One Drop of Non-Determinism in a Random Deterministic Automaton

Abstract

Every language recognized by a non-deterministic finite automaton can be recognized by a deterministic automaton, at the cost of a potential increase of the number of states, which in the worst case can go from n states to 2^n states. In this article, we investigate this classical result in a probabilistic setting where we take a deterministic automaton with n states uniformly at random and add just one random transition. These automata are almost deterministic in the sense that only one state has a non-deterministic choice when reading an input letter. In our model each state has a fixed probability to be final. We prove that for any $d \geq 1$, with non-negligible probability the minimal (deterministic) automaton of the language recognized by such an automaton has more than n^d states; as a byproduct, the expected size of its minimal automaton grows faster than any polynomial. Our result also holds when each state is final with some probability that depends on n, as long as it is not too close to 0 and 1, at distance at least $\Omega(\frac{1}{\sqrt{n}})$ to be precise, therefore allowing models with a sublinear number of final states in expectation.

2012 ACM Subject Classification Theory of computation → Formal languages and automata theory; Mathematics of computing → Discrete mathematics; Mathematics of computing → Probability and statistics

Keywords and phrases non-deterministic automaton, powerset construction, probabilistic analysis

Digital Object Identifier 10.4230/LIPIcs.STACS.2023.23

42nd Conference on Very Important Topics (CVIT 2016).
Editors: John Q. Open and Joan R. Access; Article No. 23; pp. 23:1–23:35
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

A fundamental result in automata theory is that deterministic and complete finite state automata recognize the same languages as non-deterministic finite state automata. This result can be established using the classical (accessible) subset construction [12]: starting with a non-deterministic automaton with n states, one can build a deterministic automaton with at most 2^n states that recognizes the same language. This upper bound is tight; there are regular languages recognized by an n-state non-deterministic automaton whose minimal automaton (the smallest deterministic and complete automaton that recognizes the language) has 2^n states. The number of states of the minimal automaton of a regular language is called its *state complexity*. Figure 1 shows two n-state non-deterministic automata with somewhat similar shape, and whose languages have very different state complexities. Both automata can be made deterministic by just removing the a-loop on the initial state.

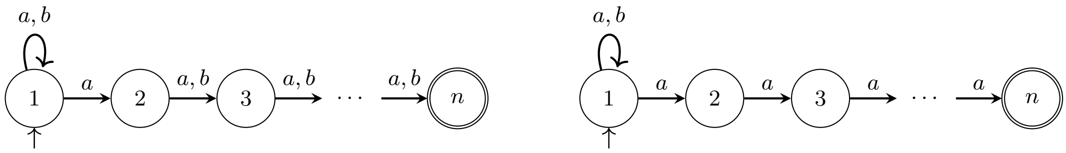

Figure 1 On the left, a non-deterministic automaton with n states recognizing the language $\mathcal{L}_\ell = \Sigma^* a \Sigma^{n-2}$. On the right, a non-deterministic automaton with n states recognizing the language $\mathcal{L}_r = \Sigma^* a^{n-1}$. The minimal automaton of $\mathcal{L}_\ell$ has 2^{n-1} states, whereas the one of $\mathcal{L}_r$ has n states.

In this article, we address the following (informal) question: if we take a random n-state deterministic automaton and add just one random transition, what can be said about the state complexity of the resulting recognized language? Does it hugely increase as for $\mathcal{L}_\ell$, or does it remain small as for $\mathcal{L}_r$?

From [3], we know that with high probability, the state complexity of the language recognized by a size-n deterministic automaton taken uniformly at random is linear. It is important as it implies that the corresponding distribution on regular languages is not degenerated: this contrasts with the case of random regular expressions where the expected state complexity of the described regular languages is constant [14] which means that the induced distribution on regular languages is concentrated on a finite number of languages.

To be more precise, our formal setting in this article is the following. Let $\Sigma = \{a, b, \ldots\}$ be a finite alphabet with $k \geq 2$ letters. For any $n \geq 1$, we consider the uniform distribution on deterministic and complete automata on Σ, with stateset $\{1, \ldots, n\}$ and with no final states (for now); the initial state is picked uniformly at random, and the action of the letters on the stateset are k uniform and independent random mappings. We also pick uniformly at random two independent states p and q, and add a transition $p \xrightarrow{a} q$, if it is not already there. Finally each state is final with a given fixed probability $f \in (0, 1)$, independently. Hence in this model an almost deterministic automaton has an expected number final states of fn. Our results still hold if we allow the probability f of being final to depend on the size n of the automaton provided that f_n has a distance to 0 and 1 in $\Omega(\frac{1}{\sqrt{n}})$. This allows us to consider a probabilistic model in which random automata have an expected number of final states as low as $\Theta(\sqrt{n})$.

Our main result is that for any $d \geq 1$ there exists a constant $c_d > 0$ such that the state complexity of the language of such a random almost deterministic automaton is greater than n^d with probability at least c_d, for n sufficiently large. That is, for any polynomial P, there is a non-negligible probability that the state complexity of the language of a random automaton is greater than $P(n)$: we will say that the state complexity is *super-polynomial* with *visible probability*. As a direct consequence, the expected state complexity is super-polynomial.

It should be noted that with the same random models for deterministic automata, one cannot hope to replace visible probability in our results with a probability that converges to 1 (high probability). Indeed random automata have, with high probability, a constant fraction of states that are not accessible from the initial state; if the source of the added transition is not accessible from the initial state, the added transition does not impact the recognized language, whose state complexity is therefore at most equal to n. Thus, we make no effort in the present paper to optimize our probabilistic lower bounds. See the conclusion for a more advanced discussion on this topic.

Related work. The study of random deterministic automata can be traced back to the work of Grusho on the size of the accessible part [11]: he established that, with high probability, a constant proportion of the states are accessible from the initial state. He also shows that with high probability there is a unique terminal strongly connected component of size approximately $\nu_k n$, for some $\nu_k > \frac{1}{2}$ that only depends on the size k of the alphabet. More structural results on the underlying graph of a random deterministic automaton were established in the work of Carayol and Nicaud [6], with a local limit law for the size of the accessible part and an application to random generation of accessible determistic automata, and more recently in the work of Cai and Devroye [5], with, in particular, a fine grain analysis of what is happening outside the large strongly connected component. In [1], Addario-Berry, Balle and Perarnau gave a precise analysis of the diameter of a random deterministic automaton, showing in particular that it is logarithmic. We will use some of these results in this paper, namely one on the size of the largest terminal strongly connected component. We will deal with the restriction to states accessible from the initial state in the powerset construction using the result of [5] that with high probability the cycles outside the accessible part are small: for any $\varepsilon > 0$, with probability at least $1 - \varepsilon$ all the non-accessible cycles have length smaller than some constant C_ε. In particular, for any $\omega(n) \to \infty$, all the cycles outside the accessible part have length at most $\omega(n)$ with high probability.

All these results on random automata focus on the underlying graph of the transition structures, without saying much about the recognized languages, and on the average complexity of textbook algorithms on automata, as we do in this article.

There are results in this line of work, and we should first mention the work of De Felice and Nicaud [9, 10], who studied the complexity of applying Brzozowski's algorithm to a random deterministic automaton. The first step of this algorithm consists in applying the powerset construction to the mirror of the automaton, obtained by reversing every transition and exchanging the role of initial and final states. Hence, as in the present article, they studied the determinization procedure of random automata, but for a model very different from ours: we add one random transition to a uniform random deterministic automaton where they consider the mirror of a uniform random deterministic automaton. However, we will still use some of their technical lemmas concerning cycles in the last part of our proof.

There are other works on random deterministic automata and their languages, which are less directly related to this article. For instance, the probability that a random accessible automaton is minimal was studied by Bassino, David and Sportiello [3], the analysis of minimization algorithms by Bassino, David and Nicaud [2, 8], etc. More recently, several papers studied the synchronization of random automata [4, 17], until the very recent work of Chapuy and Perarnau [7], establishing that most deterministic automata are synchronizing, with a word of length $O(\sqrt{n}\log n)$. We refer the interested reader to the survey of Nicaud [16] for an overview on random deterministic automata.

2 Definitions and notations

For any $n \geq 1$, let $[n] = \{1, \ldots, n\}$. If $x, y \in \mathbb{R}$ with $x \leq y$, let $[\![x, y]\!] = [x, y] \cap \mathbb{Z}$ be the set of integers that are between x and y. Let $\mathcal{E}$ be a set equipped with a size function s from $\mathcal{E}$ to $\mathbb{Z}_{\geq 0}$, and let $\mathcal{E}_n$ denote the elements of $\mathcal{E}$ having size n. A property X on $\mathcal{E}$ (that is, a subset of $\mathcal{E}$ viewed as the set of elements for which the property holds) holds with *visible probability* if there exists some constant $c > 0$ such that, for n sufficiently large, $\mathcal{E}_n$ is non-empty and $\mathbb{P}(X) \geq c$ for the uniform distribution on $\mathcal{E}_n$. By a slight abuse of notation, if X is a random variable $\mathcal{E} \to \mathbb{Z}_{\geq 0}$ we say that for the uniform distribution on $\mathcal{E}$, X is *super-polynomial with visible probability* when for any $d \geq 1$, there exists a constant $c_d > 0$, such that for n sufficiently large, $\mathcal{E}_n \neq \emptyset$ and $\mathbb{P}(X \geq n^d) \geq c_d$ for the uniform distribution on $\mathcal{E}_n$.

Recall that if u and v are two words on an ordered alphabet Σ, u is *smaller than* v *for the length-lexicographic order* if $|u| < |v|$ or they have same length and $u <_{\text{lex}} v$ for the lexicographic order.

Throughout the article, the stateset of an automaton with n states will always be $[n]$, with the exception of the powerset construction recalled just below. The alphabet will always be $\Sigma = \{a, b\}$, except in the statement of our main theorem, where we allow larger alphabets as it is trivially generalized to this case. Hence, in our setting, a *deterministic (and complete) automaton* is just a tuple (n, δ, F), where $F \subseteq [n]$ is the *set of final states* and δ is the *transition function*, a mapping from $[n] \times \Sigma$ to $[n]$. We will often write $\delta_\alpha(s) = t$ or $s \xrightarrow{\alpha} t$ instead of $\delta(s, \alpha) = t$, for $s, t \in [n]$ and $\alpha \in \Sigma$, and call this an α-transition or a transition. The transition function is classically extended to sets of states by setting $\delta(X, \alpha) = \{\delta(s, \alpha) : s \in X\}$, for $X \subseteq [n]$, and to words by setting inductively $\delta(s, w) = s$ if w is the empty word ε and $\delta(s, w\alpha) = \delta(\delta(s, w), \alpha)$. We will not need to specify the *initial state* until the end of the proof; when we finally do, it will be generated uniformly at random and independently in $[n]$. Final states are only used in the last part of our proof, so to ease the presentation, we define a *deterministic (and complete) transition structure* as being an automaton with neither initial nor final states: they are given by a pair (n, δ) where n is the number of states and δ is the transition function.

An *almost deterministic automaton* $(n, \delta, F, p \xrightarrow{a} q)$ is a deterministic automaton (n, δ, F) in which we add the additional a-transition $p \xrightarrow{a} q$. Similarly, an *almost deterministic transition structure* $(n, \delta, p \xrightarrow{a} q)$ is a deterministic transition structure (n, δ) in which we add the additional a-transition $p \xrightarrow{a} q$. For any $\alpha \in \Sigma$ and any $r \in [n]$, the transition function γ of an almost deterministic automaton $(n, \delta, F, s \xrightarrow{a} t)$ (or almost deterministic transition structure) is therefore defined by $\gamma(r, \alpha) = \{\delta(r, \alpha)\}$ if $(r, \alpha) \neq (p, a)$ and $\gamma(p, a) = \{\delta(p, a), q\}$. These automata or transition structures can be deterministic, when we already have $\delta(p, a) = q$.

The classical *powerset automaton* $\mathcal{B}$ of a possibly non-deterministic automaton $\mathcal{A} = (n, \delta, F, p \xrightarrow{a} q)$, with a transition function γ, is a deterministic automaton $\mathcal{B}$ with states in $2^{[n]}$ and transition function γ extended to sets, as defined above. If we add an initial state i_0 to $\mathcal{A}$, the initial state of $\mathcal{B}$ is $\{i_0\}$ and it recognizes the same language as $\mathcal{A}$ when a state X of $\mathcal{B}$ is final if and only at least one of its element is final in $\mathcal{A}$, i.e. $X \cap F \neq \emptyset$. We can restrict this construction to the accessible part of $\mathcal{B}$ only (from its initial state $\{i_0\}$, where i_0 is the initial state of $\mathcal{A}$) while still recognizing the same language; we call this automaton the *accessible powerset automaton* of $\mathcal{A}$.

Recall that two states r and s in a deterministic automaton $\mathcal{A}$ are *equivalent* if the languages recognized by moving the initial state to r or to s are equal. The *minimal automaton* of a regular language $\mathcal{L}$ is the deterministic and complete automaton with the

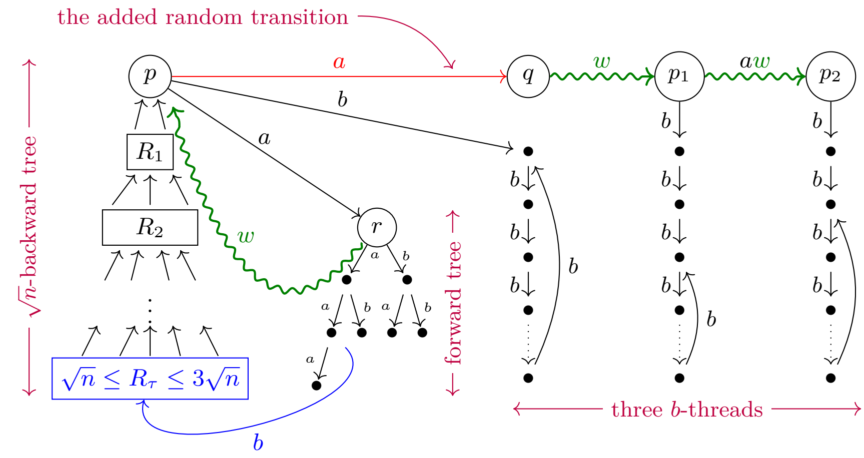

Figure 2 Illustration of the proof sketch. On the left, the backward tree from p that is detailed in Section 4.1, it has size $O(\sqrt{n})$ and contains between $\sqrt{n}$ and $3\sqrt{n}$ extremal leaves (i.e. leaves in its last level τ) to be valid. On its right, the forward tree from r, described in Section 4.2; it is a breadth-first traversal that is valid if it hits an extremal leaf of the backward tree before $O(\sqrt{n})$ states are examined. On the right the b-threads introduced in Section 4.3, obtained by reading b's from the p_i's; they are valid if they are made of previously unseen states and do not intersect.

smallest number of states that recognizes $\mathcal{L}$. The number of states of the minimal automaton of $\mathcal{L}$ is called the *state complexity* of $\mathcal{L}$. We will use the following classical property [12]:

▶ **Proposition 1.** *If there is a set of accessible states X in a deterministic automaton $\mathcal{A}$ such that the states of X are pairwise non-equivalent, then $\mathcal{A}$ has state complexity at least $|X|$.*

The following remark allows us to focus on the case of a two-letter alphabet:

▶ Remark 2. Let $\Gamma \subseteq \Sigma$ be two non-empty alphabets. If $\mathcal{L}$ is a regular language on Σ, the state complexity of $\mathcal{L}$ is at least the state complexity of $\mathcal{L} \cap \Gamma^*$.

3 Main statement and proof outline

Our main result is that the state complexity of the language recognized by a random almost deterministic automaton is super-polynomial with visible probability, when each state is final with probability f_n that is not too close to either 0 or 1:

▶ **Theorem 3.** *Let Σ be an alphabet with at least two letters. Let f_n be a map from $\mathbb{Z}_{\geq 1}$ to $(0, 1)$ such that there exists a constant $\alpha > 0$ such that $f_n \geq \frac{\alpha}{\sqrt{n}}$ and $1 - f_n \geq \frac{\alpha}{\sqrt{n}}$ for n sufficiently large. Consider an almost deterministic n-state transition structure $\mathcal{A}$ on Σ taken uniformly at random. Each state of $\mathcal{A}$ is then taken to be final with probability f_n, independently of everything else. Then with visible probability, the language recognized by $\mathcal{A}$ has super-polynomial state complexity.*

▶ **Corollary 4.** *Under the conditions of Theorem 3, the expected state complexity of the language recognized by $\mathcal{A}$ growths faster than any polynomial in n.*

The proof of Theorem 3 consists in identifying a structure and several constraints (see Figure 2) that guarantee that when performing the accessible powerset construction and adding a random set of final states, we have sufficiently many pairwise non-equivalent states.

At each step, we add a new constraint on top of those we already have, and we have to ensure that these constraints are still satisfied by sufficiently many almost deterministic transition structures. A convenient way to sketch the proof is to consider that we start with n states and no transitions, and add random transitions when needed, on the fly. More precisely, our proofs can be seen as the description of an algorithm that tries to expose the required structure by performing two types of queries on the set of still unknown transitions: either we ask what the destination of a given transition is, or we ask for all the transitions that have a given state as their destination. Thus, at any point in the algorithm, conditioned on the results of all previous queries, the destinations of all still unexposed transitions are independent and uniform among the set of states for which we have not performed the second type of query. We use this to prove that our algorithm has a non-negligible probability of success. We also have two random states p and q and will add the transition $p \xrightarrow{a} q$ at some point. We fix $d \geq 1$, the main steps of the proof are the following:

1. Generate $r = \delta_a(p)$, the target of the a-transition starting from p in the deterministic transition structure. With visible probability, $r \neq q$ and there is a word w of length $\Theta(\log n)$ such that $\delta_w(r) = p$, which can be found by generating $O(\sqrt{n})$ random transitions. We also assume that the b-transition starting at p is still unset. This step is the most technical, we explore backward from p and forward from r until we reach a common state.
2. Assuming such a w is found, we add the transition $p \xrightarrow{a} q$, which makes the automaton non-deterministic. We then iteratively generate the transitions starting from q and following the word $w(aw)^{d-1}$, and ask that the target of each such transition be a state that was not previously seen in the whole process. This happens with visible probability.
3. Let $p_0 = p$ and $p_i = \delta_{w(aw)^{i-1}}(q)$ for $i \in [d]$. If the two previous steps are successful, then $\delta_{(aw)^d}(\{p\}) = \{p_0, p_1, \ldots, p_d\}$, and the outgoing b-transition of each p_i is still unset. Then, for each p_i, we iteratively generate the b-transitions $\delta_b(p_i)$, $\delta_{bb}(p_i)$, $\ldots$ until we cycle after λ_i steps. This process is considered successful if we do not use an already set b-transition and if the $d+1$ cycles are pairwise disjoint. We furthermore ask that the λ_i are all in $\Theta(\sqrt{n})$. All these properties happen with visible probability.
4. At this stage, we have $\gamma_{(aw)^d}(\{p\}) = \{p_0, \ldots, p_d\}$; this set is composed of $d+1$ different states, and reading b's from each p_i eventually ends in a b-cycle of length ℓ_i. Given the λ_i's, each ℓ_i is a uniform element of $[\lambda_i]$, and they are independent. We now ask that the ℓ_i's are pairwise coprime, and that each of them is in $\Omega(\sqrt{n})$. This also happens with visible probability [18].
5. If everything worked so far, in the powerset construction applied to the almost deterministic transition structure there is a b-cycle of length $\prod_{i=0}^{d} \ell_i = \Omega(n^{\frac{d+1}{2}})$. We now randomly determine which states are final. If we consider a b-cycle alone in the automaton, of length $\Omega(\sqrt{n})$, its states are pairwise non-equivalent with visible probability as soon as the probability f_n that a state is final is not too close to either 0 or 1, which we assumed in our model. This property happens to be preserved when building the product automaton for the union of two one-letter cycles, provided their lengths are coprime. Consequently, the large b-cycles in the powerset construction is made of pairwise non-equivalent states with visible probability.
6. It just remains to guarantee that $\{p\}$ is accessible in the subset construction. We use the fact that with high probability, all cycles with length in $\Omega(\ln(n))$ are accessible in a random deterministic automaton [5]. By construction the cycle around p labelled aw built at step 1 has length $\Theta(\log n)$, hence p is accessible with high probability.

The first steps of the proof sketch are depicted in Figure 2, with more details and notations that will be introduced in the next section.

4 Random almost deterministic transition structures

As indicated in the presentation of the proof in Section 3, a convenient way to see a uniform random transition structure is to start with no known transition at all, and generate them on the fly, when needed: we use the fact that the targets of the $2n$ transitions in a size-n uniform transition structure are independent uniform random elements of $[n]$.

Consider for instance that we take a random state s and iteratively follow b-transitions starting from s: we generate the path $s \xrightarrow{b} \delta(s,b) \xrightarrow{b} \delta(s,bb) \xrightarrow{b} \dots$ until we cycle back on a previously seen state. In this process, we keep picking uniformly at random and independently integers in $[n]$ until we have a collision: this is exactly the setting of the classical Birthday Problem. Straightforward computations show that the expected length ℓ_s of this b-path $\mathcal{P}_s$ is in $\Theta(\sqrt{n})$, and that it is between $\sqrt{n}$ and $2\sqrt{n}$ with visible probability.

Now suppose that we want to add the condition that the target of every a-transition outgoing from a state of $\mathcal{P}_s$ is not in $\mathcal{P}_s$. We can proceed as follows: for a given fixed path $\mathcal{P}_s$ of length ℓ_s, the Birthday Problem analysis tells us that with visible probability the outgoing a-transitions do not reach $\mathcal{P}_s$. As long as $\sqrt{n} \leq \ell_s \leq 2\sqrt{n}$, we can lower bound this probability by a constant that does not depend on ℓ_s. Moreover, a given transition structure can have only one b-path from s, so we can partition the set of size-n transition structures according to their b-path, for a given s. Hence a simple computation using the law of total probabilities (or direct counting) shows that we can combine the two "with visible probability" and that, with visible probability there is a b-path $\mathcal{P}_s$ from s of length between $\sqrt{n}$ and $2\sqrt{n}$ such that every outgoing a-transition ends outside $\mathcal{P}_s$.

We detailed this reasoning because it is the main technique we will use in the sequel to build on the previous results and add new constraints, until we exhibit a shape that ensures that applying the accessible powerset construction will produce a large (super-polynomial) number of states. Also, we will rely much on properties derived from the Birthday Problem, such as:

- If we generate $O(\sqrt{n})$ elements of $[n]$, there is no collision with visible probability, even if there is a set of forbidden states of size $O(\sqrt{n})$ which make the process fail.
- If we generate $\Omega(\sqrt{n})$ elements of $[n]$, there is a collision with visible probability, even if there is a set of forbidden states of size $O(\sqrt{n})$ which make the process fail.
- If we generate random elements of $[n]$, with visible probability we hit a fixed set of states of size $\Omega(\sqrt{n})$ before a collision occurs.

4.1 Backward tree

We first look at the shape of a typical backward tree[1] from a state p in a random transition structure $\mathcal{T} = (n, \delta)$. We define $d(x,y)$ as the smallest length of a word w such that $\delta_w(x) = y$ (and ∞ if y is not accessible from x). For a given state p, we consider the backward exploration of $\mathcal{T}$ starting from p: we iteratively build the sets of states $R_i(p) = \{x : d(x,p) = i\}$. For $\tau \geq 1$, the nodes of the *backward tree* of depth τ from p are $B_\tau(p) = \cup_{i=0}^{\tau} R_i(p)$ and the edges are the transitions $x \xrightarrow{\alpha} y$ that go from a state $x \in R_i(p)$ to a state $y \in R_{i-1}(p)$, for $i \in [\tau]$.

We keep building the backward tree until the first time τ where $R_\tau(p) \geq \sqrt{n}$. If it happens, the tree is called the $\sqrt{n}$-backward tree. If the transition structure is taken uniformly at

[1] The backward tree is not a tree in the graph theoretical sense as a node at depth ℓ can have two out-going edges to two different nodes at depth $\ell - 1$.

random, there is a visible probability that $R_\tau(p)$ exists and has size at most $3\sqrt{n}$, that $\tau = \Omega(\log n)$ and that the whole $\sqrt{n}$-backward tree contains at most $O(\sqrt{n})$ nodes.

To see that, first consider $R_1(p)$. Each state $x \neq p$ can be in $R_1(p)$, if there is a transition $x \xrightarrow{a} p$ or $x \xrightarrow{b} p$ (or both) in $\mathcal{T}$. This happens with probability $\pi_n^{(1)} = \frac{2}{n} - \frac{1}{n^2} \approx \frac{2}{n}$. The cardinality of $R_1(p)$ thus follows a binomial law of parameters $n-1$ and $\pi_n^{(1)}$. In particular, in expectation it contains around 2 states.

Assume now that we know all the $R_j(p)$ for $j \leq i$ and want to compute $R_{i+1}(p)$; we suppose that $R_i(p) \neq \emptyset$. Recall that $B_i(p) = \cup_{j=0}^{i} R_j(p)$ and let $k_i = |B_i(p)|$. By definition of d, none of the states of $B_i(p)$ can be in $R_{i+1}(p)$. On the other hand, any state x of $[n] \setminus B_i(p)$ can be in $R_{i+1}(p)$, and the condition that a state is not in $B_i(p)$ is exactly that its outgoing transitions are not in $B_{i-1}(p)$. All other target states are equally likely under this conditioning, for both transitions. Hence there are $n - k_{i-1}$ possible targets for $\delta(x,a)$ and $\delta(x,b)$: the probability that at least one of them is in $R_i(p)$ is $\pi_n^{(i)} = \frac{2|R_i(p)|}{(n-k_{i-1})} - \frac{|R_i(p)|^2}{(n-k_{i-1})^2} \approx \frac{2|R_i(p)|}{n}$ if $|R_i(p)|$ and k_{i-1} are both $o(n)$. Hence the number of elements in $R_{i+1}(p)$ follows a binomial law of parameters $n - k_i$ and $\pi_n^{(i)}$. In particular, in expectation, $R_{i+1}(p)$ is roughly twice as large as $R_i(p)$, as long as they are not too big. Since binomial laws are concentrated around their means, the presentation above can be turned into a formal proof, establishing the following result.

▶ **Lemma 5.** *Let p be a random state of a random n-state deterministic transition structure. With visible probability, the $\sqrt{n}$-backward tree from p exists, has depth $\tau \in \Theta(\log n)$, contains between $\sqrt{n}$ and $3\sqrt{n}$ extremal leaves, i.e. states in $R_\tau(p)$, and has a total number of nodes in $\Theta(\sqrt{n})$.*

In [5], Cai and Devroye also consider backward trees, with a precise analysis for fixed depth (that does not depend on n) conditionally on p being in the large strongly connected component; they use approximation by a Galton-Watson branching process. This allows them to give a more precise analysis on the existence of the circuit we are building in this paper: they prove that conditioned on the fact that p is accessible, there is such a circuit with high probability. However we cannot reuse their result directly, since we need to quantify the amount of randomness used to discover the circuit: we need unset transitions to continue our construction. It is not obvious to describe the distribution of the transitions if we condition on the existence of the circuit (in particular, there can be several such circuits).

In our setting, we have a direct access to the distribution of most unseen transitions. Indeed, if we fix the $\sqrt{n}$-backward tree T_p from p and consider a state x that is not in the tree, its outgoing transitions can end either in $[n] \setminus T_p$ or at an *extremal leaf*, a leaf of maximal depth, of T_p (otherwise x would be in T_p); and every possible state has the same probability. It is a bit more complicated for transitions outgoing from a state of T_p that are not already part of the tree, but we will not use them in our construction; except for p itself, but if we condition on having T_p, its outgoing transitions ends in uniform elements of $[n]$. So as long as we do not consider a transition outgoing from a node of T_p, except p, we can easily perform our probabilistic computations given the $\sqrt{n}$-backward tree of p being T_p. Since the $\sqrt{n}$-backward tree of p of a transition structure is unique if it exists, we can use the law of total probabilities at the end to complete the proof.

Also observe that we cannot hope for a result with high probability in our setting: the probability that p has no incoming transition is $(1 - \frac{1}{n})^{2(n-1)} \approx e^{-2}$ and is therefore visible.

4.2 Forward tree and circuit using $p \xrightarrow{a} r$

We fix the $\sqrt{n}$-backward tree T_p of p that satisfies the conditions of Lemma 5. Then we generate the a-transition $p \xrightarrow{a} r$ outgoing from p: as explained in the previous section, this is a uniform random element of $[n]$. We then begin a process consisting in doing a breadth-first traversal of the transition structure starting from $r_0 := r$. We discover the states $r_0 = \delta(r, \varepsilon)$, $r_1 = \delta(r, a)$, $r_2 = \delta(r, b)$, $r_3 = \delta(r, aa)$, $r_4 = \delta(r, ab)$, $\ldots$, where the words are taken in length-lexicographic order. We continue this process until we reach either some r_i that belongs to T_p, or an already seen r_i ($r_i = r_j$ for some $j < i$). The process is successful if we halt because we hit an extremal leaf of T_p after at most $\sqrt{n}$ steps, otherwise it fails.

Let L_p be the set of extremal leaves of T_p. As mentioned above, since we only discover new states before the last step of the process, the transition considered at time $i \geq 1$ ends in a uniform random state of $([n] \setminus T_p) \cup L_p$: the fact that T_p is the $\sqrt{n}$-backward tree from p prevents transitions from ending at a node of $T_p \setminus L_p$ (the case of time 0 is easily handled separately). Hence we are in a variant of the Birthday Problem: we have a target set L_p of size $\Theta(\sqrt{n})$ and we iteratively draw random numbers of $[n] \setminus T_p \cup L_p$ until we hit L_p (success) or we see an element twice (failure). All the computations are classical even if we ask that the process halts before $\sqrt{n}$ steps. In particular $|[n] \setminus T_p \cup L_p| = n - O(\sqrt{n})$ so we do not differ much from the standard case with parameter n. This yields:

▶ **Lemma 6.** *For the uniform distribution on size-n transition structures having T_p as $\sqrt{n}$-backward tree from p, with visible probability the breadth-first traversal starting at $r := \delta_a(p)$ hits an extremal leaf of T_p before it discovers the same state twice, and it does this in at most $\sqrt{n}$ steps.*

If the conclusions of Lemma 6 hold then there is a word w of length $\Theta(\log n)$ such that $\delta_w(r) = p$, and aw labels a circuit around p: starting from p, we read a to reach r, then we follow the path that hits an extremal leaf of T_p, discovered during the breadth-first traversal; then finally go back to p using the transitions of T_p. Observe that there can be several paths that work in the last part: it is possible that both transitions outgoing from a state at distance $i+1$ from p end in states at distance i. To uniquely determine w, we choose, in this last part, the smallest for the lexicographic order. Doing this still preserves uniqueness in the following sense: for a given transition structure, there is at most one triplet (T_p, r, F_r) such that T_p is the $\sqrt{n}$-backward tree from p, $r = \delta_a(p)$, and F_r is the forward tree from r, and all the properties of Lemma 5 and Lemma 6 are satisfied. The choice of w is then fixed by (T_p, r, F_r), and the uniqueness of the triplet, which exists when all the requirement are fulfilled, allows the use of the law of total probabilities.

Let $p \in [n]$. An n-state transition structure is *p-compatible* if its $\sqrt{n}$-backward tree from p exists and satisfies the conclusions of Lemma 5, and if the breadth-first traversal from r discovers different states that are not in T_p for all labels smaller than z, and $\delta(r, z) \in L_p$, with $|z| \leq \frac{1}{2}\log_2 n$. When the transition structure $\mathcal{T}$ is p-compatible, we define its *p-substructure* as being the incomplete automaton of stateset the states of T_p, r and all the other states discovered during the breadth-first traversal until label z. Its transitions are the transitions of T_p, and all the transitions of the breadth-first search until label z (included). We have:

▶ **Proposition 7.** *With visible probability, an n-state transition structure taken uniformly at random is p-compatible, where p is also taken uniformly at random and independently in $[n]$. In this case, the p-substructure is unique, has $O(\sqrt{n})$ states, and contains a circuit around p labelled aw, where w is uniquely determined using the transitions of the p-structure only and we have $|w| \in \Theta(\log n)$.*

4.3 Discovering the b-threads

Fix a p-substructure X_p and consider the uniform distribution over n-state transition structures that are p-compatible with X_p. For this distribution, if we take a state $s \notin X_p$, its outgoing transitions end in an element of $[n] \setminus T_p \cup L_p$, uniformly at random and independently from the others transitions: the condition that the p-substructure is X_p only forbids these transitions from ending at a node of the $\sqrt{n}$-backward-tree of p that is not an extremal leaf.

We now add a random a-transition $p \xrightarrow{a} q$ to form a random almost deterministic transition structure that has X_p as p-substructure, by picking uniformly at random $q \in [n]$. Since $|X_p| \in O(\sqrt{n})$, with high probability $q \notin X_p$. We fix some $d \geq 1$ from now on, and read, letter by letter, the word $w(aw)^{d-1}$ starting from q, where aw labels the circuit around p in X_p given in Proposition 7. Since w has length $\Theta(\log n)$, the word $w(aw)^{d-1}$ has logarithmic length, and, using the Birthday Problem once again, with high probability we only discover new states that are not in X_p while reading the whole word. In this case, we name $p_0 = p$ and $p_i = \delta(q, w(aw)^{i-1})$ for $i \in [d]$. Observe that in the whole process, we never considered b-transitions starting from one of the p_i, with $0 \leq i \leq d$. Moreover, as explained above, $\delta(p_0, b)$ is a uniform random element of $[n]$ and each $\delta(p_i, b)$ is a uniform random element of $[n] \setminus T_p \cup L_p$, under our conditioning, and it is also the case for every transition outgoing from a newly discovered state.

Let us define the *b-thread* of p_i as the set of all states reached from p_i using words of the form b^j. Discovering state by state such a b-thread consists in iteratively generating the outgoing b-transition of the previous state, which is done by taking a uniform element of $[n] \setminus T_p \cup L_p$. Let us start with the b-thread of p_0. By the Birthday Problem again, with visible probability it cycles back after discovering between $\sqrt{n}$ and $2\sqrt{n}$ states while never discovering a state of X_p, since $|X_p| \in O(\sqrt{n})$. If this happens, we consider the b-thread from p_1. With visible probability, it also cycles back after discovering between $\sqrt{n}$ and $2\sqrt{n}$ states while never discovering a state of X_p or of the b-thread from p_0, as they both have size in $O(\sqrt{n})$. Since d is fixed, doing this for the b-thread starting at each p_i we obtain:

▶ **Lemma 8.** *Let $d \geq 1$. Let X_p be a p-substructure of size-n transition structures. For the uniform distribution on size-n transition structures that are p-compatible and that have X_p as p-substructure, if we add a random transition $p \xrightarrow{a} q$ by choosing q uniformly at random and independently in $[n]$, then with visible probability (i) the states discovered while following the path labeled by $w(aw)^{d-1}$ are all different and do not belong to X_p (ii) the b-threads starting at the p_i's, where $p_0 = p$ and $p_i = \delta(q, w(aw)^{i-1})$, have length between $\sqrt{n}$ and $2\sqrt{n}$, are pairwise disjoint and do not intersect X_p.*

4.4 Cycle lengths and accessibility

An almost deterministic transition structure that satisfies the conditions of Lemma 8 is called *(p, b)-compatible*, and we say that it has b-thread lengths $\vec{\lambda} = (\lambda_0, \ldots, \lambda_d)$ if the b-thread from each p_i as length λ_i. We also define its *(p, b)-substructure* as its p-substructure where we add the states along the path labeled by $w(aw)^{d-1}$ from q and the b-threads from each p_i.

Consider an almost deterministic transition structure $\mathcal{T}$ of given (p, b)-substructure $X_{p,b}$ with b-thread lengths $\vec{\lambda} = (\lambda_0, \ldots, \lambda_d)$ and cycle lengths $\vec{\ell} = (\ell_0, \ldots, \ell_d)$. If $\vec{\ell'} = (\ell'_0, \ldots, \ell'_d)$ is another vector where each $\ell'_i \in [\lambda_i]$, we can re-target the last b-transition of each b-thread so that the cycle lengths are now $\vec{\ell'}$. Thus, conditioned on $\vec{\lambda}$, each cycle length ℓ_i is a uniform random element of $[\lambda_i]$. Since $\sqrt{n} \leq \lambda_i \leq 2\sqrt{n}$, and since each $\ell_i \in [\![\frac{1}{2}\sqrt{n}, \sqrt{n}\,]\!]$, with visible probability the ℓ_i's are uniform and independent random elements of $[\![\frac{1}{2}\sqrt{n}, \sqrt{n}\,]\!]$.

To conclude this part, we generate the initial state i_0 uniformly at random. All our constraints so far hold with visible probability, and one of them implies the existence of a circuit of length $\Omega(\log n)$ around p. Cai and Devroye [5] established that with high probability such a cycle is accessible; the conjunction of a high-probability event with a visible event is still visible. This yields:

▶ **Theorem 9.** *Let $d \geq 1$. There exists a set of almost deterministic transition structures with n states and one initial state $\mathfrak{T}_n$ such that with visible probability for the uniform distribution over size-n almost deterministic transition structure with an initial state, the state p (source of the additional a-transition) is accessible from the initial state and there exists a word w of length $\Theta(\log n)$ such that $\delta(p, w(aw)^{d-1}) = \{p_0, \dots, p_d\}$ is a set of $d+1$ states, and the b-threads starting from the p_i's have lengths λ_i in $[\![\sqrt{n}, 2\sqrt{n}]\!]$ and their cycle length is in $[\![\frac{1}{2}\sqrt{n}, \sqrt{n}]\!]$. Moreover, this set $\mathfrak{T}_n$ can be built so that for the uniform distribution on $\mathfrak{T}_n$, the cycle lengths are uniform and independent random elements of $[\![\frac{1}{2}\sqrt{n}, \sqrt{n}]\!]$.*

If $\mathcal{T}$ is in the set $\mathfrak{T}_n$ and we read b's from $P = \{p_0, \dots, p_d\}$, we eventually reach the b-cycle of P in the accessible powerset transition structure of $\mathcal{T}$, and its length is $\operatorname{lcm}(\ell_0, \dots, \ell_d)$. As the ℓ_i's are uniform and independent random elements of $[\![\frac{1}{2}\sqrt{n}, \sqrt{n}]\!]$, their lcm is $\Omega(n^{\frac{d+1}{2}})$ with visible probability [10], yielding our first main consequence (before adding final states):

▶ **Corollary 10.** *For the uniform distribution on size-n almost deterministic transition structures, the accessible powerset transition structure has a super-polynomial number of states with visible probability.*

5 Adding final states

We are now ready to randomly select which states are final. In our model, for every n, each state is final with fixed probability f_n, which may depend on n as long as it is not too close to either 0 or 1: we require that a set of $\Theta(\sqrt{n})$ states contains both final and non-final states with visible probability. This holds under our condition that f_n and $1 - f_n$ are in $\Omega(\frac{1}{\sqrt{n}})$, as a variant of the Birthday Problem again.

Previously, we exhibited the existence with visible probability of $d+1$ occurrences of b-cycles in a random almost deterministic transition structure, yielding a large b-cycle when applying the powerset construction. We will focus on b-cycles in the sequel, as it turns out to be sufficient to prove our main result. It relies on the notion of primitive words, which we now recall.

Let Γ be a nonempty finite alphabet. If $w \in \Gamma^\ell$ is a word of length ℓ, we write $w = w_0 \cdots w_{\ell-1}$ and use the convention that all indices are taken modulo ℓ: for instance w_ℓ is the letter w_0. A nonempty word w is *primitive* if it is not a non-trivial power of another word: it cannot be written $w = z^k$ for some word z and some $k \geq 2$. If w is primitive, it is easily seen that every circular permutation of w is also primitive. See [15] for a more detailed account on primitive words.

Primitive words appear in our proof with the following observation. If $\mathcal{C} = (c_0, \dots, c_{\ell-1})$ is a b-cycle of states starting at c_0, its *associated word* is the size-ℓ word $v = v_0 \dots v_{\ell-1}$ of $\{0,1\}^\ell$ where $v_i = 1$ if and only if c_i is a final state. Recall that if we start the same cycle elsewhere, at c_i, the associated word $v' = v_i \cdots v_\ell v_0 \cdots v_{i-1}$ is primitive if and only if v is primitive: reading the associated word from any starting state preserves primitivity. A b-cycle is said to be *primitive* if one (equivalently, all) of its associated words is (are) primitive. Our study is based on the following statement.

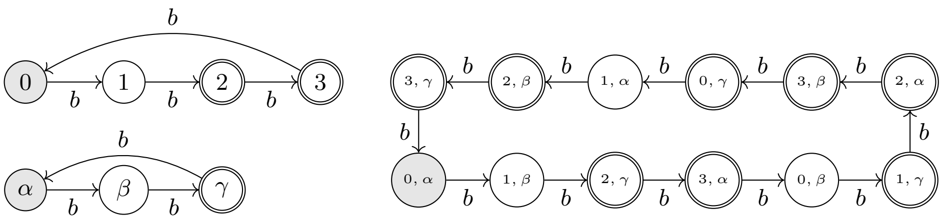

Figure 3 On the left, two primitive b-cycles (accepting states are denoted by double circles) whose associated words are 0011 (top) and 001 (bottom), starting at 0 and α, respectively. On the right, the b-cycle of $\{0, \alpha\}$ of associated word $0011 \odot 001 = 001101111011$, which is primitive by Lemma 12.

▶ **Lemma 11.** *Let $\mathcal{A}$ be a deterministic automaton on Σ and $\alpha \in \Sigma$. If $\mathcal{C}$ is a primitive α-cycle of $\mathcal{A}$, then the states of $\mathcal{C}$ are pairwise non-equivalent: the state complexity of the language recognized by $\mathcal{A}$ is at least $|\mathcal{C}|$.*

So we reduced our problem to studying the primitivity of the b-cycles we built in Section 4, and to how it exports to the associated b-cycle in the powerset construction.

5.1 Some properties of primitive words

If $w^{(1)}$ and $w^{(2)}$ are two non-empty words of respective lengths ℓ_1 and ℓ_2 on the binary alphabet $\{0, 1\}$, we denote by $w^{(1)} \odot w^{(2)}$ the word w of length $\ell = \operatorname{lcm}(\ell_1, \ell_2)$ given by $w_i = 1$ if and only if $w_i^{(1)} = 1$ or $w_i^{(2)} = 1$ (recall that the indices are taken modulo the length of the word). We will see in the sequel that this operation naturally happens when extending the notion of state equivalence from each b-cycle to the corresponding b-cycle in the powerset construction.

▶ **Lemma 12.** *Let $w^{(1)}$ and $w^{(2)}$ be two primitive words on $\{0, 1\}$ of lengths at least 2 that are coprime. Then the word $w^{(1)} \odot w^{(2)}$ is primitive.*

▶ Remark 13. Lemma 12 does not hold if the lengths are not coprime. For instance, if $w^{(1)} = 011111$ and $w^{(2)} = 1011$, then $w^{(1)} \odot w^{(2)} = \underbrace{1 \ldots 1}_{12 \text{ times}}$, which is not primitive.

From a probabilistic point of view, it is well known [15] that a uniform random word is primitive with very high probability. We rely on the following finer result.

▶ **Lemma 14** (De Felice, Nicaud [10]). *Let μ be a probability measure on $\{0, 1\}^n$ such that $\mu(0^n) = \mu(1^n) = 0$ and such that two words with the same number of 0's have same probability. Then the probability that a word is not primitive under μ is at most $\frac{2}{n}$.*

We adapt it to our needs as follows:

▶ **Corollary 15.** *Let f_n be a sequence of real numbers in $(0, 1)$ such that $f_n = \Omega(\frac{1}{\sqrt{n}})$ and $1 - f_n = \Omega(\frac{1}{\sqrt{n}})$. Let ℓ be an integer greater than $\alpha\sqrt{n}$, for a fixed α, and let w be a random binary word of length ℓ whose letters are 1's with probability f_n and 0 with probability $1 - f_n$, independently. Then w is primitive with visible probability.*

5.2 Finalizing the proof of Theorem 3

By Lemma 12, primitivity is preserved by the product $\odot$ when the lengths are coprime, so we restrict the cycle lengths built in Section 4 so that they are pairwise coprime. By Theorem 9, these lengths are uniform random elements of $[\![\frac{1}{2}\sqrt{n}, \sqrt{n}\,]\!]$, we therefore adapt a known result of probabilistic number theory to prove that it still happens with visible probability.

More precisely, Tóth established [18] that the probability that $d+1$ integer taken uniformly at random and independently in $[n]$ are pairwise coprime tends to some positive constant A_{d+1}, generalizing the folklore result that two independent random numbers in $[n]$ are coprime with probability that tends to $\frac{6}{\pi^2}$. This can be used to derive the following variant:

▶ **Corollary 16.** *Let ℓ_0, ℓ_1, …, ℓ_d be $d+1$ integers taken uniformly at random and independently in $[\![\frac{1}{2}\sqrt{n}, \sqrt{n}\,]\!]$. With visible probability, the ℓ_i's are pairwise coprime.*

Combining Corollary 15 and Corollary 16, we can extend Theorem 9 to also require that the b-cycles are primitive and their lengths are pairwise coprime. And this still happens with visible probability.

We can then conclude as follows: if all these requirements are met, the state p is accessible and there is a word z such that $\delta(p, z) = \{p_0, \dots, p_d\}$, the b-threads of the p_i's are pairwise disjoint and eventually form cycles of respective pairwise coprime lengths ℓ_i, and each such cycle is primitive. Moreover, all the ℓ_i are in $\Theta(\sqrt{n})$. By a direct induction on Lemma 12, this yields that the b-cycle of $\{p_0, \dots, p_d\}$ is primitive and has length $\Theta(\sqrt{n}^{d+1})$. By Lemma 11, the language recognized by this almost deterministic automaton has state complexity at least $\Theta(n^{\frac{d+1}{2}})$. This concludes the proof, as it holds for every fixed d.

6 Conclusion and discussion

Our main theorem states that state complexity of a random almost deterministic automaton is greater than n^d with probability at least $c_d > 0$ for n sufficiently large. One can wonder how small the constant c_d is and for which sizes the lower-bound holds. As we said in the introduction, we did not try to estimate c_d nor did we try to optimize its value in this article. Since the powerset construction quickly generates very large automata which would need to be minimized, a proper experimental study does not seem feasible. However, we did generate 1000 almost deterministic transition structures with $n = 100$ states and apply the accessible powerset construction: in 78.6% of the 1000 cases the output had more than n^3 states. This would lead us to guess that even if the constant c_3 that can be derived from our proof is very small, combinatorial explosion does occur frequently in practice.

Also, as noticed above, in our setting it is certain that the property does not hold with high probability, as there is an asymptotically constant probability that the source of the added transition is not accessible. However, this probability is roughly 20.4%, not too far from what we obtained in our experiment on size-100 structures: it is very possible that if we condition the source of the added transition to be accessible, then our result holds with high probability. However, our proof techniques, based on an intensive use of the Birthday Problem cannot prove this: completely new ideas are necessary to establish such a result.

Another natural direction is to consider the case when there are *few* final states, as $\Theta(\sqrt{n})$ final states may be considered too large for a random deterministic automaton. The extreme case is to allow exactly one final state by choosing it uniformly at random. If we do so, our analysis using primitive words fails: with high probability the b-cycles we built have no final state at all, and neither has the associated b-cycle $\mathcal{C}$ in the powerset construction. However, we are confident that our techniques can be used to capture this distribution: by studying the paths ending in this final state, we should be able to find for each b-cycle $\mathcal{C}_i$ a word w_i that maps exactly one state to the final state, and such that the w_i are all different. This would be enough to establish that the states of $\mathcal{C}$ are pairwise non-equivalent and prove the conjecture. Completely formalizing and proving this idea is an ongoing work.

References

1 Louigi Addario-Berry, Borja Balle, and Guillem Perarnau Llobet. Diameter and stationary distribution of random r-out digraphs. *Electronic journal of combinatorics*, 27(P3. 28):1–41, 2020.

2 Frédérique Bassino, Julien David, and Cyril Nicaud. Average case analysis of Moore's state minimization algorithm. *Algorithmica*, 63(1-2):509–531, 2012. doi:10.1007/s00453-011-9557-7.

3 Frédérique Bassino, Julien David, and Andrea Sportiello. Asymptotic enumeration of minimal automata. In Christoph Dürr and Thomas Wilke, editors, *29th International Symposium on Theoretical Aspects of Computer Science, STACS 2012, February 29th - March 3rd, 2012, Paris, France*, volume 14 of *LIPIcs*, pages 88–99. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2012. doi:10.4230/LIPIcs.STACS.2012.88.

4 Mikhail V. Berlinkov. On the probability of being synchronizable. In Sathish Govindarajan and Anil Maheshwari, editors, *Algorithms and Discrete Applied Mathematics - Second International Conference, CALDAM 2016, Thiruvananthapuram, India, February 18-20, 2016, Proceedings*, volume 9602 of *Lecture Notes in Computer Science*, pages 73–84. Springer, 2016. doi:10.1007/978-3-319-29221-2_7.

5 Xing Shi Cai and Luc Devroye. The graph structure of a deterministic automaton chosen at random. *Random Structures & Algorithms*, 51(3):428–458, 2017.

6 Arnaud Carayol and Cyril Nicaud. Distribution of the number of accessible states in a random deterministic automaton. In Christoph Dürr and Thomas Wilke, editors, *29th International Symposium on Theoretical Aspects of Computer Science, STACS 2012, February 29th - March 3rd, 2012, Paris, France*, volume 14 of *LIPIcs*, pages 194–205. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2012. doi:10.4230/LIPIcs.STACS.2012.194.

7 Guillaume Chapuy and Guillem Perarnau. Short synchronizing words for random automata. *CoRR*, abs/2207.14108, 2022. arXiv:2207.14108, doi:10.48550/arXiv.2207.14108.

8 Julien David. Average complexity of Moore's and Hopcroft's algorithms. *Theor. Comput. Sci.*, 417:50–65, 2012. doi:10.1016/j.tcs.2011.10.011.

9 Sven De Felice and Cyril Nicaud. Brzozowski algorithm is generically super-polynomial for deterministic automata. In Marie-Pierre Béal and Olivier Carton, editors, *Developments in Language Theory - 17th International Conference, DLT 2013, Marne-la-Vallée, France, June 18-21, 2013. Proceedings*, volume 7907 of *Lecture Notes in Computer Science*, pages 179–190. Springer, 2013. doi:10.1007/978-3-642-38771-5_17.

10 Sven De Felice and Cyril Nicaud. Average case analysis of Brzozowski's algorithm. *Int. J. Found. Comput. Sci.*, 27(2):109–126, 2016. doi:10.1142/S0129054116400025.

11 Aleksandr Aleksandrovich Grusho. Limit distributions of certain characteristics of random automaton graphs. *Mathematical Notes of the Academy of Sciences of the USSR*, 14(1):633–637, 1973.

12 J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages and Computation.* Addison-Wesley, 1979.

13 Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random Graphs.* 2000.

14 Florent Koechlin, Cyril Nicaud, and Pablo Rotondo. Simplifications of uniform expressions specified by systems. *Int. J. Found. Comput. Sci.*, 32(6):733–760, 2021.

15 Lothaire. *Combinatorics on Words.* Cambridge Mathematical Library. Cambridge University Press, 2 edition, 1997. doi:10.1017/CBO9780511566097.

16 Cyril Nicaud. Random deterministic automata. In Erzsébet Csuhaj-Varjú, Martin Dietzfelbinger, and Zoltán Ésik, editors, *Mathematical Foundations of Computer Science 2014 - 39th International Symposium, MFCS 2014, Budapest, Hungary, August 25-29, 2014. Proceedings, Part I*, volume 8634 of *Lecture Notes in Computer Science*, pages 5–23. Springer, 2014. doi:10.1007/978-3-662-44522-8_2.

17 Cyril Nicaud. The černý conjecture holds with high probability. *J. Autom. Lang. Comb.*, 24(2-4):343–365, 2019. doi:10.25596/jalc-2019-343.

18 László Tóth. The probability that k positive integers are pairwise relatively prime. *Fibonacci Quart*, 40:13–18, 2002.

A Proof of Corollary 4

Proof. Let S be the random variable that maps a random automaton to the state complexity of the language it recognizes. For any $d \geq 1$ and n sufficiently large, we have, for size-n automata: $\mathbb{E}[S] \geq n^d\, \mathbb{P}(S \geq n^d) \geq c_d\, n^d$. Using the notations of Theorem 3, the expected state complexity is at least $c_d\, n^d$ for n large enough. This concludes the proof. ◀

B Technicals lemmas

In this section, we present various technical lemmas that will be used throughout the main proof. This section can be skipped at first reading as it does not provide much in terms of context.

B.1 Birthday problem like results

▶ **Fact 17.** *The following inequalities hold for any $0 \leq x \leq 0.75$: $\exp(-2x) \leq 1 - x \leq \exp(-x)$.*

Proof. Both inequalities follow from convexity of the exponential function. The upper bounds come from comparing it with its linear approximation at $x = 0$; the upper bound is easily proved by checking the sign of the difference at $x = 0$ and at $x = 0.75$. ◀

The following lemma is classical and its proof which is given for the reader's convenience, uses standard arguments.

▶ **Lemma 18.** *Let $r(n)$, $g(n)$ and $t(n)$ be mappings from $\mathbb{N}$ to $\mathbb{N}$ such that for all $n \geq 1$, $r(n) + g(n) + t(n) \leq n$. Consider an urn with n balls numbered from 1 to n with $r(n)$ balls colored red, $g(n)$ balls colored green and the $n - r(n) - g(n)$ other balls colored white. Consider the process of repeatedly drawing a ball uniformly at random with replacement until either a red or green ball is drawn, or a ball previously drawn is drawn again. The following properties hold:*

1. *Let f_n be the probability that the process has not stopped after drawing $t(n)$ balls. If $t(n) \in O(\sqrt{n})$, $r(n) + g(n) \in O(\sqrt{n})$ then there exists a constant $c > 0$ such that $f_n \geq c$ for n large enough.*
2. *Let h_n be the probability that the process stops before $t(n)$ balls have been drawn because a green ball was drawn. If $r(n) \in O(\sqrt{n})$, $g(n) \in \Theta(\sqrt{n})$ and $t(n) \in \Omega(\sqrt{n})$, there exists a constant $c > 0$ such $h_n \geq c$ for n large enough.*
3. *Let i_n be the probability that the process stops after drawing t balls with $t \in [\![\sqrt{n}, 2\sqrt{n}]\!]$ because the t-th ball was already drawn at a previous step ℓ with $t - \ell \in [\![\frac{\sqrt{n}}{2}, \sqrt{n}]\!]$. If $r(n) = O(\sqrt{n})$ and $g(n) \in O(\sqrt{n})$, there exists a constant $c > 0$ such $i_n \geq c$ for n large enough.*

The previous properties still hold if instead of having n balls, we have $b(n) \leq n$ balls with $n - b(n) \in O(\sqrt{n})$.

Proof. Property 1. Assume that $t(n) \in O(\sqrt{n})$, $d(n) = r(n) + g(n) \in O(\sqrt{n})$. For $n \geq 1$, the probability f_n satisfies:

$$f_n = \prod_{k=1}^{t(n)} \left(1 - \frac{d(n) + k - 1}{n}\right)$$

Indeed the probability that the process has not stopped at step $k \in [1, t(n)]$ knowing that it did not stop in the previous $k-1$ steps is $1 - \frac{d(n)+k-1}{n}$ which is the probability of not drawing a red ball or a green ball or one the $k-1$ previously drawn balls which are all distinct and not red or green.

As $d(n)+t(n) \in o(n)$, for n large enough $0 \leq t(n)-1+d(n) \leq 0.75n$ and using Lemma 17, we have:

$$f_n \geq \exp\left(-\frac{2}{n}\sum_{k=1}^{t(n)} d(n)+k-1\right)$$
$$= \exp\left(-\frac{t(n)^2+2t(n)d(n)}{n}+o(1)\right)$$

By assumption $t(n)^2 + 2t(n)d(n)$ is in $O(n)$, so the term inside the exponential can be bounded from below by a real constant $-c_1$ for n large enough, and $f_n \geq e^{-c_1} > 0$. Taking $c = e^{-c_1}$ concludes the proof.

Property 2. Assume that $r(n) \in O(\sqrt{n}), g(n) \in \Theta(\sqrt{n})$ and $t(n) \in \Omega(\sqrt{n})$. For $n \geq 1$ and $\ell \in [1, t(n)]$, we write h_n^ℓ the probability that process stops after drawing ℓ balls because the ℓ-th ball drawn is green. We have:

$$h_n^\ell = \prod_{k=0}^{\ell-2}\left(1 - \frac{g(n)+r(n)+k}{n}\right) \cdot \frac{g(n)}{n}.$$

Indeed the product on the left captures the probability that the process has not stopped before step ℓ (cf. the proof of Property 1) and $\frac{g(n)}{n}$ is the probability to draw a green ball. Using the law of total probabilities, we have $h_n = \sum_{\ell=1}^{t(n)} h_n^\ell$. As $t(n) \in \Omega(\sqrt{n})$, there exists a constant $d > 0$ such that $t(n) \geq d\sqrt{n}$ for n large enough. In particular, for n large enough, we have:

$$h_n \geq \sum_{\ell=\lceil \frac{d}{2}\sqrt{n}\rceil}^{\lfloor d\sqrt{n}\rfloor} h_n^\ell$$

For $\ell \in [\![\frac{d}{2}\sqrt{n}, d\sqrt{n}]\!]$, we have for n large enough:

$$\begin{array}{rcl} h_n^\ell & = & \prod_{k=0}^{\ell-2}\left(1 - \frac{g(n)+r(n)+k}{n}\right) \cdot \frac{g(n)}{n} \\ & \geq & \prod_{k=0}^{\lfloor d\sqrt{n}\rfloor-2}\left(1 - \frac{g(n)+r(n)+k}{n}\right) \cdot \frac{g(n)}{n} \end{array}$$

By Property 1 (taking $t(n) = \lfloor d\sqrt{n}\rfloor - 1$), there exists a constant $c > 0$ such that for n large enough $\prod_{k=0}^{\lfloor d\sqrt{n}\rfloor-2}\left(1 - \frac{g(n)+r(n)+k}{n}\right) \geq c$ and as $g(n) \in \Theta(\sqrt{n})$, there exists a constant $c'' > 0$ such that $g(n) \geq c'\sqrt(n)$ for n large enough and hence for n large enough $h_n^\ell \geq \dfrac{c''}{\sqrt{n}}$ for some constante $c'' > 0$. It follows that for n large enough:

$$g_n \quad \geq \quad \sum_{\ell=\lceil \frac{d}{2}\sqrt{n}\rceil}^{\lfloor d\sqrt{n}\rfloor} g_n^\ell \geq (\lfloor d\sqrt{n}\rfloor - \lceil \tfrac{d}{2}\sqrt{n}\rceil)\frac{c''}{\sqrt{n}} \geq \frac{d}{4}c'' > 0.$$

Property 3. Assume that $r(n) \in O(\sqrt{n})$ and $g(n) \in O(\sqrt{n})$. For $n \geq 1$ and $\ell \in [\![\sqrt{n}, 2\sqrt{n}]\!]$, we write i_n^ℓ the probability that process stops after drawing ℓ balls because the

ℓ-th ball has been drawn at a previous time ℓ' with $\ell - \ell' \in [\![\frac{\sqrt{n}}{2}, \sqrt{n}]\!]$. Let us denote by m_ℓ the number of possible values for ℓ'. For n sufficiently large, $m_\ell \geq \frac{\sqrt{n}}{4}$. We have:

$$i_n^\ell = \prod_{k=0}^{\ell-2} \left(1 - \frac{g(n)+r(n)+k}{n}\right) \cdot \frac{m_\ell}{n}$$

Indeed the production on the left captures the probability that the process has not stopped before step ℓ (cf. the proof of Property 1) and $\frac{m_\ell}{n}$ is the probability to draw one of balls drawn at a time ℓ' with $\ell - \ell' \in [\![\frac{\sqrt{n}}{2}, \sqrt{n}]\!]$. Using the law of total probabilities, we have $i_n = \sum_{\ell \in [\![\sqrt{n}, 2\sqrt{n}]\!]} i_n^\ell$.

For n large enough and $\ell \in [\![\sqrt{n}, 2\sqrt{n}]\!]$,

$$\begin{array}{rcl} i_n^\ell & = & \prod_{k=0}^{\ell-2} \left(1 - \frac{g(n)+r(n)+k}{n}\right) \cdot \frac{m_\ell}{n} \\ & \geq & \prod_{k=0}^{\lfloor 2\sqrt{n} \rfloor - 2} \left(1 - \frac{g(n)+r(n)+k}{n}\right) \cdot \frac{\sqrt{n}}{4n} \end{array}$$

By Property 1 (taking $t(n) = \lfloor 2\sqrt{n} \rfloor - 1$), there exists a constant $c > 0$ such that for n large enough $\prod_{k=0}^{\lfloor 2\sqrt{n} \rfloor - 2} \left(1 - \frac{g(n)+r(n)+k}{n}\right) \geq c$ and therefore, $i_n^\ell \geq \frac{c'}{\sqrt{n}}$ for some constant $c' > 0$. It follows that for n large enough:

$$i_n = \sum_{\ell \in [\![\sqrt{n}, 2\sqrt{n}]\!]} i_n^\ell \geq \left(\lfloor 2\sqrt{n} \rfloor - \lceil \sqrt{n} \rceil\right) \cdot \frac{c'}{\sqrt{n}} \geq \frac{c}{2} > 0.$$

◀

B.2 Concentration results for some binomial distributions

In this section, we give some concentration inequalities for random variables following binomial distributions occurring when drawing the backward tree in a random transition structure. Recall that in this article, we denote by $\mathrm{Bin}(n, p)$ the binomial distribution with n trials each having a probability p of success. These inequalities, derived in Lemma 20, are in fact specialization of the classical Chernoff's inequalities (see for instance [13, Th. 2.1]).

▶ **Theorem 19** (Chernoff inequalities for binomial law). *For a random variable X with the distribution $\mathrm{Bin}(n, p)$, we have, with $\mathbb{E} = np$:*

$$\begin{array}{rclr} \mathbb{P}(X \geq \mathbb{E}(X) + \lambda) & \leq & \exp\left(\frac{-\lambda^2}{2(\mathbb{E}(X) + \frac{\lambda}{3})}\right) & \text{for } \lambda \geq 0; \\ \mathbb{P}(X \leq \mathbb{E}(X) - \lambda) & \leq & \exp\left(\frac{-\lambda^2}{2\mathbb{E}(X)}\right) & \text{for } \lambda \geq 0. \end{array}$$

▶ **Lemma 20.** *For $n \geq 1$, $f \geq 0$ and $t \geq 1$ with $f + t < n$, consider a random variable $X_n^{f,t}$ following the binomial distribution $\mathrm{Bin}\left(n - f - t, \frac{2t}{n-f} - \frac{t^2}{(n-f)^2}\right)$.*
Let $\alpha > 0$ and $\beta > 0$. There exists a constant $\gamma > 0$ such that for all $t < \alpha\sqrt{n}$, $f < \beta\sqrt{n}$ and n sufficiently large,

$$\mathbb{P}(X_n^{f,t} \geq 3t) \leq e^{-\gamma t} \quad \textit{and} \quad \mathbb{P}\left(X_n^{f,t} \leq \frac{3t}{2}\right) \leq e^{-\gamma t}.$$

Proof. The expected value of $X_n^{f,t}$ is:

$$\mathbb{E}(X_n^{f,t}) = (n-f-t)\left(\frac{2t}{n-f} - \frac{t^2}{(n-f)^2}\right) = 2t - \frac{3t^2}{n-f} + \frac{t^3}{(n-f)^2}$$

Let $\delta = t + \frac{3t^2}{n-f} - \frac{t^3}{(n-f)^2}$. Notice that $\mathbb{E}(X_n^{f,t}) + \delta = 3t$. For n sufficiently large, $\delta \geq 0$ as $t \in O(\sqrt{n})$, and we can apply Theorem 19 to obtain the following bound:

$$\mathbb{P}(X_n^{f,t} \geq 3t) = \mathbb{P}(X_n^{f,t} \geq \mathbb{E}(X_n^{f,t}) + \delta) \leq \exp\left(\frac{-\delta^2}{2(\mathbb{E}(X_n^{f,t}) + \frac{\delta}{3})}\right)$$

We have:

$$\frac{-\delta^2}{2(\mathbb{E}(X_n^{f,t}) + \frac{\delta}{3})} = -\frac{3t}{14}\frac{\left(1 + \frac{3t}{n-f} - \frac{t^2}{(n-f)^2}\right)^2}{1 - \frac{6t}{7(n-f)} + \frac{2t^3}{7(n-f)^2}} = -\frac{3t}{14}\underbrace{\frac{\left(1 + O(\frac{1}{\sqrt{n}})\right)^2}{1 + O(\frac{1}{\sqrt{n}})}}_{\geq \frac{2}{3} \text{ for } n \text{ sufficiently large}}$$

Hence for n sufficiently large, we have $\mathbb{P}(X_n^{f,t} \geq 3t) \leq e^{-\frac{t}{7}}$.

Similarly, let $\beta = \frac{t}{2} - \frac{3t^2}{n-f} + \frac{t^3}{(n-f)^2}$. For n sufficiently large, $\beta \geq 0$ and we can apply Theorem 19 to obtain the following bound:

$$\mathbb{P}\left(X_n^{f,t} \leq \frac{3t}{2}\right) = \mathbb{P}(X_n^{f,t} \leq \mathbb{E}(X_n^{f,t}) - \beta) \leq \exp\left(\frac{-\beta^2}{2\mathbb{E}(X_n^{f,t})}\right)$$

We have:

$$\frac{-\beta^2}{2\mathbb{E}(X_n^{f,t})} = -\frac{t}{16}\frac{\left(1 - \frac{6t}{n-f} + \frac{2t^2}{(n-f)^2}\right)^2}{1 - \frac{3t}{2(n-f)} + \frac{t^2}{2(n-f)^2}} = -\frac{t}{16}\underbrace{\frac{\left(1 + O(\frac{1}{\sqrt{n}})\right)^2}{1 + O(\frac{1}{\sqrt{n}})}}_{\geq \frac{1}{2} \text{ for } n \text{ large enough}}.$$

Hence for n sufficiently large: $\mathbb{P}(X_n^{f,t} \leq \frac{3t}{2}) \leq e^{-\frac{t}{32}}$. ◀

▶ **Lemma 21.** *For all $n \geq 1$, consider a the random variable X_n following the binomial distribution $\text{Bin}(n, \frac{2}{n} - \frac{1}{n^2})$. It converges in law to a Poisson Law of parameter 2: for $\ell \geq 0$, $\lim_{n\to\infty} \mathbb{P}(X_n = \ell) = \frac{2^\ell}{\ell!}e^{-2} > 0$.*

Proof. Let $\ell \geq 0$ and $p_n := \frac{2}{n} - \frac{1}{n^2}$. For all $n \geq \ell$,

$$\mathbb{P}(X_n = \ell) = \binom{n}{\ell} p_n^\ell (1-p_n)^{n-\ell} = \binom{n}{\ell}\left(\frac{p_n}{1-p_n}\right)^\ell (1-p_n)^n.$$

As ℓ is fixed, when $n \to \infty$, $\binom{n}{\ell} \sim \frac{n^\ell}{\ell!}$, $\left(\frac{p_n}{1-p_n}\right)^\ell \sim p_n^\ell \sim \frac{2^\ell}{n^\ell}$ and $(1-p_n)^n \sim e^{-2}$. ◀

C Proof of Theorem 9

The aim of the section is to give a detailed proof of Theorem 9. The proof follows the general sketch presented in the article. Recall that in the article we consider a process for generating almost deterministic transition structures which is decomposed into different phases: drawing the transition $p \xrightarrow{a} q$ to be added, drawing the $\sqrt{n}$-backward tree from p, drawing the forward tree p up-to a certain depth, ... Each phase can succeed or fail, we prove for every phase that it succeeds with visible probability conditioned by the fact that the previous phases succeeded. To make it easier to work with these conditioning, we introduce the notion of *transition structure templates* which are incomplete transition structures where the source and target of the extra transition are possibly distinguished and where some states are marked as closed to enforce that no new incoming transitions can enter these states. Instead of conditioning to the success of the previous phases, we define a set of templates which ensure that the previous phases have succeeded and prove that this set of template occurs with visible probabilities.

Once the terminology has been introduced in Section C.1, we will present the detailed outline of the proof in Section C.2 and give the proof in the remaining sections.

C.1 Transition structure templates

A *deterministic transition structure template* $\mathcal{A}$ (or template $\mathcal{A}$ for short) is given by a tuple $(n, \delta, \mathrm{src}(\mathcal{A}), \mathrm{dst}(\mathcal{A}), \mathrm{Closed}(\mathcal{A}))$ where:

- n is the number of states (and $[n]$ is the stateset),
- δ is a partial mapping from $[n] \times \Sigma$ to $[n]$,
- $\mathrm{src}(\mathcal{A}) \in [n]$ and $\mathrm{dst}(\mathcal{A}) \in [n] \cup \{\perp\}$ are two distinguished states which will respectively be the source and target of the newly added a-transition. We allow $\mathrm{dst}(\mathcal{A})$ to be undefined which we signal using the symbol $\perp$,
- $\mathrm{Closed}(\mathcal{A}) \subseteq [n]$ is a distinguished set of states called *closed states*. Closed states will play a role when we define what it means for a template $\mathcal{B}$ to extend a template $\mathcal{A}$.

The support $\mathrm{Support}(\mathcal{A})$ of a template $\mathcal{A}$ is the set of states that are either the source or the target of a transition of $\mathcal{A}$. We denote by Aut_n the set of templates with n states. Remark that all the templates are deterministic as to ease the presentation, we do not add the extra a-transition but mark in the template its source and (possibly) its target.

We now define what it means for a template $\mathcal{B}$ to extend a template $\mathcal{A}$.

▶ **Definition 22** (Extension relation between templates). *For two templates $\mathcal{A}$ and $\mathcal{B}$ with n states, the template $\mathcal{B}$* extends *the template $\mathcal{A}$, denoted by $\mathcal{A} \subseteq \mathcal{B}$ if for all $\alpha \in \Sigma$ and all states r and $s \in [n]$, we have:*

- $r \xrightarrow[\mathcal{A}]{a} s$ *implies* $r \xrightarrow[\mathcal{B}]{a} s$,
- $r \xrightarrow[\mathcal{B}]{a} s$ *implies either* $r \xrightarrow[\mathcal{A}]{a} s$ *or s is not closed in $\mathcal{A}$ (i.e., $s \notin \mathrm{Closed}(\mathcal{A})$),*
- $\mathrm{Closed}(\mathcal{A}) \subseteq \mathrm{Closed}(B)$,
- $\mathrm{src}(\mathcal{A}) = \mathrm{src}(B)$ *and* $\mathrm{dst}(\mathcal{A}) = \mathrm{dst}(B)$ *(if $\mathrm{dst}(\mathcal{A})$ is defined).*

A template $\mathcal{A}$ is *complete* if its transition function is total, $\mathrm{dst}(\mathcal{A})$ is defined and all its states are closed. We denote by CAut_n the set of complete template with n states over the input alphabet Σ. Remark that complete templates with n states are in bijection with almost deterministic transition structures by adding the transition $\mathrm{src}(\mathcal{A}) \xrightarrow{a} \mathrm{dst}(\mathcal{A})$ to a complete template $\mathcal{A}$. We choose to work only with templates to simplify the statements of the various intermediary results.

For a fixed template $\mathcal{B} \in \mathrm{Aut}_n$, the uniform distribution amongst the complete template in CAut_n extending $\mathcal{B}$ can easily be described as shown in the following lemma.

▶ **Lemma 23.** *Let $\mathcal{B} \in \mathrm{Aut}_n$. To draw uniformly at random a complete template $\mathcal{A} \in \mathrm{CAut}_n$ given that $\mathcal{A}$ extends $\mathcal{B}$, it is enough to start from $\mathcal{B}$ and draw independently the target of each missing transition, uniformly at random in the set $[n] \setminus \mathrm{Closed}(\mathcal{B})$.*

For a set $\mathfrak{B}$ of templates (possibly having a different number of states), we denote by $\mathfrak{B}_n$, the subset of $\mathfrak{B}$ containing only the templates in $\mathfrak{B}$ with n states. In the following, we will use gothic letters such as $\mathfrak{B}, \mathfrak{C}, \dots$ to denote sets of templates.

▶ **Definition 24** (Proper set of templates). *A set $\mathfrak{B}$ of templates is called* proper *if for all $n \geq 1$, for all template $\mathcal{A} \in \mathrm{Aut}_n$, $\mathcal{A}$ extends at most one template in $\mathfrak{B}_n$.*

We say that a template $\mathcal{A}$ with n states extends a proper set $\mathfrak{B}$ if it extends (exactly) one template in $\mathfrak{B}_n$. Remark that as $\mathfrak{B}$ is proper, $\mathbb{P}(\mathcal{A} \in \mathrm{CAut}_n \text{ extends } \mathfrak{B}) = \sum_{\mathcal{B} \in \mathfrak{B}_n} \mathbb{P}(\mathcal{A} \in \mathrm{CAut}_n \text{ extends } \mathcal{B})$.

We now define what it means for a proper set of templates to occur with visible probability.

▶ **Definition 25** (Proper set occurring with visible probability). *A proper set of templates $\mathfrak{B}$ is said to* occur with visible probability *if there exists a constant $c > 0$ such that for n sufficiently large, the probability that a complete template picked uniformly at random from CAut_n extends a template in $\mathfrak{B}$ is at least c (i.e., $\mathbb{P}(\mathcal{A} \in \mathrm{CAut}_n \text{ extends } \mathfrak{B}) \geq c$) for n sufficiently large).*

▶ **Definition 26** (Proper set occurring with visible probability in another proper set). *A proper set of templates $\mathfrak{C}$ is said to occur with visible probability in a proper set $\mathfrak{B}$ if there exists a constant $c > 0$ such that for n sufficiently large, for all template $\mathcal{B} \in \mathfrak{B}_n$, the probability that a complete template $\mathcal{A}$ picked uniformly at random in the complete templates extending $\mathcal{B}$ also extends $\mathfrak{C}$ is at least c (i.e., $\mathbb{P}(\mathcal{A} \in \mathrm{CAut}_n \text{ extends } \mathfrak{C} | \mathcal{A} \text{ extends } \mathcal{B}) \geq c$).*

Using the law of total probability, we obtain the following lemma which will be used throughout the proof to establish that our different sets of templates occur with visible probability.

▶ **Lemma 27.** *Let $\mathfrak{B}$ and $\mathfrak{C}$ be two proper sets of templates. Assume that:*

1. *$\mathfrak{B}$ occurs with visible probability,*
2. *$\mathfrak{C}$ occurs with visible probability in $\mathfrak{B}$.*

Then the set $\mathfrak{C}$ also occurs with visible probability.

Proof. Let $c_{\mathfrak{B}} > 0$ and $c_{\mathfrak{C}} > 0$ be the constants witnessing that $\mathfrak{B}$ occurs with visible probability and that $\mathfrak{C}$ occurs with visible probability in $\mathfrak{B}$. As $\mathfrak{B}$ is proper, we can use the law of total probabilities and for n sufficiently large, we have

$$\begin{array}{rl} & \mathbb{P}(\mathcal{A} \in \mathrm{CAut}_n \text{ extends } \mathfrak{C}_n) \\ \geq & \sum_{\mathcal{B} \in \mathfrak{B}_n} \mathbb{P}(\mathcal{A} \in \mathrm{CAut}_n \text{ extends } \mathfrak{C}_n \mid \mathcal{A} \text{ extends } \mathcal{B}) \cdot \mathbb{P}(\mathcal{A} \in \mathrm{CAut}_n \text{ extends } \mathcal{B}) \\ \geq & c_{\mathfrak{C}} \cdot \sum_{\mathcal{B} \in \mathfrak{B}_n} \mathbb{P}(\mathcal{A} \in \mathrm{CAut}_n \text{ extends } \mathcal{B}) \\ = & c_{\mathfrak{C}} \cdot \mathbb{P}(\mathcal{A} \in \mathrm{CAut}_n \text{ extends } \mathfrak{B}_n) \\ \geq & c_{\mathfrak{C}} \cdot c_{\mathfrak{B}} \end{array}$$

◀

C.2 Proof outline

Although the outline of the proof follows the outline presented in the paper, the formalization introduces some nuances and as a results, the intermediary lemmas are not identical but of course the statement of the Theorem 9 is completely equivalent. In Section

We will define proper sets of templates denoted by $\mathfrak{B}$, $\mathfrak{F}$ and $\mathfrak{L}^{(d)}$ for all $d \geq 1$ which intuitively capture the automaton constructed in Section 4.1, Section 4.2 and Section 4.3 of the proof-sketch in the main part of the article.

1. A template $\mathcal{A}$ in $\mathfrak{B}_n$ will be reduced to its $\sqrt{n}$-backward tree from $\mathrm{src}(\mathcal{A})$ which will have a depth in $\Theta(\ln(n))$, a size in $O(\sqrt{n})$ and a number of extremal leaves in $\Theta(\sqrt{n})$. The states appearing in the $\sqrt{n}$-backward tree that are not extremal leaves will be closed in $\mathcal{A}$ and $\mathrm{dst}(\mathcal{A})$ will be undefined.
2. A template $\mathcal{A}$ in $\mathfrak{F}$ will extend some $\mathcal{B} \in \mathfrak{B}$ with $\mathrm{Closed}(\mathcal{A}) = \mathrm{Closed}(\mathcal{B})$ and there will exist a $w \in \Sigma^*$ such that $\mathrm{src}(\mathcal{A}) \xRightarrow[\mathcal{A}]{aw} \mathrm{src}(\mathcal{A})$. If we take $w_{\mathcal{A}}$ minimal in the length-lexicographic order with this property, we will have $|w_{\mathcal{A}}| \in \Theta(\sqrt{n})$. In addition, we will ensure that $\mathrm{Support}(\mathcal{A}) \in O(\sqrt{n})$, $\mathrm{src}(\mathcal{A})$ will have no outgoing b-transition, $\mathrm{dst}(\mathcal{A})$ will still be undefined.
3. For $d \geq 1$, a template $\mathcal{A}$ in $\mathfrak{L}$ will extend some $\mathcal{B} \in \mathfrak{F}$ with $\mathrm{Closed}(\mathcal{A}) = \mathrm{Closed}(\mathcal{B})$. The state $\mathrm{dst}(\mathcal{A})$ is defined and not in $\mathrm{Support}(\mathcal{B}) \cup \mathrm{Closed}(\mathcal{B})$. The transitions in $\mathcal{A}$ which are not in $\mathcal{B}$ are all outside of $\mathrm{Support}(\mathcal{B})$ and can be partitioned into a simple path form $\mathrm{dst}(\mathcal{A})$ labeled by $w_B(aw_B)^{d-1}$, a b-thread from $r_0 = \mathrm{src}(\mathcal{A})$, a b-thread from $r_i = w_{\mathcal{B}}(aw_{\mathcal{B}})^{i-1}$ for $i \in [1, d]$. The lengths of the b-threads are in $[\sqrt{n}, 2\sqrt{n}]$ and the cycle length of these threads are in $[\frac{1}{2}\sqrt{n}, \sqrt{n}]$.

In the following sections, we define these proper sets formally and prove that they occur with visible probability. Finally in Section C.6, we restate Theorem 9 in terms of these proper sets and prove it.

C.3 Backward tree

For $c \geq 1$, we will define the set of templates $\mathfrak{B}^c$. We will show that $\mathfrak{B}^c$ is proper for all $c \geq 1$ (cf. Lemma 29) and that for c sufficiently large, $\mathfrak{B}^c$ occurs with visible probability (cf. Proposition 30). For the following sections, we will take $\mathfrak{B}$ equal to $\mathfrak{B}^{c_0}$ for a fixed c_0 large enough to guaranty the occurrence with visible probability.

Recall that for $\mathcal{A}$ a template with n states and $k \geq 0$, we denote by $R^{\ell}_{\mathcal{A}}$ the set of states $s \in [n]$ such that $d_{\mathcal{A}}(s, \mathrm{src}(\mathcal{A})) = \ell$.

For $c \geq 1$, we define the set $\mathfrak{B}^c$ as the set of all templates $\mathcal{A} \in \mathrm{Aut}_n$ with $n \geq c+1$ such that:

1. $|R^1_{\mathcal{A}}| = c$,
2. there exists a unique $\ell_{\mathcal{A}} \geq 1$ such that $|R^{\ell_{\mathcal{A}}}_{\mathcal{A}}| \geq \sqrt{n}$,
3. for all $k \in [2, \ell_{\mathcal{A}}]$, $\frac{3}{2}|R^{k-1}_{\mathcal{A}}| \leq |R^k_{\mathcal{A}}| \leq 3|R^{k-1}_{\mathcal{A}}|$,
4. for all transition $s \xrightarrow[\mathcal{A}]{\alpha} t$ in $\mathcal{A}$ with $\alpha \in \Sigma$, there exists $k \in [1, \ell_{\mathcal{A}}]$ such that $s \in R^k_{\mathcal{A}}$ and $t \in R^{k-1}_{\mathcal{A}}$: in particular, there are no other transition in $\mathcal{A}$ than the ones building the backward-tree up to depth $\ell_{\mathcal{A}}$.
5. $\mathrm{Closed}(\mathcal{A}) = \bigcup_{k \in [0, \ell_{\mathcal{A}}-1]} R^k_{\mathcal{A}}$ and $\mathrm{dst}(\mathcal{A})$ is undefined.

For $c \geq 1$, $\mathfrak{B}^c$ contains templates that are reduced to their $\sqrt{n}$-backward tree which is of size $O(\sqrt{n})$ with $\Theta(\sqrt{n})$ extremal leaves and a depth in $\Theta(\ln(n))$.

▶ **Lemma 28.** *For all $c \geq 1$ and for all $\mathcal{A} \in \mathfrak{B}^c$, we have:*

- $\mathrm{Closed}(\mathcal{A}) \in O(\sqrt{n})$,
- $|R^{\ell_{\mathcal{A}}}_{\mathcal{A}}| \in O(\sqrt{n})$,
- $|\mathrm{Support}(\mathcal{A})| = |\mathrm{Closed}(\mathcal{A})| + |R^{\ell_{\mathcal{A}}}_{\mathcal{A}}| \in O(\sqrt{n})$,
- $\ell_{\mathcal{A}} \in \Theta(\ln(n))$.

Proof. For all $k \geq 2$, $|R^k_{\mathcal{A}}| \geq (\frac{3}{2})^{k-1}c$ and $|R^k_{\mathcal{A}}| \leq 3^{k-1}c$. As $|R^{\ell_{\mathcal{A}}}_{\mathcal{A}}| \geq \sqrt{n}$, it follows that $3^{\ell_{\mathcal{A}}-1}c \geq \sqrt{n}$ and as $R^{\ell_{\mathcal{A}}-1}_{\mathcal{A}} < \sqrt{n}$, $(\frac{3}{2})^{\ell_{\mathcal{A}}-2}c \geq \sqrt{n}$ and $\ell_{\mathcal{A}} \in \Theta(\ln(n))$. As $|R^{\ell_{\mathcal{A}}-1}_{\mathcal{A}}| < \sqrt{n}$, $R^{\ell_{\mathcal{A}}}_{\mathcal{A}} \leq 3\,|R^{\ell_{\mathcal{A}}-1}_{\mathcal{A}}| \leq 3\sqrt{n}$.

Using the fact that for all $k \geq 2$, $|R^k_{\mathcal{A}}| \leq (\frac{2}{3})^{\ell_{\mathcal{A}}-1-k}|R^{\ell_{\mathcal{A}}-1}_{\mathcal{A}}|$,

$$\begin{array}{rcl} |\mathrm{Closed}(\mathcal{A})| = \sum_{k\in[0,\ell_{\mathcal{A}}-1]} |R^k_{\mathcal{A}}| & = & 1 + c + \sum_{k\in[2,\ell_{\mathcal{A}}-1]} |R^k_{\mathcal{A}}| \\ & \leq & 1 + c + \sum_{k\in[2,\ell_{\mathcal{A}}-1]} (\frac{2}{3})^{\ell_{\mathcal{A}}-1-k}\sqrt{n} \\ & \leq & 1 + c + 3\sqrt{n} \in O(\sqrt{n}) \end{array}$$

◀

We now prove that $\mathfrak{B}^c$ is proper.

▶ **Lemma 29.** *For all $c \geq 1$, $\mathfrak{B}^c$ is a proper set of templates.*

Proof. Let $c \geq 1$. Let $\mathcal{A}, \mathcal{B} \in \mathfrak{B}^c_n$. Assume that there exists a complete template $\mathcal{C} \in \mathrm{CAut}_n$ such that $\mathcal{C}$ extends both $\mathcal{A}$ and $\mathcal{B}$. Let $\mathrm{src} = \mathrm{src}(\mathcal{A}) = \mathrm{src}(\mathcal{B}) = \mathrm{src}(\mathcal{C})$.

By induction on k, let us prove that $R^k_{\mathcal{A}} = R^k_{\mathcal{C}}$ for all $k \in [0, \ell_{\mathcal{A}}]$. For $k = 0$, the property trivially holds. Assume that for some $k \geq 1$, we have shown that for all $i < k$, $R^i_{\mathcal{A}} = R^i_{\mathcal{C}}$, we will show that $R^k_{\mathcal{A}} = R^k_{\mathcal{C}}$. We first show that $R^k_{\mathcal{A}} \subseteq R^k_{\mathcal{C}}$. Let $s \in R^k_{\mathcal{A}}$. By definition of $R^k_{\mathcal{A}}$ there exists a transition $s \xrightarrow[\mathcal{A}]{\alpha} t$ with $t \in R^{k-1}_{\mathcal{A}}$. As $\mathcal{C}$ extends $\mathcal{A}$, this transition also belongs to $\mathcal{C}$ (i.e., $s \xrightarrow[\mathcal{C}]{\alpha} t$) and hence $d_{\mathcal{C}}(s, \mathrm{src}) \leq k$. We cannot have $d_{\mathcal{C}}(s, \mathrm{src}) = i < k$ as $R^i_{\mathcal{C}} = R^i_{\mathcal{A}}$ by induction hypothesis. Hence $d_{\mathcal{C}}(s, \mathrm{src}) = k$ and $R^k_{\mathcal{A}} \subseteq R^k_{\mathcal{C}}$. We now show that $R^k_{\mathcal{C}} \subseteq R^k_{\mathcal{A}}$. Let $s \in R^k_{\mathcal{C}}$; there must exist a transition $s \xrightarrow[\mathcal{C}]{\alpha} t$ with $t \in R^{k-1}_{\mathcal{C}} = R^{k-1}_{\mathcal{A}}$. As $R^{k-1}_{\mathcal{A}} \subseteq \mathrm{Closed}(\mathcal{A})$, this transition must also belong to $\mathcal{A}$ (i.e., $s \xrightarrow[\mathcal{A}]{\alpha} t$) and $d_{\mathcal{A}}(s, \mathrm{src}) \leq k$. We cannot have $d_{\mathcal{A}}(s, \mathrm{src}) = i < k$ as $R^i_{\mathcal{A}} = R^i_{\mathcal{C}}$ by induction hypothesis. Hence $d_{\mathcal{A}}(s, \mathrm{src}) = k$ and $R^k_{\mathcal{C}} \subseteq R^k_{\mathcal{A}}$.

Similarly we have that $R^k_{\mathcal{B}} = R^k_{\mathcal{C}}$ for all $k \in [0, \ell_{\mathcal{B}}]$. This implies that $\ell_{\mathcal{A}} = \ell_{\mathcal{B}} = \ell$ and for all $k \in [0, \ell]$, $R^k_{\mathcal{A}} = R^k_{\mathcal{B}} = R^k_{\mathcal{C}}$. In particular $\mathrm{Closed}(\mathcal{A}) = \mathrm{Closed}(\mathcal{B})$.

For all $k \in [0, \ell - 1]$, for all $t \in R^k_{\mathcal{A}}$ and for all $s \in [n]$, we have $s \xrightarrow[\mathcal{A}]{\alpha} t$ if and only $s \xrightarrow[\mathcal{C}]{\alpha} t$ because $R^k_{\mathcal{A}} \subseteq \mathrm{Closed}(\mathcal{A})$. For all $k \in [0, \ell - 1]$, for all $t \in R^k_{\mathcal{B}}$, $s \in [n]$ and $\alpha \in \Sigma$, we have $s \xrightarrow[\mathcal{A}]{\alpha} t$ if and only $s \xrightarrow[\mathcal{C}]{\alpha} t$ because $R^k_{\mathcal{B}} \subseteq \mathrm{Closed}(\mathcal{A})$.

As all transitions in $\mathcal{A}$ and $\mathcal{B}$ target a state in some $R^k_{\mathcal{A}} = R^k_{\mathcal{B}}$ for $k \in [0, \ell]$ (by Condition 3 in the definition of $\mathfrak{B}^c$), we have shown that $\mathcal{A} = \mathcal{B}$. ◀

▶ **Proposition 30.** *For c sufficiently large, $\mathfrak{B}^c$ occurs with visible probability.*

The remainder of this section is devoted to the proof of Proposition 30.

Proof. Let $c \geq 1$. For a fixed state $p \in [n]$, we will describe a process to draw a complete template $\mathcal{A}$ uniformly at random CAut_n with $\mathrm{src}(\mathcal{A}) = p$. Intuitively this process starts by drawing the transitions from $R^1_{\mathcal{A}}$ to $\mathrm{src}(\mathcal{A})$, then from $R^2_{\mathcal{A}}$ to $R^1_{\mathcal{A}}$, and so on until $R^k_{\mathcal{A}}$ becomes empty or its size becomes greater than $\sqrt{n}$. Once all such transitions have been drawn, the missing transitions are drawn. After proving that this process generates complete templates with uniform probability (amongst the complete templates having $\mathrm{src}(\mathcal{A}) = p$), we use it to obtain a lower-bound δ_c, that only depends on c, for the probability that a random complete template extends $\mathfrak{B}^c$ for n sufficiently large. Finally we show that for c sufficiently large, $\delta_c > 0$.

◦ **Description of the process**

The process builds the template by steps starting with a template with no transitions. At the start of step $i \geq 1$, the process will have created a template $\mathcal{A}_{i-1}$ and two disjoint sets of states S_{i-1} and R_{i-1}. During step i, the process will add transitions to $\mathcal{A}_{i-1}$ to construct $\mathcal{A}_i$ and two disjoint sets of states S_i and R_i. The process will maintain the invariant that the states that are not in $S_i \cup R_i$ do not have out-going transition in $\mathcal{A}_i$. And we will show that for all $i \geq 1$, the set R_i will be equal to the set $R^i_{\mathcal{A}}$ of the template $\mathcal{A}$ produced by the process and $S_i = \bigcup_{0 \leq k \leq i-1} R_k$. Observe that the set B_i defined in the main article is just $B_i = R_i \cup S_i$, but we do not need it in the appendices.

Initially, we take for $\mathcal{A}_0$ a template with n states with no transitions and $\mathrm{src}(\mathcal{A}_0) = p$, $R_0 = \{\mathrm{src}(\mathcal{A}_0)\}$ and $S_0 = \emptyset$.

During step $i + 1 \geq 1$, for each state $s \notin S_i \cup R_i$ and each $\alpha \in \Sigma$, we decide with probability $\frac{|R_i|}{n-|S_i|}$ if we add the α-transition out-going from s. If the transition is added, we draw its target uniformly at random in R_i.
We denote by $\mathcal{A}_{i+1}$ the resulting template. We take R_{i+1} to be the set of states for which a transition was added at this step and take $S_{i+1} = S_i \cup R_i$. If R_{i+1} is empty, $|R_{i+1}| \geq \sqrt{n}$ or $S_{i+1} \cup R_{i+1} = [n]$, we move to the final step.

If we enter the final step after step ℓ, we draw $\mathrm{dst}(\mathcal{A})$ uniformly at random in $[n]$. Then we consider all states $s \in [n]$ and all $\alpha \in \Sigma$ such that the α-transition outgoing from s is missing and we add it as follows:

- if s is equal to $\mathrm{src}(\mathcal{A}_\ell) = p$, we draw the target of the transition uniformly at random in $[n]$,
- if s belongs to R_i for some $i \in [\ell]$, we draw the target uniformly at random in $[n] \setminus S_i$,
- and otherwise if $s \in [n] \setminus S_{\ell+1}$ (with $S_{\ell+1} = S_\ell \cup R_\ell$), we draw a target uniformly at random in $[n] \setminus S_\ell$.

◦ **Proof that the process generates according to the uniform distribution**

Let us show this process constructs a complete template with $\mathrm{src}(\mathcal{A}) = p$ according to the uniform distribution. For this, we fix a complete template $\mathcal{B}$ with $\mathrm{src}(\mathcal{B}) = p$ and $\mathrm{dst}(\mathcal{B}) = q$ and show that it is produced with probability $\left(\frac{1}{n}\right)^{2n+1}$.

Let $\ell \geq 1$ be the maximal value such that $R^\ell_{\mathcal{B}}$ either is empty or $|R^\ell_{\mathcal{B}}| \geq \sqrt{n}$. We only consider the case where $|R^\ell_{\mathcal{B}}| \geq \sqrt{n}$. The analysis for the other cases are similar. In particular, the process enters the final step after step ℓ.

▷ Claim 31. The process can generate $\mathcal{B}$ in the final step if and only if for all $i \in [\ell]$, $R_i = R^i_{\mathcal{B}}$ and the transitions added during step i are precisely the transitions in $\mathcal{B}$ going from $R^i_{\mathcal{B}}$ to $R^{i-1}_{\mathcal{B}}$.

Proof of Claim 31. For the direct implication, assume that $\mathcal{B}$ can be generated in the final step.

Toward a contradiction assume that there exists $i \in [\ell]$ such that $R_i \neq R^i_{\mathcal{B}}$ and take i to be minimal. Assume that there exists $s \in R^i_{\mathcal{B}} \setminus R_i$. By minimality of i, s does not belong to any R_k for $k < i$, so at the end of step i, s has no out-going transition ; as no out-going transition to $S_i = \bigcup_{k<i} R_k = \bigcup_{k<i} R^k_{\mathcal{B}}$ will be added by the process for s, s has not out-going to $\bigcup_{k<i} R^k_{\mathcal{B}}$ which contradicts the fact that $s \in R^i_{\mathcal{B}}$.

Similarly assume that there exists $s \in R_i \setminus R^i_{\mathcal{B}}$. As all transitions of $\mathcal{A}_i$ also belong to $\mathcal{B}$, $d_{\mathcal{A}_i}(s,p) \geq d_{\mathcal{B}}(s,p)$, hence s belongs to $R^k_{\mathcal{B}}$ for $k < i$. By minimality of k, s would belong to R_k for $k-1$ which contradicts the fact that s belongs to R_i.

We have now shown that $R_i = R^i_{\mathcal{B}}$ for all $i \in [\ell]$.

As all transitions added by the process will belong to $\mathcal{B}$, it is enough to show that all transitions of $\mathcal{B}$ from R_i to R_{i-1} are added by the process. Assume toward a contradiction, that for some $i \in [\ell]$, there exists a $s \in R^i_{\mathcal{B}}$ and $t \in R^{i-1}_{\mathcal{B}}$ and $\alpha \in \Sigma$ such that $s \xrightarrow[\mathcal{B}]{\alpha} t$ but $s \xrightarrow{\alpha} t$ is not added during step i. As the transition $R^{i-1}_{\mathcal{B}} = R_i$ it can only be added at step i or in the final step. However in the final step as it source belongs to $R^{\mathcal{B}}_i = R_i$ its target is drawn from $[n] \setminus S_i$ which excludes $R_{i-1} = R^{i-1}_{\mathcal{B}}$ and establishes the contradiction, therefore proving the direct implication.

For the converse implication, assume that for all $i \in [1, \ell+1]$, $R_i = R^i_{\mathcal{B}}$ and the transitions added during step i are precisely the transitions in $\mathcal{B}$ going from $R^i_{\mathcal{B}}$ to $R^{i-1}_{\mathcal{B}}$. Consider a transition $s \xrightarrow[\mathcal{B}]{\alpha} t$ of $\mathcal{B}$ which is missing in $\mathcal{A}_\ell$. If $s = p$ it can be added in the final step, if $s \in R^k_{\mathcal{B}} = R_k$ for some $k \leq \ell$ then its target t cannot belongs to $R^{k'}_{\mathcal{B}}$ for $k' < k-1$ and it cannot belong to $R^k_{\mathcal{B}}$ by assumption, so it can be added by the process. Similarly if s does not belong to any $R^k_{\mathcal{B}}$ (and hence to any R_k), its target can only belong to $[n] \setminus \bigcup_{k\in[\ell-1]} R^k_{\mathcal{B}}$ and can be drawn by the process. ◀

For all $i \in [\ell]$, we denote by $r_i > 0$ the size of $R^i_{\mathcal{B}}$ and by t_i the number of transitions going from $R^i_{\mathcal{B}}$ to $R^{i-1}_{\mathcal{B}}$ in $\mathcal{B}$. We also take $s_0 = 0$, $s_i = 1 + \sum_{k=1}^{i-1} r_k$ for $i \in [1, \ell]$. Assuming that the process can still generate the template $\mathcal{B}$ at the beginning of step i, we will have $|R_i| = r_i$ and $|S_i| = s_i$ for all $i \in [\ell]$.

For $i \in [\ell]$, let p_i be the probability that we can still generate $\mathcal{B}$ at the end of step i knowing that $\mathcal{B}$ could still be generated at the beginning of step i. This probability corresponds to the probability of adding exactly the transitions of $\mathcal{B}$ that go from $R^i_{\mathcal{B}}$ to $R^{i-1}_{\mathcal{B}}$ during step i. Each of the t_i transitions from $R^i_{\mathcal{B}}$ to $R^{i-1}_{\mathcal{B}}$ is added with probability $\frac{|R_{i-1}|}{n-|S_{i-1}|}$ and has a probability $\frac{1}{|R_{i-1}|}$ to have the correct target and there are t_i such transitions. There are $2n - 2 - 2r_1 - \ldots - 2r_{i-1} - t_i = 2n - 2s_i - t_i$ other transitions considered in this step, which are not added with probability $1 - \dfrac{|R_{i-1}|}{n - |S_{i-1}|}$.

We have for all $i \in [\ell]$,

$$p_i = \left(\frac{1}{n - s_{i-1}}\right)^{t_i} \left(1 - \frac{r_{i-1}}{n - s_{i-1}}\right)^{2n-2s_i-t_i}.$$

For the final step and for all $i \in [\ell]$, let q_i be the probability of drawing, in the final step, the missing transitions whose source belongs to R_i in accordance with $\mathcal{B}$ knowing that when entering the final step, the process can still produce $\mathcal{B}$. There are $2r_i - t_i$ missing transitions with source in R_i, each having a probability $\frac{1}{n-s_i}$ to be drawn. So we have for all $i \in [\ell]$,

$$q_i = \left(\frac{1}{n - s_i}\right)^{2r_i - t_i}.$$

Let γ be the probability of drawing $\mathrm{dst}(B)$ and the missing transitions whose source is either $\mathrm{src}(\mathcal{B})$ or a state which does not belong to one of the R_i according to $\mathcal{B}$ again assuming that when entering the final step $\mathcal{B}$ can still be generated:

$$\gamma = \frac{1}{n}\frac{1}{n^2}\left(\frac{1}{n - s_\ell}\right)^{2n-2s_{\ell+1}}.$$

The probability $p_{\mathcal{B}}$ that the process generates $\mathcal{B}$ is:

$$p_{\mathcal{B}} = \left(\prod_{i=1}^{\ell} p_i\right) \cdot \left(\prod_{i=1}^{\ell} q_i\right) \cdot \gamma = \left(\prod_{i=1}^{\ell} p_i q_i\right) \cdot \gamma.$$

Remark that for all $i \in [1, \ell]$, $s_i + r_i = s_{i+1}$, hence :

$$p_i q_i = (n - s_i)^{2n-2s_{i+1}} (n - s_{i-1})^{-2n+2s_i}.$$

Hence

$$\prod_{i=1}^{\ell} p_i q_i = (n-s_\ell)^{2n-2s_{\ell+1}} n^{-2n+2}.$$

It follows that:

$$p_{\mathcal{B}} = (n-s_\ell)^{2n-2s_{\ell+1}} n^{-2n+2} \frac{1}{n^3} (\frac{1}{n-s_\ell})^{2n-2s_{\ell+1}} = \left(\frac{1}{n}\right)^{2n+1}.$$

So we have proved that the process generates according to the uniform distribution.

◦ **Lower-bound for the probability to extend $\mathfrak{B}^c$**

We now want to show that for c large enough, there exists a constant $\delta_c > 0$ such that for n large enough the following event, denoted by $X^{(n)}$, occurs with probability at least δ_c:

- the process enters the final step after step ℓ because $r_\ell \geq \sqrt{n}$ for some $\ell \geq 2$,
- $r_1 = c$ and $r_i \in [\frac{3}{2} r_{i-1}, 3r_{i-1}]$ for all $i \in [2, \ell]$,
- during the final step, all missing transitions for vertices in R_i are drawn in $[n] - S_\ell$ for $i \in [0, \ell - 1]$.

Using Claim 31, assuming that $X^{(n)}$ occurs, the template $\mathcal{A}$ drawn at the end of the process extends the template $\mathcal{A}_\ell$ drawn at the end of step ℓ if we set $\text{Closed}(\mathcal{A}_\ell) = S_\ell$ and hence it extends $\mathfrak{B}^c$. Therefore, this is enough to establish the proposition.

We first need some notations to describe the different sizes for the set R_i's that can occur during the process.

For $k \in [1, n]$, $V_k^{(n)}$ denotes the set of k-tuples $(r_1, \ldots, r_k) \in \mathbb{N}^k$ such that $r_1 + \ldots + r_k \leq n$ with $r_1 + \cdots + r_{k-1} < n$ and $0 < r_i < \sqrt{n}$ for all $i \leq k-1$. We take $V^{(n)} = \bigcup_{k=1}^{n} V_k^{(n)}$ which corresponds to the sizes for the sets R_i that can occur at the end of step k.

For $k \in [n]$ and $\overline{r} = (r_1, \ldots, r_k) \in V_k^{(n)}$, we say that $\overline{r}$ succeeds for c if $r_1 = c$ and $r_i \in [\frac{3}{2} r_{i-1}, 3r_{i-1}]$ for all $i \in [2, k]$ and $r_k \geq \sqrt{n}$.

For $k \in [n]$ and $\overline{r} = (r_1, \ldots, r_k) \in V_k^{(n)}$, we say that $\overline{r}$ fails for c if either $k = 1$ and $r_1 \neq c$, or $k \geq 2$, $r_1 = c$, $r_i \in [\frac{3}{2} r_{i-1}, 3r_{i-1}]$ for all $i \in [2, k-1]$ and $r_k \notin [\frac{3}{2} r_{k-1}, 3r_{k-1}]$.

▶ Remark 32. If $\overline{r} = (r_1, \ldots, r_k) \in V_n^{(n)}$ succeeds or fails for c then $s_{k-1} = 1 + \sum_{i=1}^{k-1} r_i \leq 4\sqrt{n}$ for n large enough. Indeed $s_{k-1} \leq 1 + c + \sum_{i=2}^{k-1} (\frac{2}{3})^{k-1-i} r_{k-1} \leq 1 + c + r_{k-1} \sum_{j=0}^{\infty} (\frac{2}{3})^j \leq 1 + c + 3\sqrt{n}$.

For $\overline{r} = (r_1, \ldots, r_k) \in V^{(n)}$, we consider the event $E^{(n)}(\overline{r})$ that the process reaches the end of step k having drawn sets $R_1, \ldots, R_k$ of respective sizes $r_1, \ldots, r_k$. For a subset M of $V^{(n)}$, we denote by $E^{(n)}(M) = \bigcup_{\overline{r} \in M} E^{(n)}(\overline{r})$. Instead of writing $\mathbb{P}(E^{(n)}(M))$, we will simply write $\mathbb{P}(M)$.

▷ Claim 33. For $n \geq 10$, $\mathbb{P}(\overline{r} \in V^{(n)}$ succeeds for $c) + \mathbb{P}(\overline{r} \in V^{(n)}$ fails for $c) = 1$.

Proof. Let $n \geq 10$. Let $P^{(n)}$ be the set of $\overline{r} = (r_1, \ldots, r_k) \in V^{(n)}$ with $k \leq n$ such that either $r_k \geq \sqrt{n}$, $r_k = 0$ or $r_1 + \ldots + r_k = n$. $P^{(n)}$ denotes the set of tuples of sizes that can occur when the process enters the final step. In particular, $\mathbb{P}(\overline{r} \in P^{(n)}) = 1$.

The key property here is that for any $\overline{r} = (r_1, \ldots, r_k) \in P^{(n)}$, either $\overline{r}$ succeeds for c or there exists a prefix $\overline{r'}$ of $\overline{r}$ which fails for c. To see this, consider $\overline{r} = (r_1, \ldots, r_k) \in P^{(n)}$ such that no prefix of $\overline{r}$ fails for c. If $r_k \geq \sqrt{n}$, $\overline{r}$ succeeds. If $r_k < \sqrt{n}$, we must have $r_1 + \cdots + r_k = n$ with $r_1 = c$ and for all $i \in [2, k]$, $r_i \in [\frac{3}{2} r_{i-1}, 3r_{i-1}]$. We will see that this situation cannot occur. Indeed we have:

$$n = r_1 + \cdots + r_k \leq r_k \left(1 + \frac{2}{3} + \cdots + \left(\frac{2}{3}\right)^{k-1}\right) \leq r_k \cdot \sum_{i=0}^{\infty} \left(\frac{2}{3}\right)^i = 3r_k \leq 3\sqrt{n}$$

Hence for $n \geq 10$, this situation cannot occur.

Using the law of total probabilities,

$$\begin{aligned}\mathbb{P}(\overline{r} \in P^{(n)}) &= \mathbb{P}(\overline{r} \in P^{(n)} \text{ succeeds for } c) + \sum_{\substack{\overline{s} \in V^{(n)} \\ \overline{s} \text{ fails for } c}} \sum_{\substack{\overline{r} \in P^{(n)}: \\ \overline{s} \text{ prefix of } \overline{r}}} \mathbb{P}(\overline{r}) \\ &= \mathbb{P}(\overline{r} \in V^{(n)} \text{ succeeds for } c) + \sum_{\substack{\overline{s} \in V^{(n)} \\ \overline{s} \text{ fails for } c}} \mathbb{P}(\overline{s}) \\ &= \mathbb{P}(\overline{r} \in V^{(n)} \text{ succeeds for } c) + \mathbb{P}(\overline{r} \in V^{(n)} \text{ fails for } c)\end{aligned}$$

◀

In the rest of the proof, we will use the following fact which directly follows from the definition of the process.

▷ Claim 34. Assuming that the process enters step k having previously drawn sets $R_1, \ldots, R_{k-1}$ of respective size $r_1, \ldots, r_{k-1}$, the size r_k of the set R_k drawn in step k follows the distribution $\text{Bin}\left(n - s_{k-1} - r_{k-1}, \dfrac{2r_{k-1}}{n - s_{k-1}} - \dfrac{r_{k-1}^2}{(n - s_{k-1})^2}\right)$.

Proof of Claim 34. Each state s in $[n] \setminus S_{k-1} \cup R_{k-1}$ has two missing transitions, each having probability $\frac{r_{k-1}}{n-s_{k-1}}$ to add s to R_k. Hence each of these $n - s_{k-1} - r_{k-1}$ states belongs to R_k with probability $\dfrac{2r_{k-1}}{n - s_{k-1}} - \dfrac{r_{k-1}^2}{(n - s_{k-1})^2}$. ◀

▷ Claim 35. For c large enough, there exists a constant $\gamma > 0$ such that for n large enough, the probability that the process reaches the final step without failing is:

$$\mathbb{P}(\overline{r} \in V^{(n)} \text{ succeeds for } c) \geq \gamma.$$

Proof of Claim 35. Using Claim 33,

$$\mathbb{P}(\overline{r} \in V^{(n)} \text{ succeeds for } c) = 1 - \sum_{k=1}^{n} \mathbb{P}(\ \overline{r} \in V_k^{(n)} \text{ fails for } c)$$

If we denote by $p_c^{(n)}$ the probability that R_1 has size c, we have $\mathbb{P}(\ \overline{r} \in V_1^{(n)} \text{ fails for } c) = 1 - p_c^{(n)}$. For $k \geq 2$,

$$\mathbb{P}(\ \overline{r} \in V_k^{(n)} \text{ fails for } c) = \sum_{\substack{\overline{r} \in V_k^{(n)} \\ \text{fails for } c}} \mathbb{P}(E^{(n)}(r_1, \ldots, r_k) | E^{(n)}(r_1, \ldots, r_{k-1})) \mathbb{P}(E^{(n)}(r_1, \ldots, r_{k-1}))$$

For $(r_1, \ldots, r_k)$ which fails for c, we either have $r_k < \frac{3}{2} r_{k-1}$ or $r_k > 3r_{k-1}$. By Remark 32, $s_{k-1} < 4\sqrt{n}$ for n large enough. Combining Claim 34 and Lemma 20 (with $t = r_{k-1}$ and $f = s_{k-1}$), there exists a constant $\beta > 0$ such that for n large enough:

$$\mathbb{P}(E^{(n)}(r_1, \ldots, r_k) | E^{(n)}(r_1, \ldots, r_{k-1})) \leq 2e^{-\beta r_{k-1}} \leq 2e^{-\beta (\frac{3}{2})^{k-2} c}$$

Hence for n large enough,

$$\begin{aligned}\mathbb{P}(\ \overline{r} \in V_k^{(n)} \text{ fails for } c) &= \sum_{\substack{\overline{r} \in V_k^{(n)} \\ \text{fails for } c}} \mathbb{P}(E^{(n)}(r_1, \ldots, r_k) | E^{(n)}(r_1, \ldots, r_{k-1})) \mathbb{P}(E^{(n)}(r_1, \ldots, r_{k-1})) \\ &\leq 2e^{-\beta (\frac{3}{2})^{k-2} c} \underbrace{\sum_{\substack{\overline{r} \in V_k^{(n)} \\ \text{fails for } c}} \mathbb{P}(E^{(n)}(r_1, \ldots, r_{k-1}))}_{\leq p_c^{(n)}} \\ &\leq 2e^{-\beta (\frac{3}{2})^{k-2} c} p_c^{(n)}\end{aligned}$$

Hence, for n large enough,

$$\begin{aligned}\mathbb{P}(\ \overline{r} \in V^{(n)} \text{ succeeds for } c) &\geq 1-(1-p_c^{(n)})-\textstyle\sum_{k=2}^{n} 2e^{-\beta(\frac{3}{2})^{k-2}c}p_c^{(n)} \\ &\geq p_c^{(n)}(1-2\underbrace{\sum_{i=0}^{\infty} e^{-\beta(\frac{3}{2})^{i}c}}_{=\lambda_c})\end{aligned}$$

As for $i \geq 5$, it holds that $(\frac{3}{2})^i \geq \frac{3}{2}i$, we have:

$$\begin{aligned}\lambda_c &\leq \textstyle\sum_{i=0}^{4} e^{-\beta(\frac{3}{2})^{i}c} + \sum_{i=5}^{\infty}(e^{-\beta(\frac{3}{2})c})^i \\ &= \textstyle\sum_{i=0}^{4} e^{-\beta(\frac{3}{2})^{i}c} + \dfrac{e^{-\beta 5(\frac{3}{2})c}}{1-e^{-\beta(\frac{3}{2})c}}\end{aligned}$$

Hence λ_c tends to 0 as c tends to infinity. In particular, for c large enough, $\lambda_c < 1/4$. If we take such a c, we can conclude using Lemma 21, which ensures that $p_c^{(n)}$ tends to a constant as n tends to infinity. ◀

▷ Claim 36. There exists a constant $\beta > 0$ such that for n large enough and for all $\overline{r} \in V_k^{(n)}$ that does not fail: $\mathbb{P}(X^{(n)}|E^{(n)}(\overline{r})) \geq \beta$.

Proof of Claim 36. Let $\overline{r} = (r_1, \ldots, r_k) \in V_k^{(n)}$ such that $\overline{r}$ does not fail. Assume that the process reaches the final step having generated sets $R_1, \ldots, R_k$ of respective sizes $r_1, \ldots, r_k$. Let $t_1, \ldots, t_{k-1}$ be the number of transitions missing for the states in $R_1, \ldots, R_{k-1}$ at the beginning of the final step. Remark that $t_i \leq 2r_i$ for all $i \in [1, k-1]$.

The probability p that for all $i \in [0, k-1]$, none of the missing transitions with a source in R_i has its target in S_k is:

$$\begin{aligned}p &= \left(\tfrac{n-s_k}{n}\right)^2 \cdot \textstyle\prod_{i=1}^{k-1}\left(\tfrac{n-s_k}{n-s_i}\right)^{t_i} \\ &\geq \left(\tfrac{n-s_k}{n}\right)^{2+t_1+\cdots+t_{k-1}} \\ &\geq \left(\tfrac{n-s_k}{n}\right)^{2s_{k-1}} \geq \left(\tfrac{n-4\sqrt{n}}{n}\right)^{8\sqrt{n}},\end{aligned}$$

as by Remark 32, $s_k \leq 4\sqrt{n}$ for n large enough. We have the following lower-bound for p independently of the t_i's and the r_i's for n large enough.

$$p \geq \underbrace{\left(\frac{n-4\sqrt{n}}{n}\right)^{8\sqrt{n}}}_{\to e^{-32}>0}$$

Hence there exists $\beta > 0$ such that for n large enough $\mathbb{P}(X^{(n)}|E^{(n)}(\overline{r})) \geq \beta$. ◀

For c large enough and n large enough, we have:

$$\begin{aligned}\mathbb{P}(X^{(n)}) &= \sum_{\substack{\overline{r}\in V^{(n)} \\ \overline{r} \text{ succeeds for } c}} \mathbb{P}\left(X^{(n)} \cap E^{(n)}(\overline{r})\right) \\ &= \sum_{\substack{\overline{r}\in V^{(n)} \\ \overline{r} \text{ succeeds for } c}} \mathbb{P}\left(X^{(n)}|E^{(n)}(\overline{r})\right) \cdot \mathbb{P}(E^{(n)}(\overline{r})) \\ &\geq \beta \cdot \left(\sum_{\substack{\overline{r}\in V^{(n)} \\ \overline{r} \text{ does not fail for } c}} \mathbb{P}(E^{(n)}(\overline{r}))\right) && \text{by Claim 36} \\ &\geq \gamma\beta && \text{by Claim 35}\end{aligned}$$

This concludes the proof of Proposition 30. ◀

C.4 Forward tree

We denote by $\mathfrak{F}$ the set of templates $\mathcal{A} \in \mathrm{Aut}_n$ such that:

1. $\mathcal{A}$ extends some (unique) template $\mathcal{B} \in \mathfrak{F}$ with ℓ_B the first index such that $R_B^{\ell_B} \geq \sqrt{n}$,
2. $\mathrm{Closed}(\mathcal{A}) = \mathrm{Closed}(\mathcal{B})$,
3. Let Paths be the set of words of the form aw with $w \in \Sigma^*$ with $|w| \leq \frac{\ln(n)}{2}$. Let $u_1, \dots, u_m$ be an enumeration of the words in Paths in an increasing length-lexicographic order. There exists $t_{\mathcal{A}} \in [1, m]$ such that:
   a. for all $i \neq j \in [1, t_{\mathcal{A}}]$, $\delta_{\mathcal{A}}(\mathrm{src}(\mathcal{A}), u_i) \neq \delta_{\mathcal{A}}(\mathrm{src}(\mathcal{A}), u_j)$
   b. for all $i \in [1, t_{\mathcal{A}} - 1]$, $\delta_{\mathcal{A}}(\mathrm{src}(\mathcal{A}), u_i) \notin R_{\mathcal{B}}^{\ell_{\mathcal{B}}}$ and $\delta_{\mathcal{A}}(\mathrm{src}(\mathcal{A}), u_{t_{\mathcal{A}}}) \in R_{\mathcal{B}}^{\ell_{\mathcal{B}}}$.
   c. every transition belonging to $\mathcal{A}$ but not $\mathcal{B}$ is of the form $\delta_{\mathcal{A}}(\mathrm{src}(\mathcal{A}), u) \xrightarrow{\alpha} \delta_{\mathcal{A}}(\mathrm{src}(\mathcal{A}), u\alpha)$ with $u\alpha = u_i$ for some $i \in [1, t_{\mathcal{A}}]$.

▶ **Lemma 37.** *For all $c \geq 1$ and for all $\mathcal{A} \in \mathfrak{F}$, we have:*

1. $\mathrm{Closed}(\mathcal{A}) \in O(\sqrt{n})$, $\mathrm{Support}(\mathcal{A}) \in O(\sqrt{n})$,
2. *there exists a word $w \in \Sigma^+$ such that $\mathrm{src}(\mathcal{A}) \xRightarrow[\mathcal{A}]{w} \mathrm{src}(\mathcal{A})$ and if we take $w_{\mathcal{A}}$ to be the smallest such word for the length-lexicographic ordering, we have $|w_{\mathcal{A}}| \in \Theta(\ln(n))$.*

Proof. For the proof of Property 1. Let $\mathcal{A}$ be a template in $\mathfrak{F}$. By definition $\mathcal{A}$ extends some $\mathcal{B} \in \mathfrak{B}$. By Lemma 28, $|\mathrm{Closed}(\mathcal{B})| \in O(\sqrt{n})$ and hence $|\mathrm{Closed}(\mathcal{A})| = |\mathrm{Closed}(\mathcal{B})| \in O(\sqrt{n})$. As at most $|\mathrm{Paths}| \leq \sqrt{n}$ transitions belong to $\mathcal{A}$ and not $\mathcal{B}$, $|\mathrm{Support}(\mathcal{A})| \leq |\mathrm{Support}(\mathcal{B})| + \sqrt{n} \in O(\sqrt{n})$ by Lemma 28.

For the proof of Property 2, remark that as $\mathrm{Closed}(\mathcal{B}) = \bigcup_{k \in [0, \ell_{\mathcal{B}} - 1]} R_{\mathcal{B}}^k$ and $\mathcal{A}$ extends $\mathcal{B}$, for all $k \in [0, \ell_{\mathcal{B}}]$, $R_{\mathcal{A}}^k = R_{\mathcal{B}}^k$.

We know that $\delta_{\mathcal{A}}(\mathrm{src}(\mathcal{A}), u_{t_{\mathcal{A}}}) \in R_{\mathcal{B}}^{\ell_{\mathcal{B}}} = R_{\mathcal{A}}^{\ell_{\mathcal{B}}}$ and hence there exists a word $v \in \Sigma^{\ell_{\mathcal{B}}}$ such that $\mathrm{src}(\mathcal{A}) \xRightarrow[\mathcal{A}]{u_{t_{\mathcal{A}}} v} \mathrm{src}(\mathcal{A})$. Hence $|w_{\mathcal{A}}| \leq \ell_{\mathcal{B}} + \sqrt{n} + 1 \in O(\sqrt{n})$ by Lemma 28.

Let $w \in \Sigma^+$ be a word such that $\mathrm{src}(\mathcal{A}) \xRightarrow[\mathcal{A}]{w} \mathrm{src}(\mathcal{A})$. As $\mathrm{src}(\mathcal{A})$ has no outgoing b-transition in $\mathcal{A}$, $w = au$ for some $u \in \Sigma^*$. Let $t = \delta_{\mathcal{A}}(\mathrm{src}(\mathcal{A}), a)$. As $\delta_{\mathcal{A}}(t, u) = \mathrm{src}(\mathcal{A})$, t belongs to $R_{\mathcal{A}}^{\ell}$ for some $\ell \geq 0$. As $\mathrm{src}(\mathcal{A})$ has no out-going transitions in $\mathcal{B}$, the transition $\mathrm{src}(\mathcal{A}) \xrightarrow{a} t$ was added in $\mathcal{A}$ and as for all $k \in [0, \ell_{\mathcal{B}} - 1]$, $R_{\mathcal{A}}^k = R_{\mathcal{B}}^k$ is closed in $\mathcal{B}$, it follows that $\ell \geq \ell_{\mathcal{B}}$. Hence $|w| \geq 1 + \ell_{\mathcal{B}} \in \Omega(\ln(n))$ by Lemma 28. Hence $|w_{\mathcal{A}}| \in \Omega(\ln(n))$.

◀

▶ **Lemma 38.** *The set of templates $\mathfrak{F}$ is proper.*

Proof. Let $\mathcal{A}, \mathcal{B} \in \mathfrak{F}$ be two templates and a complete template $\mathcal{C}$ such that $\mathcal{C}$ extends both $\mathcal{A}$ and $\mathcal{B}$.

As $\mathfrak{B}$ is proper, there exists a unique automaton $\mathcal{D}$ in $\mathfrak{B}$ such that $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ all extend the same template $\mathcal{D}$. Let $\mathrm{src} = \mathrm{src}(\mathcal{A}) = \mathrm{src}(\mathcal{B}) = \mathrm{src}(\mathcal{C}) = \mathrm{src}(\mathcal{D})$.

Let $t = \min(t_{\mathcal{A}}, t_{\mathcal{B}})$. For all $i \in [1, t]$, $\delta_{\mathcal{A}}(\mathrm{src}, u_i) = \delta_{\mathcal{C}}(\mathrm{src}, u_i) = \delta_{\mathcal{B}}(\mathrm{src}, u_i)$. Hence $t_{\mathcal{A}} = t_{\mathcal{B}}$. By Condition 3.a of the definition of $\mathfrak{F}$, this implies that $\mathcal{A} = \mathcal{B}$. ◀

▶ **Proposition 39.** *The set of template $\mathfrak{F}$ occurs with visible probability.*

Proof. By Lemma 27, it is enough to show that $\mathfrak{F}$ occurs with visible probability in $\mathfrak{B}$.

Let $\mathcal{B}$ in $\mathfrak{B}_n$. We want to provide a lower-bound for $\mathbb{P}(\mathcal{A} \in \mathrm{CAut}_n \text{ extends } \mathfrak{F} | \mathcal{A} \text{ extends } \mathcal{B})$.

By Lemma 23, to draw uniformly at random a complete automaton $\mathcal{A} \in \mathrm{CAut}_n$ knowing that it extends $\mathcal{B}$, it is enough to start from $\mathcal{B}$ and draw independently $\mathrm{dst}(\mathcal{A})$ uniformly at random in $[n]$ and the targets of all missing transitions uniformly at random in $[n] \setminus \mathrm{Closed}(\mathcal{B})$.

We will now describe a process which draws the target of the missing transitions in $\mathcal{B}$ in a particular order but still independently and uniformly at random in $[n] \setminus \mathrm{Closed}(B)$ and draws $\mathrm{dst}(\mathcal{A})$ uniformly at random in $[n]$.

The process starts with the template $\mathcal{B}$ and at each step draws the target of a transition which is missing so far or does nothing at this step. If $\mathcal{C}$ is the automaton built at some step of the process, we will say that we try to draw the transition for a word $u\alpha \in \Sigma^+$ with $u \in \Sigma^*$ and $\alpha \in \Sigma$ from $s \in [n]$ to mean that that if $\delta_{\mathcal{C}}(s, u)$ is defined and $\delta_{\mathcal{C}}(s, u\alpha)$ is not, we draw the target of the missing α-labelled transition outgoing from $\delta_C(s, u)$ uniformly at random in $[n] \setminus \mathrm{Closed}(B)$ and otherwise we do nothing.

Recall that Paths denotes the set of words of the form aw with $w \in \Sigma^*$ with $|w| \leq \frac{\ln(n)}{2}$, and that $u_1, \ldots, u_m$ is an enumeration increasing for the length-lexicographic order of the words in Paths. The process tries to draw the transitions for the words $u_1, \ldots, u_m$ successively. Then it draws $\mathrm{dst}(\mathcal{A})$ uniformly at random in $[n]$ and the target of each missing transition uniformly at random in $[n] \setminus \mathrm{Closed}(\mathcal{B})$.

Consider an urn containing the $b(n)$ states in $[n] \setminus \mathrm{Closed}(\mathcal{B})$ where the $g(n)$ states in $R_{\mathcal{B}}^{\ell_{\mathcal{B}}}$ are colored green. The probability that the process described above produces a template extending $\mathfrak{F}$ is equal to the probability of drawing a green without picking the same ball twice when drawing with replacement in the urn in less than $t(n) = |\mathrm{Paths}| \in O(\sqrt{n})$ draws. As $n - b(n) = |\mathrm{Closed}(\mathcal{B})| \in O(\sqrt{n})$ and $g(n) \in \Theta(\sqrt{n})$ (by Lemma 28), we can use Property 2 of Lemma 18 to conclude that there exists a constant $\gamma > 0$ such that for n sufficiently large, $\mathbb{P}(\mathcal{A} \in \mathrm{CAut}_n \text{ extends } \mathfrak{F} | \mathcal{A} \text{ extends } \mathcal{B}) \geq \gamma$. This concludes the proof. ◀

C.5 Discovering the b-threads

Let $n \geq 0$ and $d \geq 1$. Consider a template $\mathcal{B} \in \mathfrak{F}_n$ with $w_{\mathcal{B}} \in \Sigma^*$ the smallest word for the length-lexicographic order such that $\mathrm{src}(\mathcal{B}) \xrightarrow[\mathcal{B}]{aw_{\mathcal{B}}} \mathrm{src}(\mathcal{B})$. For all $(d+1)$-tuples $\overline{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_d) \in \mathbb{N}^{d+1}$ and $\overline{\ell} = (\ell_0, \ell_1, \ldots, \ell_d) \in \mathbb{N}^{d+1}$, we say that a template $\mathcal{A} \in \mathrm{Aut}_n$ is $(\mathcal{B}, \overline{\lambda}, \overline{\ell})$-shaped if:

1. $\mathcal{A}$ extends $\mathcal{B}$ with $\mathrm{Closed}(\mathcal{A}) = \mathrm{Closed}(\mathcal{B})$,
2. $\mathrm{dst}(\mathcal{A})$ is defined and does not belong to $\mathrm{Support}(\mathcal{B})$,
3. the transitions in $\mathcal{A}$ that are not in $\mathcal{B}$ are all outside of $\mathrm{Support}(\mathcal{B})$ and can be partitioned into the following disjoint sets:
   - a simple path form $\mathrm{dst}(\mathcal{A})$ labeled by $w_B(aw_B)^{d-1}$,
   - a b-thread from $r_0 = \mathrm{src}(\mathcal{A})$ of length λ_0,
   - a b-thread from $r_i = w_{\mathcal{B}}(aw_{\mathcal{B}})^{i-1}$ of length λ_i with a cycle length ℓ_i for $i \in [1, d]$.

For $d \geq 1$, the set of templates $\mathfrak{L}^d$ contains all $(B, \overline{\lambda}, \overline{\ell})$-shaped template $\mathcal{A}$ with $\mathcal{B} \in \mathfrak{F}_n$, $\overline{\lambda} \in [\![\sqrt{n}, 2\sqrt{n}]\!]^{d+1}$ and $\overline{\ell} \in [\![\frac{\sqrt{n}}{2}, \sqrt{n}]\!]^{d+1}$.

▶ **Lemma 40.** *For all $d \geq 0$, the set $\mathfrak{L}_d$ is proper.*

Proof. Let $\mathcal{A} \neq \mathcal{B} \in \mathfrak{L}^d$ be two templates. Assume that there exists a complete template $\mathcal{C}$ such that $\mathcal{C}$ extends both $\mathcal{A}$ and $\mathcal{B}$.

As $\mathfrak{F}$ is proper, there exists a unique automaton $\mathcal{D}$ in $\mathfrak{B}$ such that $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ all extend the same template $\mathcal{D}$.

Let $\mathrm{src} = \mathrm{src}(\mathcal{A}) = \mathrm{src}(\mathcal{B}) = \mathrm{src}(\mathcal{C}) = \mathrm{src}(\mathcal{D})$ and $\mathrm{dst} = \mathrm{dst}(\mathcal{A}) = \mathrm{dst}(\mathcal{B}) = \mathrm{dst}(\mathcal{C})$.

By a direct induction on the length of the words, we can show that for all $u \in b^*$, $\delta_{\mathcal{A}}(\mathrm{src}, u) = \delta_{\mathcal{C}}(\mathrm{src}, u) = \delta_{\mathcal{B}}(\mathrm{src}, u)$. Similarly we can show for all non-empty prefix u of a word in $\{w_{\mathcal{D}}(aw_{\mathcal{D}})^i b^n | i \in [0, d-1]\}$ that $\delta_{\mathcal{A}}(\mathrm{dst}, u) = \delta_{\mathcal{C}}(\mathrm{dst}, u) = \delta_{\mathcal{B}}(\mathrm{dst}, u)$.

With the definition of $\mathfrak{L}^d$, this implies that $\mathcal{A} = \mathcal{B}$. ◀

▶ **Proposition 41.** *For $d \geq 1$, the set of templates $\mathfrak{L}^d$ occurs with visible probability.*

Proof. Let $d \geq 1$. By Lemma 27 and Proposition 39, it is enough to show that $\mathfrak{L}^d$ occurs with visible probability in $\mathfrak{F}$.

Consider an automaton $\mathcal{B} \in \mathfrak{F}_n$. Let $p_{\mathcal{B}}$ be the probability that a complete templates $\mathcal{A} \in \mathrm{CAut}_n$ drawn uniformly at random from the set of complete template extending $\mathcal{B}$ extends $\mathfrak{L}^d$. By Lemma 23, to draw uniformly at random a complete template $\mathcal{A}$ extending $\mathcal{B}$, it is enough to independently draw uniformly at random $\mathrm{dst}(\mathcal{A})$ in $[n]$ and the target of each transition missing in $\mathcal{B}$ in the set $[n] \setminus \mathrm{Closed}(\mathcal{B})$.

We will now describe a process which draws the target of the missing transitions in $\mathcal{B}$ in a particular order but still independently and uniformly at random in $[n] \ \setminus \mathrm{Closed}(\mathcal{B})$.

The process starts with the automaton $\mathcal{B}$ and at each step draws the target of a transition which is missing so far or does nothing at this step. If $\mathcal{C}$ is the automaton built at some step of the process, we will say that we try to draw the transition for a word $u\alpha \in \Sigma^+$ from some state $s \in [n]$ to mean that if $\delta_{\mathcal{C}}(s, u)$ is defined and $\delta_{\mathcal{C}}(s, u\alpha)$ is not, we draw the target of the missing α-labelled transition outgoing from $\delta_{\mathcal{C}}(s, u)$ uniformly at random in $[n] \setminus \mathrm{Closed}(\mathcal{B})$ and otherwise we do nothing.

The process is decomposed into following phases:

- In step 0, we draw uniformly at random $\mathrm{dst}(A)$ in $[n]$. If $\mathrm{dst}(\mathcal{A})$ belongs to $\mathrm{Support}(\mathcal{B})$, the process is said to fail at step 0.
- In step 1, we successively try to draw the transition from $\mathrm{dst}(\mathcal{A})$ for all the non-empty prefixes of the word $w_{\mathcal{B}}(aw_{\mathcal{B}})^{d-1}$ by increasing length. If the target of one of the added transition belongs $\mathrm{Support}(\mathcal{B})$ or is drawn twice during this step, we say that the process fails at step 1.
- In step 2, we successively try to draw the transition from $\mathrm{src}(\mathcal{A})$ for the words $b, bb, \ldots, b^n$. If the b-thread from $\mathrm{src}(\mathcal{A})$ contains a state in $\mathrm{Support}(B)$ or drawn in the previous steps or if its length is not in $[\sqrt{n}, 2\sqrt{n}]$ and its cycle length is not in $[\frac{\sqrt{n}}{2}, \sqrt{n}]$, we say that the process fails at step 2.
- For $i \in [1, d]$, in step $i + 2$, we similarly try to draw the b-thread from $w_{\mathcal{B}}(aw_{\mathcal{B}})^{i-1}$ with the same failure condition.
- Finally we draw the target of all missing transitions in some fix order and take of all the states to be closed.

If we do not take the failure into account, this process generates uniformly at random complete templates extending $\mathcal{B}$. If the process does not fail at any step, the complete template drawn is $(\mathcal{B}, \overline{\lambda}, \overline{\ell})$-shaped with $\overline{\lambda} \in [\![\sqrt{n}, 2\sqrt{n}]\!]^{d+1}$ and $\overline{\ell} \in [\![\frac{\sqrt{n}}{2}, \sqrt{n}]\!]^{d+1}$.

The probability p that the process does not fail at any step is equal to $p_0 \cdot p_1 \cdots p_{d+2}$ where p_0 is the probability that the process does not fail during step 0 and for all $i \in [1, d+2]$, p_i is the probability that the process does not fail at step i knowing that it did not fail during the previous steps.

Using Lemma 18, we will show that for all $i \in [0, d + 2]$ there exists a constant $c_i > 0$ only depending on d such that for n large enough $p_i \geq c_i$. This will imply that there exists a constant $c > 0$, such that for n sufficiently large $p_{\mathcal{B}} \geq p \geq c$ which will conclude the proof.

Recall that by Lemma 37, we have $|w_{\mathcal{B}}| \in \Theta(\ln(n))$, $|\mathrm{Support}(\mathcal{B})| \in O(\sqrt{n})$ and $|\mathrm{Closed}(\mathcal{B})| \in O(\sqrt{n})$.

The probability that the process does not fail in step 0 is $\frac{n - |\mathrm{Support}(\mathcal{B})|}{n} = 1 + O(\frac{1}{\sqrt{n}})$.

For the probability p_1 that the process does not fail in step 1 assuming it did not fail in step 0, we let $t(n) = |w_{\mathcal{B}}(aw_{\mathcal{B}})^{d-1}| \in O_d(\sqrt{n})$. In step 1, we draw the target of at most $t(n)$

transitions in the set $[n] \setminus \mathrm{Closed}(B)$ of size $b(n) \leq n$ with $n - b(n) = |\mathrm{Closed}(B)| \in O(\sqrt{n})$. The process does not fail if we never draw the same state twice nor a state in Support(B) whose size $r(n)$ is in $O(\sqrt{n})$. By Property 1 of Lemma 18, there exists a constant $c_1 > 0$ only depending on d such that $p_1 \geq c_1$ for n large enough.

Let $i \in [2, d+2]$. We consider the probability p_i that the process does not fail at step i knowing it did not fail during the previous steps. Let F_i denote the set of states drawn in the previous phases. As the process is assumed not to have failed in the previous steps, it holds that $|F_i| \leq O_d(\sqrt{n})$. In step i, we draw the b-thread starting from r_{i-2}. For the process not to fail in step i, we need to draw states with replacement in $[n] \setminus \mathrm{Closed}(B)$ of size $b(n) \leq n$ with $n - b(n) \in O(\sqrt{n})$ without drawing a state in $F_i \cup \mathrm{Support}(\mathcal{B})$ of size $r(n) \in O_d(\sqrt{n})$ with the first repetition occurring at time $\lambda \in [\![\sqrt{n}, 2\sqrt{n}]\!]$ and the state drawn twice was first drawn at a time ℓ with $\lambda - \ell \in [\![\frac{\sqrt{n}}{2}, \sqrt{n}]\!]$ (with convention that r_{i-2} was drawn at time 0). By Property 3 of Lemma 18 there exists a constant $c_i > 0$ only depending on d such that $p_i \geq c_i$ for n sufficiently large. ◀

C.6 Restatement of Theorem 9 and its proof

For $d \geq 1$, let $\mathfrak{T}^d$ denote the set of almost deterministic transition structures with initial state $\mathcal{A} = (n, \delta_{\mathcal{A}}, p \xrightarrow{a} q, i_0)$ such that p is reachable from the initial state i_0 and the complete template in CAut_n (i.e., $(n, \delta_{\mathcal{A}}, \mathrm{src}(\mathcal{A}) = p, \mathrm{dst}(\mathcal{A}) = q)$) extends $\mathfrak{L}^d$.

We can now prove Theorem 9 which is slightly reformulated below.

▶ **Theorem 42** (Reformulation of Theorem 9). *Let $d \geq 1$. The set of almost deterministic transition $\mathfrak{T}^d$ occurs with visible probability for the uniform distribution over size-n almost deterministic transition structure. Furthermore, for all $\mathcal{A} = (n, \delta_{\mathcal{A}}, p \xrightarrow{a} q, i_0)$, the state p is reachable from i_0 and there exists a word w of length $\Theta(\log n)$ such that $\delta(p, w(aw)^{d-1}) = \{p_0, \ldots, p_d\}$ is a set of $d+1$ states, and the b-threads starting from the p_i's have lengths λ_i in $[\![\sqrt{n}, 2\sqrt{n}]\!]$ and their cycle length is in $[\![\frac{1}{2}\sqrt{n}, \sqrt{n}]\!]$.*

Moreover for the uniform distribution on $\mathfrak{T}_n$, the cycle lengths are uniform and independent random elements of $[\![\frac{1}{2}\sqrt{n}, \sqrt{n}]\!]$.

Proof. For a complete template $\mathcal{A} \in \mathrm{CAut}_n$, we denote by $\mathrm{SSC}_{\max}(\mathcal{A})$ the terminal strongly connected component with maximal size and, if there are several possible, the one containing the smallest state.

For all $n \geq 1$, we consider the following events that can occur when drawing uniformly at random complete templates $\mathcal{A}$ in CAut_n:

- all states of $\mathcal{A}$ can reach $\mathrm{SSC}_{\max}(\mathcal{A})$ (event R_n),
- $\mathcal{A}$ extends $\mathfrak{L}^d$ (event T_n),
- all cycles outside of $\mathrm{SSC}_{\max}(\mathcal{A})$ have length at most $\ln(\ln(n))$ (event C_n).

In [11], Grusho established that $\lim_{n\to\infty} \mathbb{P}(R_n) = 1$ and in [5, Theorem 2], Cai and Devroye proved that $\lim_{n\to\infty} \mathbb{P}(C_n) = 1$. In Proposition 41, we have shown that there exists a constant $c > 0$, such that for n sufficiently large, $\mathbb{P}(T_n) \geq c$.

Using the union-bound property on the complements, we have:

$$\mathbb{P}(R_n \cap T_n \cap C_n) \geq \mathbb{P}(T_n) - \underbrace{(\mathbb{P}(R_n^c) + \mathbb{P}(C_n^c))}_{\to 0}$$

Hence for n large enough, $\mathbb{P}(R_c \cap T_n \cap C_n) \geq \frac{c}{2}$.

So if we draw uniformly at random a complete template $\mathcal{A} \in \mathrm{CAut}_n$ and an initial state i_0, then with visible probability, we have that $\mathcal{A}$ extends $\mathfrak{L}^d$, all the states can reach $\mathrm{SSC}_{\max}(\mathcal{A})$

and all cycles outside of $\mathrm{SSC}_{\max}(\mathcal{A})$ have length at most $\ln(\ln(n))$. As $\mathcal{A}$ extends $\mathfrak{L}^d$, there is a cycle going through $\mathrm{src}(A)$ with a length in $\Theta(\sqrt{n})$. So for n large enough, $\mathrm{src}(A)$ must belong to $\mathrm{SSC}_{\max}(\mathcal{A})$ and is therefore reachable from the initial state i_0 (or any other state).

It only remains to prove that for the uniform distribution on $\mathfrak{T}_n$, the cycle lengths are uniform and independent random elements of $[\![\frac{1}{2}\sqrt{n}, \sqrt{n}]\!]$.

Let $\mathcal{B} \in \mathfrak{F}_n$, $\overline{\lambda} \in [\![\sqrt{n}, 2\sqrt{n}]\!]^{d+1}$ and $\overline{\ell}, \overline{\ell'} \in [\![\frac{1}{2}\sqrt{n}, \sqrt{n}]\!]^{d+1}$, the set of almost deterministic transition structures with initial state that are $(\mathcal{B}, \overline{\lambda}, \overline{\ell})$-shaped is in one-to-one correspondence with the set of almost deterministic transition structures with initial state that are $(\mathcal{B}, \overline{\lambda}, \overline{\ell'})$-shaped. Indeed as $\mathfrak{L}^d$ is proper, the transformation modifying the cycle length of the different b-threads from ℓ to ℓ' while preserving the thread length λ is a one-to-one.

Let $\overline{\ell} \in [\![\frac{1}{2}\sqrt{n}, \sqrt{n}]\!]^{d+1}$. Consider the probability $p_{\overline{\ell}}$ that an almost deterministic transition structure with initial state $\mathcal{A}$ taken uniformly at random from $\mathfrak{T}^d$ is $(\mathcal{B}, \overline{\lambda}, \overline{\ell})$-shaped for some $\mathcal{B} \in \mathfrak{F}$ and some $\overline{\lambda} \in [\![\sqrt{n}, 2\sqrt{n}]\!]^{d+1}$. As $\mathfrak{L}^d$ is proper, we can use the law of total probabilities:

$$
\begin{aligned}
p_{\ell} &= \sum_{\substack{\mathcal{B}\in\mathfrak{F} \\ \overline{\lambda}\in[\![\frac{1}{2}\sqrt{n},\sqrt{n}]\!]^{d+1}}} \mathbb{P}(\mathcal{A} \in \mathfrak{T}^d \text{ is } (\mathcal{B}, \overline{\lambda}, \overline{\ell})\text{-shaped}) \\
&= \sum_{\substack{\mathcal{B}\in\mathfrak{F} \\ \overline{\lambda}\in[\![\frac{1}{2}\sqrt{n},\sqrt{n}]\!]^{d+1}}} \mathbb{P}(\mathcal{A} \in \mathfrak{T}^d \text{ is } (\mathcal{B}, \overline{\lambda}, \overline{\ell'})\text{-shaped}) \\
&= p_{\ell'}
\end{aligned}
$$

◀

C.7 Proofs of the auxiliary lemmas and propositions

In our proof of Theorem 9, we have established the proof of all the auxiliary lemmas and propositions presented in the article. For completeness, we will briefly describe where these lemmas and propositions have been established.

▶ **Lemma 43** (Restatement of Lemma 5)**.** *Let p be a random state of a random n-state deterministic transition structure. With visible probability, the $\sqrt{n}$-backward tree from p exists, has depth $\tau \in \Theta(\log n)$, contains between $\sqrt{n}$ and $3\sqrt{n}$ extremal leaves, i.e. states in $R_\tau(p)$, and has a total number of nodes in $\Theta(\sqrt{n})$.*

Proof. This is a direct consequence of the fact that $\mathfrak{B}$ occurs with visible probability (cf. Lemma 28 and Proposition 30). ◀

▶ **Lemma 44** (Restatement of Lemma 6)**.** *For the uniform distribution on size-n transition structures having T_p as $\sqrt{n}$-backward tree from p, with visible probability the breadth-first traversal starting at $r := \delta_a(p)$ hits an extremal leaf of T_p before it discovers the same state twice, and it does this in at most $\sqrt{n}$ steps.*

Proof. The proof of this lemma is almost identical to proof of Proposition 39 which shows that $\mathfrak{F}$ occurs with visible probability in $\mathfrak{B}$. ◀

▶ **Proposition 45** (Restatement of Proposition 7)**.** *With visible probability, an n-state transition structure taken uniformly at random is p-compatible, where p is also taken uniformly at random and independently in $[n]$. In this case, the p-substructure is unique, has $O(\sqrt{n})$ states, and contains a circuit around t labelled aw, where w is uniquely determined using the transitions of the p-structure only and we have $|w| \in \Theta(\log n)$.*

Proof. This is a direct consequence of the fact that $\mathfrak{F}$ is proper and occurs with visible probability (cf. Proposition 39 and Lemma 38). ◀

▶ **Lemma 46** (Restatement of 8). *Let $d \geq 1$. Let X_p be a p-substructure of size-n transition structures. For the uniform distribution on size-n transition structures that are p-compatible and that have X_p as p-substructure, if we add a random transition $p \xrightarrow{a} q$ by choosing q uniformly at random and independently in $[n]$, then with visible probability (i) the states discovered while following the paths labeled by $w(aw)^{d-1}$ are all different and do not belong to X_p (ii) the b-threads starting at the p_i's, where $p_0 = p$ and $p_i = \delta(s, a(aw)^{d-1})$, have length between $\sqrt{n}$ and $2\sqrt{n}$, are pairwise disjoint and do not intersect X_p.*

Proof. This is essentially proved when establishing that $\mathfrak{L}_d$ occurs with visible probability in $\mathfrak{F}$. ◀

D Proofs of Section 5

D.1 Proof of Lemma 11

Proof. Let p and q be two different states of $\mathcal{C}$. Let x and y be the associated words of $\mathcal{C}$ starting at p and q, respectively. Let k be the smallest integer such that $\delta_{\alpha^k}(p) = q$ and let u be the prefix of length k of x, and v be the associated suffix: $x = uv$. Then $y = vu$. Assume by contradiction that p and q are equivalent. This implies that $x = y$, as the automata obtained by placing the initial states either on p or q recognize the same elements of $\{\alpha\}^*$. Hence $uv = vu$, and therefore u and v are the power of the same word by a classical result on primitive words [15, Prop. 1.3.2 page 8]. This is in contradiction with the fact that $\mathcal{C}$ is primitive. ◀

D.2 Proof of Lemma 12

Proof. Let $w = w^{(1)} \odot w^{(2)}$. Assume by contradiction that there exists some word z and some $k \geq 2$ such that $w = z^k$. Let p be a prime number that divides k, we have $w = (z^{k/p})^p$. This yields that p divides $\ell = \mathrm{lcm}(\ell_1, \ell_2) = \ell_1 \times \ell_2$ and that for every non-negative integer i, $w_i = w_{i+\ell/p}$ (indices taken modulo ℓ). Obviously, p divides either ℓ_1 or ℓ_2, but not both. By symmetry, assume that it divides ℓ_1: $\ell_1 = pr$ and $\ell/p = r\ell_2$.

Since $w^{(2)}$ has length at least 2 and is primitive, there exists an index $i_0 \in \{0, \dots, \ell_2 - 1\}$ such that $w^{(2)}_{i_0} = 0$. Define $i_j = i_0 + j\ell_2$, for any $j \geq 0$. As indices in $w^{(2)}$ are taken modulo ℓ_2, we have $w^{(2)}_{i_j} = 0$ for all $j \geq 0$. Therefore, $w^{(1)}_{i_j} = 1$ if and only if $w_{i_j} = 1$. Thus $w^{(1)}_{i_j} = w^{(1)}_{i_j + r\ell_2}$ for all $j \geq 0$. Moreover, $r\ell_2$ is not a multiple of ℓ_1: let $\alpha \geq 1$ be the largest integer such that p^α divides ℓ_1, then p^α does not divide $r\ell_2$. Let $s := r\ell_2 \mod \ell_1$, we just established that $s \neq 0$, so we have the non-trivial relation $w^{(1)}_{i_j} = w^{(1)}_{i_j+s}$ for all $j \geq 0$. Recall that $i_j = i_0 + j\ell_2$. As ℓ_1 and ℓ_2 are coprime, the i_j take all values modulo ℓ_1 when j ranges from 0 to $(\ell_1 - 1)$ and i_j stays between 0 and $\mathrm{lcm}(\ell_1, \ell_2)$ doing so. Hence, for all $k \in \{0, \ell_1 - 1\}$, $w^{(1)}_k = w^{(1)}_{k+s}$, for some $s > 0$. This is a contradiction with the fact that $w^{(1)}$ is primitive, concluding the proof. ◀

D.3 Proof of Corollary 15

Proof. Let X be the event that $w = 0^\ell$ or $w = 1^\ell$. We have $\mathbb{P}(X) = f_n^\ell + (1 - f_n)^\ell$. Since changing the 0's in 1's and the 1's in 0's preserves primitivity, we can assume by symmetry that $f_n \leq \frac{1}{2}$. By hypothesis, there exists some constant $\beta > 0$ such that $\frac{\beta}{\sqrt{n}} \leq f_n$ and

$\frac{\beta}{\sqrt{n}} \leq 1 - f_n$ hence, as $f_n \leq \frac{1}{2}$, we have

$$f_n^\ell \leq \frac{1}{2^{\alpha\sqrt{n}}} \text{ and } (1-f_n)^\ell \leq \left(1 - \frac{\beta}{\sqrt{n}}\right)^{\alpha\sqrt{n}}.$$

Since $(1-\frac{\beta}{\sqrt{n}})^{\alpha\sqrt{n}} = e^{-\alpha\beta} + O(\frac{1}{\sqrt{n}})$, there exists some constant $\delta < 1$ such that $\mathbb{P}(X) \leq \delta$, for n sufficiently large.

Let W be a random word under our distribution. For any $w \in \{0,1\}^\ell$, the conditional probability that W values w given that $W \notin \{0^\ell, 1^\ell\}$ is

$$\mathbb{P}(W = w \mid \overline{X}) = \begin{cases} 0 & \text{if } w = 0^\ell \text{ or } w = 1^\ell, \\ \frac{\mathbb{P}(w)}{1-\mathbb{P}(X)} & \text{otherwise.} \end{cases}$$

Hence we are in the settings of Lemma 14, and the probability that w is not primitive, given that $w \notin \{0^\ell, 1^\ell\}$ is at most $\frac{2}{\ell}$. This concludes the proof since:

$$\mathbb{P}(w \text{ not primitive}) = \mathbb{P}(X) + \mathbb{P}(w \text{ not primitive} \mid \overline{X})\mathbb{P}(\overline{X}) \leq \delta + \frac{2}{\ell}.$$

This concludes the proof. ◀

D.4 Proof of Corollary 16

We first state Tóth's theorem:

▶ **Theorem 47** (Tóth [18]). *For any $d \geq 2$, there exists some constant $A_d > 0$ such that d integers taken uniformly at random and independently in $[n]$ are pairwise coprime with probability $A_d + O(\frac{\log^{d-1} n}{n})$.*

Now we can prove Corollary 16:

Proof. By Theorem 47, there are $N_n := A_{d+1}\lfloor\frac{1}{2}\sqrt{n}\rfloor^{d+1}(1+o(1))$ tuples of $[\![1, \frac{1}{2}\sqrt{n}]\!]$ whose coordinates are pairwise coprimes and $M_n := A_{d+1}\lfloor\sqrt{n}\rfloor^{d+1}(1+o(1))$ tuples of $[\![1, \sqrt{n}]\!]$ whose coordinates are pairwise coprimes. We conclude the proof by remarking that $M_n - N_n$ is asymptotically equivalent to $A_{d+1}(1 - \frac{1}{2^{d+1}})n^{\frac{d+1}{2}}$. ◀