



Number of Prefixes in Trace Monoids: Clique Polynomials and Dependency Graphs

Cyril Banderier¹✉  and Massimiliano Goldwurm² 

¹ Laboratoire d'Informatique de Paris-Nord (UMR CNRS 7030),
Université Paris-Nord, Villetaneuse, France
cb@lipn.fr

² Dipartimento di Matematica “Federigo Enriques”,
Università degli Studi di Milano, Milan, Italy

Abstract. We present some asymptotic properties on the average number of prefixes in trace languages. Such languages are characterized by an alphabet and a set of commutation rules, also called complement alphabet, which can be encoded by an independency graph or by its complement, called dependency graph. One key technical result, which has its own interest, concerns general properties of graphs and states that “if an undirected graph admits a transitive orientation, then the multiplicity of the root of minimum modulus of its clique polynomial is smaller or equal to the number of connected components of its complement graph”. As a consequence, under the same hypothesis of transitive orientation of the independency graph, one obtains the relation $E[T_n] = O(E[W_n])$, where the random variables T_n and W_n represent the number of prefixes in traces of length n under two different fundamental probabilistic models:

- the uniform distribution among traces of length n (for T_n),
- the uniform distribution among words of length n (for W_n).

These two quantities are related to the time complexity of algorithms for solving classical membership problems on trace languages.

Keywords: Trace monoids · Clique polynomials · Möbius functions · Automata theory · Analytic combinatorics · Patterns in words

1 Introduction

In computer science, trace monoids have been introduced by Mazurkiewicz [22] as a model of concurrent events, describing which action can permute or not with another action (we give a formal definition of traces and trace monoids in Sect. 2, see also [14] for a treatise on the subject). In combinatorics, they are related to the fundamental studies of the “monoïde partiellement commutatif” introduced by Cartier and Foata in [10], and to its convenient geometrical view as heap of pieces proposed by Viennot in [25].

Several classical problems in language theory (recognition of rational and context-free trace languages, determination of the number of representative words of a given trace, computing the finite state automaton recognizing these words) can be solved by algorithms that work in time and space proportional to (or strictly depending on) the number of prefixes of the input trace [3, 6–8, 15, 23]. This is due to the fact that prefixes represent the possible decompositions of a trace in two parts and hence they are natural indexes for computations on traces.

This motivates the analysis of the number of prefixes of a trace of given length both in the worst and in the average case. In the average case analysis, two natural sequences of random variables play a key role:

- $\{T_n\}_{n \in \mathbb{N}}$, the number of prefixes of traces of length n generated at random under the equidistribution of traces of given size;
- $\{W_n\}_{n \in \mathbb{N}}$, the number of prefixes of traces of length n generated at random under the equidistribution of representative words of given size.

For some families of trace monoids, the asymptotic average, variance, and limit distributions of $\{T_n\}$ and $\{W_n\}$ are known [6, 7, 19–21]. It is interesting that they rely on the structural properties of an underlying graph (the independency graph, defined in Sect. 2). For example, it is known that, for every trace monoid \mathcal{M} , the maximum number of prefixes of a trace of length n is of the order $\Theta(n^\alpha)$, where α is the size of the largest clique in the concurrent alphabet defining \mathcal{M} [8]. We summarize further such results in Sect. 3. In analytic combinatorics (see [17] for an introduction to this field), it remains a nice challenge to get a more universal description of the possible asymptotics of T_n and W_n .

In this work we prove that, if the concurrent alphabet (Σ, \mathcal{C}) admits a transitive orientation, then

$$E[T_n] = O(E[W_n]).$$

This is obtained by showing a general property of undirected graphs, which in our context is applied to the concurrent alphabet (Σ, \mathcal{C}) and its complement (Σ, \mathcal{C}^c) . Such a property states that, for any undirected graph \mathcal{G} admitting a transitive orientation of its edges, the number of connected components of its complement is greater or equal to the multiplicity of the root of smallest modulus in the clique polynomial of \mathcal{G} . The interest for the present discussion mainly relies on the use of finite state automata and on classical tools of formal languages to study properties of integer random variables in particular the asymptotic behaviour of their moments.

The paper is organized as follows: in Sect. 2 we recall the basic definitions on trace monoids; in Sect. 3 we summarize some asymptotic results on the random variables T_n and W_n ; in Sects. 4 and 5, we present our main results on cross-sections of trace monoids, clique polynomials, and a new bound relating the asymptotic behaviour of T_n and W_n ; we then conclude with possible future extensions of our work.

2 Notation and Preliminary Notions

For the reader not already familiar with the terminology of trace languages, we present in this section the key notions used in this article (see e.g. [14] for more details on all these notions).

Given a finite alphabet Σ , as usual Σ^* denotes the free monoid of all words over Σ , ε is the empty word and $|w|$ is the length of a word w for every $w \in \Sigma^*$. We recall that, for any $w \in \Sigma^*$, a prefix of w is a word $u \in \Sigma^*$ such that $w = uv$, for some $v \in \Sigma^*$. Also, for any finite set \mathcal{S} , we denote by $\#\mathcal{S}$ the cardinality of \mathcal{S} .

A **concurrent alphabet** is then a pair (Σ, \mathcal{C}) , where $\mathcal{C} \subseteq \Sigma \times \Sigma$ is a symmetric and irreflexive relation over Σ . Such a pair can alternatively be defined by an undirected graph, which we call **independency graph**, where Σ is the set of nodes and $\{\{a, b\} \mid (a, b) \in \mathcal{C}\}$ is the set of edges. Its complement (Σ, \mathcal{C}^c) is called **dependency graph**. As the notions of concurrent alphabet and independency graph are equivalent, in the sequel we indifferently refer to either of them. Informally, a concurrent alphabet lists the pairs of letters which can commute.

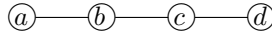
The **trace monoid** generated by a concurrent alphabet (Σ, \mathcal{C}) is defined as the quotient monoid $\Sigma^* / \equiv_{\mathcal{C}}$, where $\equiv_{\mathcal{C}}$ is the smallest congruence extending the equations $\{ab = ba : (a, b) \in \mathcal{C}\}$, and is denoted by $\mathcal{M}(\Sigma, \mathcal{C})$ or simply by \mathcal{M} . Its elements are called **traces** and its subsets are named **trace languages**. In other words, a trace is an equivalence class of words with respect to the relation $\equiv_{\mathcal{C}}$ given by the reflexive and transitive closure of the binary relation $\sim_{\mathcal{C}}$ over Σ^* such that $uabv \sim_{\mathcal{C}} ubav$ for every $(a, b) \in \mathcal{C}$ and every $u, v \in \Sigma^*$. For any $w \in \Sigma^*$, we denote by $[w]$ the trace represented by w ; in particular $[\varepsilon]$ is the empty trace, i.e. the unit of \mathcal{M} . Note that the product of two traces $r, s \in \mathcal{M}$, where $r = [x]$ and $s = [y]$, is the trace $t = [xy]$, which does not depend on the **representative words** $x, y \in \Sigma^*$ and we denote the product by $t = s \cdot r$. The length of a trace $t \in \mathcal{M}$, denoted by $|t|$, is the length of any representative word. For any $n \in \mathbb{N}$, let $\mathcal{M}_n := \{t \in \mathcal{M} : |t| = n\}$ and $m_n := \#\mathcal{M}_n$.

Note that if $\mathcal{C} = \emptyset$ then \mathcal{M} reduces to Σ^* , while if $\mathcal{C} = \{(a, b) \in \Sigma \times \Sigma \mid a \neq b\}$ then \mathcal{M} is the commutative monoid of all monomials with letters in Σ .

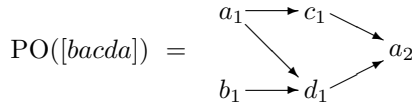
Any trace $t \in \mathcal{M}$ can be represented by a **partial order** over the multiset of letters of t , denoted by $\text{PO}(t)$. It works as follows: first, consider a word w satisfying $t = [w]$. Then, for any pair of letters (a, b) of w , let a_i be the i -th occurrence of the letter a and b_j the j -th occurrence of the letter b . The partial order is then defined as $a_i < b_j$ whenever a_i precedes b_j in all representative words of $[w]$. (See Example 1 hereafter.)

A **prefix** of a trace $t \in \mathcal{M}$ is a trace p such that $t = p \cdot s$ for some $s \in \mathcal{M}$. Clearly, any prefix of t is a trace $p = [u]$ where u is a prefix of a representative of t . It is easy to see that if p is a prefix of t then the $\text{PO}(u)$ is an order ideal of $\text{PO}(t)$ and can be represented by the corresponding antichain. We recall that an **antichain** of a partial order set (\mathcal{S}, \leq) is a subset $A \subseteq \mathcal{S}$ such that $a \leq b$ does not hold for any pair of distinct elements $a, b \in A$, while an order ideal in (\mathcal{S}, \leq) is a subset $\{a \in \mathcal{S} \mid \exists b \in A \text{ such that } a \leq b\}$ for some antichain A of (\mathcal{S}, \leq) . For every $t \in \mathcal{M}$, we denote by $\text{Pref}(t)$ the set of all prefixes of t .

Example 1. Let \mathcal{M} be the trace monoid characterized by the following independence graph:



That is, one has $ab = ba, bc = cb, cd = dc$. Then, the trace $[bacda]$ (i.e., the equivalence class of the word $bacda$) is the set of words $\{bacda, badca, abdca, abceda, acbda\}$. The corresponding partially ordered set is given by the following diagram



where an arrow from x_i to y_j means that x_i always precedes y_j and where we omitted the arrows implied by transitivity. The set of prefixes is given by

$$\text{Pref}([bacda]) = \{[\varepsilon], [a], [b], [ab], [ac], [abc], [abd], [abcd], [abcda]\}.$$

In this set, we now overline the letters belonging to the antichain of each prefix: $\{[\varepsilon], [\bar{a}], [\bar{b}], [\bar{a}\bar{b}], [\bar{a}\bar{c}], [\bar{a}\bar{b}\bar{c}], [\bar{a}\bar{b}\bar{d}], [\bar{a}\bar{b}\bar{c}\bar{d}], [\bar{a}\bar{b}\bar{c}\bar{d}\bar{a}]\}$. ■

Recognizable, rational and context-free trace languages are well defined by means of linearization and closure operations over traditional string languages; their properties and in particular the complexity of their membership problems are widely studied in the literature (see for instance [8, 14, 15, 23]).

For any alphabet Σ and trace monoid \mathcal{M} , we denote by $\mathbb{Z}\langle\langle\Sigma\rangle\rangle$ the set of formal series on words (they are thus series in noncommutative variables) and by $\mathbb{Z}\langle\langle\mathcal{M}\rangle\rangle$ the set of formal series on traces (they are thus series in partially commutative variables), and $\mathbb{Z}\llbracket z \rrbracket$ stands for ring of classical power series in the variable z . These three distinct rings (with the operations of sum and Cauchy product, see [5, 14, 24]) will be used in Sects. 4 and 5.

3 Asymptotic Results for the Number of Prefixes

Several algorithms are presented in the literature for the recognition of rational and context-free trace languages, or for other problems like computing the number of representative words of a trace, that take a trace t as input and then carry out some operations on all prefixes of t [3, 6–9, 15, 23]. Thus, their time and space complexity strictly depend on the number of prefixes of t and in many cases they work just in time $\Theta(\#\text{Pref}(t))$. Now, it follows from [8] that

$$\max\{\#\text{Pref}(t) : t \in \mathcal{M}_n\} = \Theta(n^\alpha), \tag{1}$$

where α is the size of the largest clique in the independency graph of \mathcal{M} . It is thus essential to get a more refined analysis of the asymptotic behaviour of $\#\text{Pref}(t)$ under natural distribution models in order to obtain a better understanding of the average complexity of all these algorithms.

In this section, we recall the main results on the number of prefixes of a random trace, under two different probabilistic models.

3.1 Probabilistic Analysis on Equiprobable Words

A main goal of the present contribution is to compare the random variables T_n and W_n , defined by

$$T_n = \# \text{Pref}(t) \quad \text{and} \quad W_n = \# \text{Pref}([w]), \tag{2}$$

where t is uniformly distributed over \mathcal{M}_n , while w is uniformly distributed over Σ^n . Clearly the properties of T_n and W_n immediately yield results on time complexity of the algorithms described in [3,6,7] assuming, respectively, equiprobable input traces of length n and equiprobable representative words of length n . Since every trace of length n has at least $n + 1$ prefixes, a first crude asymptotic bound is

$$n + 1 \leq T_n \leq dn^\alpha, \quad n + 1 \leq W_n \leq dn^\alpha \quad (\forall n \in \mathbb{N}),$$

for a suitable constant $d > 0$, where α is defined as in (1). More precise results on the moments of W_n are studied in [6,7,20]:

$$E[W_n^j] = \Theta(n^{jk}) \quad \forall j \in \mathbb{N}, \tag{3}$$

where k is the number of connected components of the dependency graph of \mathcal{M} . This relation is obtained by constructing suitable bijections between each moment of W_n and the set of words of length n in a regular language [6]. These bijections also allow proving a first order cancellation of the variance, i.e. $\text{var}(W_n) = O(n^{2k-1})$ [20]. Further, when the dependency graph is transitive, this leads to two different limit laws, either chi-squared or Gaussian, according whether all the connected components of (Σ, C^c) have the same size or not [19].

3.2 Probabilistic Analysis on Equiprobable Traces

Now, in order to analyse T_n (the number of prefixes of a random trace of size n), it is useful to introduce the **generating function** of the trace monoid \mathcal{M} :

$$M(z) := \sum_{n \in \mathbb{N}} m_n z^n, \quad \text{with } m_n := \#\mathcal{M}_n = \#\{t \in \mathcal{M} : |t| = n\}.$$

The **Möbius function** of \mathcal{M} is defined as $\mu_{\mathcal{M}} := \sum_{t \in \mathcal{M}} \mu_{\mathcal{M}}(t) t$, where

$$\mu_{\mathcal{M}}(t) = \begin{cases} 1 & \text{if } t = [\varepsilon], \\ (-1)^n & \text{if } t = [a_1 a_2 \cdots a_n], \\ & \text{where all } a_i \in \Sigma \text{ are distinct and } (a_i, a_j) \in \mathcal{C} \text{ for any } i \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

It is in fact a polynomial belonging to $\mathbb{Z}\langle\langle \mathcal{M} \rangle\rangle$. As established by Cartier and Foata in [10], an important property of $\mu_{\mathcal{M}}$ is that

$$\xi_{\mathcal{M}} \cdot \mu_{\mathcal{M}} = \mu_{\mathcal{M}} \cdot \xi_{\mathcal{M}} = 1, \tag{4}$$

where $\xi_{\mathcal{M}} = \sum_{t \in \mathcal{M}} t$ is the characteristic series of \mathcal{M} . Here, $\xi_{\mathcal{M}}$ can be seen as a partially commutative analogue of $M(z)$.

Now, let $p_{\mathcal{M}} \in \mathbb{Z}[z]$ be the commutative analogue of $\mu_{\mathcal{M}}$. It then follows that

$$p_{\mathcal{M}}(z) = 1 - c_1z + c_2z^2 - \dots + (-1)^\alpha c_\alpha z^\alpha, \tag{5}$$

where c_i is the number of cliques of size i in the independency graph of \mathcal{M} . For this reason, we call $p_{\mathcal{M}}$ the **clique polynomial** of the independency graph (Σ, C) . Its properties are studied in several papers (see for instance [18, 21]). In particular, the commutative analogue of Eq. (4) is then

$$M(z) \cdot p_{\mathcal{M}}(z) = p_{\mathcal{M}}(z) \cdot M(z) = 1. \tag{6}$$

This entails that $M(z) = (p_{\mathcal{M}}(z))^{-1}$, a fundamental identity which can also be derived by an inclusion-exclusion principle.

As it is known from [21] that $p_{\mathcal{M}}$ has a unique root ρ of smallest modulus (and clearly $\rho > 0$ via Pringsheim’s theorem, see [17]), one gets $m_n = \#\mathcal{M}_n = c\rho^{-n}n^{\ell-1} + O(\rho^{-n}n^{\ell-2})$, where $c > 0$ is a constant and ℓ is the multiplicity of ρ in $p_{\mathcal{M}}(z)$. We observe that the existence of a unique root of smallest modulus for $p_{\mathcal{M}}(z)$ is not a consequence of the strict monotonicity of the sequence $\{m_n\}$. Indeed, if one considers $M(z) = \frac{1}{(1-z^3)(1-z)^2}$, one has $m_{n+3} = ((n + 5)m_n + 2m_{n+1} + 2m_{n+2})/(n + 3)$ so the sequence $\{m_n\}$ is strictly increasing; however, the polynomial $(1-z^3)(1-z)^2$ has 3 distinct roots of smallest modulus. Therefore, such a $M(z)$ *cannot* be the generating function of a trace monoid.

In our context, clique polynomials are particularly relevant as they are related to the average value of the number of prefixes of traces [7, 21]. In fact, for any trace monoid \mathcal{M} , we have $E[T_n] = \frac{P_n}{m_n}$, where $P_n = \sum_{t \in \mathcal{M}_n} \#\text{Pref}(t)$. Since $\xi_{\mathcal{M}}^2 = \sum_{t \in \mathcal{M}} \#\text{Pref}(t)t$, from (4) and (6) its commutative analogue becomes $\sum_n P_n z^n = p_{\mathcal{M}}(z)^{-2}$ and hence $P_n = \Theta(\rho^{-n}n^{2\ell-1})$, which proves

$$E[T_n] = \Theta(n^\ell), \tag{7}$$

where ℓ is the multiplicity of the smallest root of $p_{\mathcal{M}}(z)$.

4 Cross-Sections of Trace Monoids

Cross-sections are standard tools to study the properties of trace monoids by lifting the analysis at the level of free monoids. Intuitively, a cross-section of a trace monoid \mathcal{M} is a language \mathcal{L} having exactly one representative string for each trace in \mathcal{M} . Thus, the generating function of \mathcal{L} coincides with $M(z)$ and hence it satisfies equality (6). As a consequence, by choosing an appropriate regular cross-section \mathcal{L} , one can use the property of a finite state automaton recognizing \mathcal{L} to study the singularities of $M(z)$, i.e. the roots of $p_{\mathcal{M}}(z)$.

Formally, a **cross-section** of a trace monoid \mathcal{M} over a concurrent alphabet (Σ, C) is a language $\mathcal{L} \subseteq \Sigma^*$ such that

- for each trace $t \in \mathcal{M}$, there exists a word $w \in \mathcal{L}$ such that $t = [w]$,
- for each pair of words $x, y \in \mathcal{L}$, if $[x] = [y]$ then $x = y$.

Among all cross-sections of \mathcal{M} , it is convenient to consider a canonical one. A natural one is based on a normal form using the lexicographic order [1]. Alternatively, one can see it as based on the orientations of edges in the independency graph of \mathcal{M} , as used in [12, 13] to study properties of Möbius functions in trace monoids. It works as follows. Let \leq be any total order on the alphabet Σ and let \leq^* be the lexicographic linear order induced by \leq over Σ^* . We denote by $<_c$ the binary relation over Σ such that $a <_c b$ if $(a, b) \in \mathcal{C}$ and $a \leq b$. Thus, $<_c$ is an orientation of the independency graph of \mathcal{M} . We now consider the following cross-section of \mathcal{M} : the language \mathcal{L}_{\leq} of all minimal lexicographic representatives of traces in \mathcal{M} , i.e. $\mathcal{L}_{\leq} = \{w \in \Sigma^* \mid w \leq^* y \text{ for every } y \in [w]\}$. Moreover, \mathcal{L}_{\leq} is regular, as it satisfies the equality

$$\mathcal{L}_{\leq} = \Sigma^* \setminus \bigcup_{\substack{(a,b) \in \mathcal{C} \\ a <_c b}} \Sigma^* b \mathcal{C}_a^* a \Sigma^*, \tag{8}$$

where $\mathcal{C}_a := \{c \in \Sigma \mid (a, c) \in \mathcal{C}\}$ is the set of letters allowed to commute with a . Thus, \mathcal{L}_{\leq} is the set of all words in Σ^* that do not contain any factor of the form bva where $a <_c b$ and $v \in \mathcal{C}_a^*$. Then, for any $w \in \Sigma^*$, in order to verify whether $w \in \mathcal{L}_{\leq}$, one can read the letters of w in their order, updating at each step the family of letters $a \in \Sigma$ forming a “forbidden” factor of the form bva , with $a <_c b$, $v \in \mathcal{C}_a^*$. If one of these letters is met then w is rejected, otherwise it is accepted.

To formalize the definition, for each $b \in \Sigma$, the predecessors of b are $\text{Pred}(b) = \{a \in \Sigma \mid a <_c b\}$. Define the finite state automaton \mathcal{A} as the 4-tuple $(2^\Sigma, \emptyset, \delta, F)$, where the set of states is 2^Σ , i.e. the power set of Σ , the initial state is the empty set \emptyset , $F = \{\mathcal{S} \in 2^\Sigma \mid \mathcal{S} \neq \Sigma\}$ is the family of final states and the transition function $\delta : (2^\Sigma \times \Sigma) \rightarrow 2^\Sigma$ is given by

$$\delta(\mathcal{S}, b) = \begin{cases} \Sigma & \text{if } b \in \mathcal{S} \\ \text{Pred}(b) \cup (\mathcal{S} \cap \mathcal{C}_b) & \text{otherwise} \end{cases} \quad (\forall \mathcal{S} \subseteq \Sigma, \forall b \in \Sigma).$$

Note that, during a computation, the current state \mathcal{S} represents the set of forbidden letters. At the beginning, all input letters are allowed, as \emptyset is the initial state, while Σ is a trap state, where all letters are forbidden. In a general step, if $\mathcal{S} \subseteq \Sigma$ is the current state and $b \notin \mathcal{S}$ is an input letter, the new set of forbidden letters must be obtained from $\mathcal{S} \cup \text{Pred}(b)$ by removing those elements that do not commute with b . This justifies the above definition of δ and it is clear that \mathcal{A} recognizes \mathcal{L}_{\leq} .

Moreover, the state set of the above automaton can be reduced to the states $\mathcal{S} \subsetneq \Sigma$ reachable from \emptyset . Setting

$$Q = \{\mathcal{S} \subseteq \Sigma \mid \mathcal{S} \neq \Sigma, \exists w \in \Sigma^* : \delta(\emptyset, w) = \mathcal{S}\},$$

the entries of the transition matrix \tilde{A} of the automaton \mathcal{A} are given by:

$$\tilde{A}_{\mathcal{S}, \mathcal{S}'} = \sum_{b \in \Sigma : \delta(\mathcal{S}, b) = \mathcal{S}'} b \quad (\forall \mathcal{S}, \mathcal{S}' \in Q).$$

The commutative analogue in $\mathbb{N}[[z]]$ of this transition matrix has therefore all its entries which are monomials of degree one in z . Factorizing by z , this commutative analogue can thus be written zA , for a matrix A we call the *adjacency matrix* of \mathcal{A} . Note that A strictly depends on both the concurrent alphabet (Σ, \mathcal{C}) and the total order \leq over Σ .

As a consequence, since \mathcal{A} recognizes a cross-section of \mathcal{M} , denoting by π and η , respectively, the characteristic (column) vectors of \emptyset and Q , the generating function $M(z)$ is given by

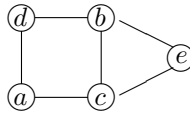
$$M(z) = \sum_{n=0}^{+\infty} \pi' A^n \eta z^n = \pi' (I - zA)^{-1} \eta, \tag{9}$$

where I is the identity matrix of size $\#Q \times \#Q$ and π' is the transposed of π . This identity, together with relation (6) proves the following proposition.

Proposition 1 (Factorisation property). *For any trace monoid \mathcal{M} with a concurrent alphabet (Σ, \mathcal{C}) , let \leq be a total order on Σ , let A be the adjacency matrix of the automaton \mathcal{A} recognizing the cross-section \mathcal{L}_{\leq} of \mathcal{M} , and assume I , π and η defined as in (9). Then, $M(z)$ and $p_{\mathcal{M}}(z)$ satisfy the identities*

$$M(z) = \pi' (I - zA)^{-1} \eta = \frac{\pi' \text{adj}(I - zA) \eta}{\det(I - zA)}, \quad p_{\mathcal{M}}(z) = \frac{\det(I - zA)}{\pi' \text{adj}(I - zA) \eta}. \tag{10}$$

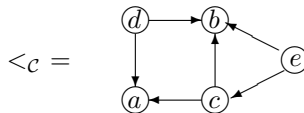
Example 2. Consider the concurrent alphabet (Σ, \mathcal{C}) defined by the graph



Then, the clique polynomial and the generating function of \mathcal{M} are given by

$$p_{\mathcal{M}}(z) = 1 - 5z + 6z^2 - z^3, \quad M(z) = \sum_{n=0}^{+\infty} m_n z^n = \frac{1}{1 - 5z + 6z^2 - z^3}.$$

The standard ordering (a, b, c, d, e) on Σ induces the following (non-transitive) orientation $<_{\mathcal{C}}$ over the independency graph



Thus the predecessors of each letter are given by $\text{Pred}(a) = \text{Pred}(b) = \emptyset$, $\text{Pred}(c) = \text{Pred}(d) = \{a, b\}$, $\text{Pred}(e) = \{b, c\}$ and the transition matrix of \mathcal{A} is defined by the following table, where rows and columns are labelled by the states of \mathcal{A} :

$$\tilde{A} = \begin{array}{c|cccc} & \emptyset & \{a, b\} & \{b, c\} & \{c\} \\ \hline \emptyset & a + b & c + d & e & 0 \\ \{a, b\} & 0 & c + d & e & 0 \\ \{b, c\} & 0 & d & e & a \\ \{c\} & 0 & d & e & a + b \end{array}$$

From that $I - zA$ is easily computed (where A is the adjacency matrix of \mathcal{A}):

$$I - zA = \begin{bmatrix} 1 - 2z & -2z & -z & 0 \\ 0 & 1 - 2z & -z & 0 \\ 0 & -z & 1 - z & -z \\ 0 & -z & -z & 1 - 2z \end{bmatrix}$$

and, accordingly, $\det(I - zA) = 1 - 7z + 16z^2 - 13z^3 + 2z^4 = (1 - 2z)p_{\mathcal{M}}(z)$. ■

Proposition 2. *For any trace monoid \mathcal{M} over a concurrent alphabet (Σ, C) and any total order \leq on Σ , all roots of the clique polynomial $p_{\mathcal{M}}(z)$ are reciprocals of eigenvalues of the corresponding adjacency matrix A . More precisely, the clique polynomial of any independency graph (Σ, C) is of the form*

$$p_{\mathcal{M}}(z) = \prod_{i=1}^{\alpha} (1 - x_i z)$$

where α is the size of the maximum clique in (Σ, C) and all x_i 's are eigenvalues of a adjacency matrix A .

Proof (sketch). The result follows from Proposition 1 by refining equalities (10) and recalling that all roots of clique polynomials are different from 0. □

We observe that the reverse property does not hold in general, i.e. it may occur that an eigenvalue of A is not the reciprocal of a root of $p_{\mathcal{M}}(z)$. However, as shown in the following section, such a reverse sentence is true whenever the graph (Σ, C) admits a transitive orientation.

5 Concurrent Alphabets with Transitive Orientation

Now let us consider a trace monoid \mathcal{M} such that its independency graph (Σ, C) admits a transitive orientation. Then, we may fix a total order \leq on Σ such that $<_C$ is transitive. In this case, the definition of cross-section \mathcal{L}_{\leq} and of the automaton \mathcal{A} can be simplified, since the set of “forbidden” factors of the form bwa , with $a <_C b$ and $w \in \mathcal{C}_a^*$, can be reduced to the simple set of words $S = \{\tau\sigma \in \Sigma^2 \mid \sigma <_C \tau\}$. To prove this property, consider a forbidden factor of the above form bwa , with $a <_C b$ and $w \in \mathcal{C}_a^*$; thus any symbol c occurring in w must verify $(a, c) \in C$. As a consequence, either $a <_C c$ or $c <_C a$: in the first case

ca belongs to \mathcal{S} while, in the second case, by transitivity of $<_c$ we have $c <_c b$ and hence bc is in \mathcal{S} .

Thus, identity (8) can be simplified as $\mathcal{L}_{\leq} = \Sigma^* \setminus \bigcup_{a <_c b} \Sigma^* ba \Sigma^*$. Moreover, the state set of the automaton \mathcal{A} can be reduced to $Q = \{\text{Pred}(a) \mid a \in \Sigma\}$ and the transition function now assumes values $\delta(\mathcal{S}, b) = \text{Pred}(b)$, for every $\mathcal{S} \in Q$ and every $b \in \Sigma \setminus \mathcal{S}$.

Proposition 3. *Let (Σ, \mathcal{C}) be a concurrent alphabet with an associated independency graph admitting a transitive orientation $<_c$. Let \leq be a total order on Σ extending $<_c$. Also assume that the dependency graph (Σ, \mathcal{C}^c) is connected. Then the adjacency matrix A is primitive.*

Proof (sketch). Under these hypotheses, by the simplifications above, it turns out that the state diagram of the automaton \mathcal{A} (defined by \leq) is strongly connected and has at least one loop. □

The hypothesis of transitivity for $<_c$ cannot be avoided to guarantee that A is primitive. For instance, in Example 2 the dependency graph (Σ, \mathcal{C}^c) is connected but the orientation $<_c$ of (Σ, \mathcal{C}) is not transitive, and in fact observe that the corresponding transition matrix is not irreducible and hence A is not primitive. Nevertheless, the smallest root of $p_{\mathcal{M}}(z)$ is simple and then the same concurrent alphabet satisfies the following theorem.

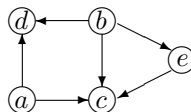
Theorem 4. *Let (Σ, \mathcal{C}) be a concurrent alphabet. If its independency graph admits a transitive orientation $<_c$, then one has $\ell \leq k$, where ℓ and k denote, respectively, the multiplicity of the smallest root of $p_{\mathcal{M}}(z)$ and the number of connected components of the dependency graph (Σ, \mathcal{C}^c) .*

Proof (sketch). First, it is well-known [18, 21] that $p_{\mathcal{M}}(z)$ is always the product of the clique polynomials of all independency subalphabets given by the connected components of (Σ, \mathcal{C}^c) . Then, each of these clique polynomials (using the additional condition that one has a transitive orientation) has a smallest root of multiplicity 1: this follows from Proposition 3 and a commutative analogue of a result in [11] stating that, when (Σ, \mathcal{C}) has a transitive orientation, its clique polynomial equals $\det(I - zA)$. □

Applying the previous theorem to relations (3) and (7), one gets the following.

Theorem 5. *Let (Σ, \mathcal{C}) be a concurrent alphabet. If its independency graph admits a transitive orientation $<_c$, then the random variables counting the number of prefixes in traces (as defined in (2)) satisfy $E[T_n] = O(E[W_n])$.*

Example 3. Consider the concurrent alphabet (Σ, \mathcal{C}) and the orientation $<_c$ of Example 2. Note that (Σ, \mathcal{C}) is connected but $<_c$ is not transitive and in fact A is not primitive. However, (Σ, \mathcal{C}) admits a (different) orientation that is transitive, given by



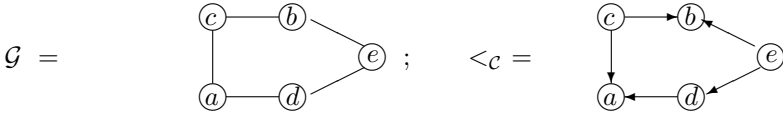
A total order extending the previous orientation is $c < d < a < e < b$. Computing matrix A with respect to this total order we obtain

$$I - zA = \begin{bmatrix} 1 - 2z & -z & -z & -z \\ 0 & 1 - z & -z & -z \\ -z & -z & 1 - z & -z \\ 0 & -z & 0 & 1 - z \end{bmatrix},$$

and hence $\det(I - zA) = 1 - 5z + 6z^2 - z^3 = p_{\mathcal{M}}(z)$. ■

The following example considers an independency graph of \mathcal{M} that does not admit any transitive orientation. In this case $p_{\mathcal{M}}(z)$ is a proper factor of $\det(I - zA)$, but its smallest root is again simple and hence $\ell \leq k$ is still true even if the hypothesis of Theorem 4 is not satisfied.

Example 4. Consider the concurrent alphabet corresponding to the following independency graph \mathcal{G} , associated to the following partial order $<_c$:



Thus the transition matrix, defined according to Sect. 4, is given by the following table:

$\tilde{A} =$	\emptyset	$\{a, b\}$	$\{a\}$	$\{b, d\}$	$\{d\}$
\emptyset	$a + b$	c	d	e	0
$\{a, b\}$	0	c	d	e	0
$\{a\}$	b	c	d	e	0
$\{b, d\}$	0	c	0	e	a
$\{d\}$	b	c	0	e	a

Accordingly, one has $\det(I - zA) = 1 - 6z + 10z^2 - 5z^3 = (1 - z)p_{\mathcal{M}}(z)$. ■

6 Conclusion

We have investigated the fundamental role played by the clique polynomial in asymptotic studies of trace monoids. Building on the factorization property (stated in Proposition 1), we got a link between the multiplicity of its smallest root and the number of connected components of some associated graph (Theorem 4). This, in turn, is the key for a new asymptotic relation between the number of prefixes in traces of length n : $E[T_n] = O(E[W_n])$ (Theorem 5), where T_n and W_n correspond to two natural models (uniform distribution over traces and over words). In the long version of this article, we plan to extend

these analyses to more general cases (including concurrent alphabets without transitive orientation).

Several other problems remain open in our context and could be at the centre of future investigations. The first one concerns the adjacency matrix A defined in Sect. 4, which does not seem to be studied too much in the previous literature; in particular, in all our examples $\det(I - zA)$ is a clique polynomial, even when the concurrent alphabet (Σ, \mathcal{C}) does not admit any transitive orientation. For this purpose, similarly to the approach used in [11] and in our proof of Theorem 4, it is possible to adapt a noncommutative approach building on links to words with forbidden patterns (see [2]). We plan to use these links to tackle the asymptotic behaviour of the variance and higher moments of $\{T_n\}$, and the limit distributions of both $\{T_n\}$ and $\{W_n\}$ for all trace monoids.

In conclusion, all these studies are further illustration of the nice interplay between complex analysis (analytic combinatorics) and the structural properties of formal languages, as also illustrated e.g. in [4, 5, 16, 17, 19, 20].

References

1. Anisimov, A.V., Knuth, D.E.: Inhomogeneous sorting. *Int. J. Comput. Inf. Sci.* **8**, 255–260 (1979). <https://doi.org/10.1007/BF00993053>
2. Asinowski, A., Bacher, A., Banderier, C., Gittenberger, B.: Analytic combinatorics of lattice paths with forbidden patterns, the vectorial kernel method, and generating functions for pushdown automata. *Algorithmica* **82**, 1–43 (2020). <https://doi.org/10.1007/s00453-019-00623-3>
3. Avellone, A., Goldwurm, M.: Analysis of algorithms for the recognition of rational and context-free trace languages. *RAIRO Theoret. Inform. Appl.* **32**, 141–152 (1998)
4. Banderier, C., Drmota, M.: Formulae and asymptotics for coefficients of algebraic functions. *Comb. Probab. Comput.* **24**, 1–53 (2015)
5. Berstel, J., Reutenauer, C.: *Rational Series and Their Languages*. Springer, Heidelberg (1988)
6. Bertoni, A., Goldwurm, M.: On the prefixes of a random trace and the membership problem for context-free trace languages. In: Huguet, L., Poli, A. (eds.) *AAECC 1987*. LNCS, vol. 356, pp. 35–59. Springer, Heidelberg (1989). <https://doi.org/10.1007/3-540-51082-6.68>
7. Bertoni, A., Goldwurm, M., Sabadini, N.: Analysis of a class of algorithms for problems on trace languages. In: Beth, T., Clausen, M. (eds.) *AAECC 1986*. LNCS, vol. 307, pp. 202–214. Springer, Heidelberg (1988). <https://doi.org/10.1007/BFb0039193>
8. Bertoni, A., Mauri, G., Sabadini, N.: Equivalence and membership problems for regular trace languages. In: Nielsen, M., Schmidt, E.M. (eds.) *ICALP 1982*. LNCS, vol. 140, pp. 61–71. Springer, Heidelberg (1982). <https://doi.org/10.1007/BFb0012757>
9. Breveglieri, L., Crespi Reghizzi, S., Goldwurm, M.: Efficient recognition of trace languages defined by repeat until loops. *Inf. Comput.* **208**, 969–981 (2010)
10. Cartier, P., Foata, D.: *Problèmes combinatoire de commutation et réarrangements*. Lecture Notes in Mathematics, vol. 85. Springer, Heidelberg (1969). <https://doi.org/10.1007/BFb0079468>

11. Choffrut, C., Goldwurm, M.: Determinants and Möbius functions in trace monoids. *Discrete Math.* **194**, 239–247 (1999)
12. Diekert, V.: Transitive orientations, Möbius functions, and complete semi-thue systems for free partially commutative monoids. In: Lepistö, T., Salomaa, A. (eds.) ICALP 1988. LNCS, vol. 317, pp. 176–187. Springer, Heidelberg (1988). https://doi.org/10.1007/3-540-19488-6_115
13. Diekert, V.: Möbius functions and confluent semi-commutations. *Theor. Comput. Sci.* **108**, 25–43 (1993)
14. Diekert, V., Rozenberg, G. (eds.): *The Book of Traces*. World Scientific, Singapore (1995)
15. Duboc, C.: *Commutations dans les monoïdes libres: un cadre théorique pour l'étude du parallélisme*. Thèse, Faculté des Sciences de l'Université de Rouen (1986)
16. Flajolet, P.: Analytic models and ambiguity of context-free languages. *Theor. Comput. Sci.* **49**, 283–309 (1987)
17. Flajolet, P., Sedgewick, R.: *Analytic Combinatorics*. Cambridge University Press, Cambridge (2009)
18. Fisher, D., Solow, A.: Dependence polynomials. *Discrete Math.* **82**, 251–258 (1990)
19. Goldwurm, M.: Some limit distributions in analysis of algorithms for problems on trace languages. *Int. J. Found. Comput. Sci.* **1**(3), 265–276 (1990)
20. Goldwurm, M.: Probabilistic estimation of the number of prefixes of a trace. *Theor. Comput. Sci.* **92**, 249–268 (1992)
21. Goldwurm, M., Santini, M.: Clique polynomials have a unique root of smallest modulus. *Inf. Process. Lett.* **75**, 127–132 (2000)
22. Mazurkiewicz, A.: *Concurrent program schemes and their interpretations*, DAIMI Rep. PB 78, Aarhus University, Aarhus (1977)
23. Rytter, W.: Some properties of trace languages. *Fund. Inform.* **7**, 117–127 (1984)
24. Salomaa, A., Soittola, M.: *Automata-Theoretic Aspects of Formal Power Series*. Springer, New York (1978). <https://doi.org/10.1007/978-1-4612-6264-0>
25. Viennot, G.X.: Heaps of pieces, I : Basic definitions and combinatorial lemmas. In: Labelle, G., Leroux, P. (eds.) *Combinatoire énumérative*. LNM, vol. 1234, pp. 321–350. Springer, Heidelberg (1986). <https://doi.org/10.1007/BFb0072524>