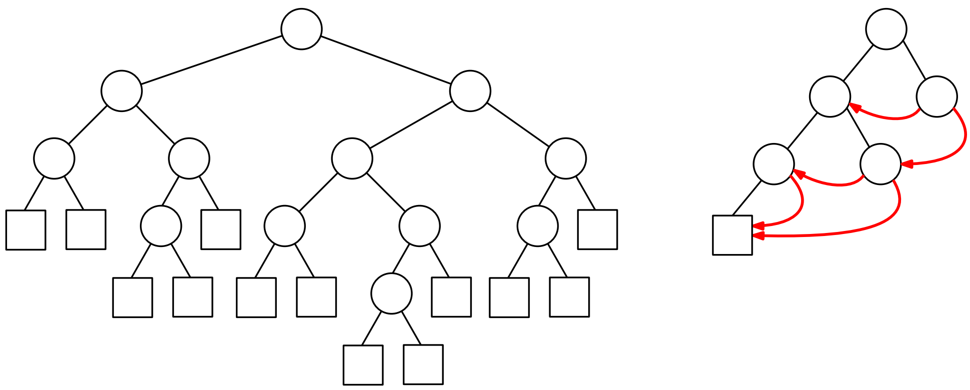

Figure 1: A binary tree with 12 internal nodes is transformed into a compacted tree with 5 internal nodes. Each unique subtree corresponds to exactly one node in the compacted tree.

The presence of a stretched exponential in the asymptotics of counting sequences is not common, although recently more and more instances emerge. For example, they appear in integer partitions in number theory, in pushed Dyck paths (a Dyck path of maximum height H is given a weight y^{-h} for some $y > 1$) in combinatorics/statistical mechanics, or in cogrowth sequences in group theory; see [4] and references therein.

*Laboratoire Bordelais de Recherche en Informatique, Université de Bordeaux, France.
†Laboratoire d'Informatique Gaspard-Monge, Université Gustave-Eiffel, France.
‡Institute of Discrete Mathematics and Geometry, TU Wien, Austria.
[1] On-line Encyclopedia of Integer Sequences: https://oeis.org

Furthermore, next to compacted binary trees also other DAG-classes admit stretched exponentials. With the same methods we were able to show that relaxed binary trees (which are compacted binary trees in which subtrees do not need to be unique; see OEIS A082161) satisfy a result similar to Theorem 1 and admit the same stretched exponential [4]. In addition, these objects are closely related to deterministic finite automata (DFAs) recognizing a finite binary language as shown in Figure 2 and these also admit a stretched exponential [3]; see OEIS A331120. Note that until now, the problem of counting such automata, even asymptotically, was widely open; see for example [2].

Theorem 2 ([3]). *The number $m_{2,n}$ of non-isomorphic minimal DFAs on a binary alphabet recognizing a finite language with $n+1$ states satisfies*

$$m_{2,n} = \Theta\left(n!\, 8^n e^{3a_1 n^{1/3}} n^{7/8}\right)$$

for $n \to \infty$, where $a_1 \approx -2.338$ is the largest root of the Airy function.

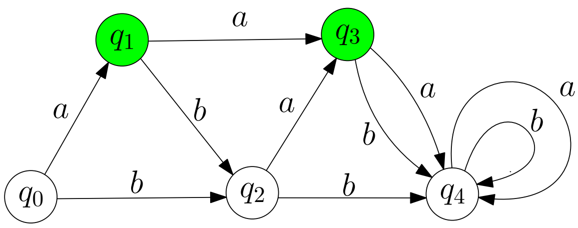

Figure 2: The unique minimal DFA with 5 states for the language $\{a, aa, ba, aba\}$. Here, q_0 is the initial state, q_1 and q_3 are the final states, and q_4 is the unique dead state.

To prove these results, we have developed the following method: First, we map the problem bijectively to a path enumeration problem which is described by a two-parameter recurrence. Second, we use empirical methods to estimate the values of all terms defined by the recurrence. Third, we prove asymptotically tight bounds by induction derived from the previous estimates.

Our method is very versatile and has already found applications in phylogenomics solving the counting problem of tree-child networks, which thus also admit stretched exponentials [6]. In the future, our results allow the analysis of shape parameters such as the average height or depth. Furthermore, we will apply our method to other counting problems arising, e.g., in lambda terms or queuing theory, and possibly prove the presence of more stretched exponentials.

References

[1] Mireille Bousquet-Mélou, Markus Lohrey, Sebastian Maneth, and Eric Noeth. XML compression via directed acyclic graphs. *Theory of Computing Systems*, 57(4):1322–1371, 2015. [doi].

[2] Michael Domaratzki. Enumeration of formal languages. *Bulletin of the EATCS*, 89:117–133, 2006.

[3] Andrew Elvey Price, Wenjie Fang, and Michael Wallner. Asymptotics of Minimal Deterministic Finite Automata Recognizing a Finite Binary Language. In *AofA 2020*, volume 159 of *LIPIcs*, pages 11:1–11:13. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020. [doi].

[4] Andrew Elvey Price, Wenjie Fang, and Michael Wallner. Compacted binary trees admit a stretched exponential. *J. Combin. Theory Ser. A*, 177:105306, 2021. [doi].

[5] Philippe Flajolet, Paolo Sipala, and Jean-Marc Steyaert. Analytic variations on the common subexpression problem. In *Automata, languages and programming*, pages 220–234. Springer, 1990. [doi].

[6] Michael Fuchs, Guan-Ru Yu, and Louxin Zhang. On the asymptotic growth of the number of tree-child networks. *arXiv preprint arXiv:2003.08049*, 2020.

[7] Antoine Genitrini, Bernhard Gittenberger, Manuel Kauers, and Michael Wallner. Asymptotic enumeration of compacted binary trees of bounded right height. *J. Combin. Theory Ser. A*, 172:105177, 2020. [doi].