Recognizable Series on Graphs and Hypergraphs

Raphaël Bailly[†a], Guillaume Rabusseau[†b,c], François Denis[b]

[a] *Université Paris-1 Panthéon-Sorbonne - SAMM,* rbailly@univ-paris1.fr
[b] *Aix-Marseille Université - LIF - Qarma,* firstname.lastname@lif.univ-mrs.fr
[c] *McGill University - SOCS - RLLab,* guillaume.rabusseau@mail.mcgill.ca

Abstract

We introduce the notion of *Hypergraph Weighted Model* (HWM), a computational model that generically associates a tensor network to a graph or a hypergraph and then computes a value by generalized tensor contractions directed by its hyperedges. A series r defined on a hypergraph family is said to be recognizable if there exists a HWM that computes it. This model generalizes the notion of recognizable series on strings and trees. We present some examples on non classical graphs families such as circular strings and pictures and we study properties of the model such as closure properties and recognizability of finite support series. We conclude by a section exploring the learnability of HWMs defined over the family of circular strings.

Keywords: Recognizable Series, Graphs, Hypergraphs, Tensor Networks

1. Introduction

Real-valued functions whose domains are composed of syntactical structures, such as strings, trees or graphs, are widely used in computer science. One way to handle such functions is by means of devices computing them. Weighted automata, which are able to jointly analyze the structure of a syntactical input and to compute an associated output value, are such a computing device. They give rise to the notion of *recognizable series*. Weighted automata have been defined for strings and trees, but their extension to graphs is challenging.

Alternatively, recognizable series defined on strings and trees have equivalent algebraic characterizations. Roughly speaking, defining an automaton

[†]equal contributions

is equivalent to associating linear or multilinear operators with each element of the underlying alphabet (according to their arity); then the automaton's computation boils down to composing the operators associated with the input. We show in this paper that this algebraic formalism can be naturally extended to graphs (and hypergraphs).

More precisely, we define the notion of *Hypergraph Weighted Model* (HWM), a computational model that generically associates a tensor network to a hypergraph and that computes a value by successive generalized tensor contractions directed by its hyperedges. We say that a series r defined on a hypergraph family is HWM-recognizable if there exists a HWM M that computes it: we then denote r by r_M.

Tensors can be seen as multi-arrays composed of elements taken from a commutative semiring $\mathbb{K}$ and as it is defined, a HWM computes a series that takes its values in $\mathbb{K}$. It would be interesting to develop a study of HWM in this most general case. However, as we are mainly interested in numerical applications of HWMs, we will make the supposition that $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.

We show that when $\mathbb{K} = \mathbb{C}$, the model can be quite simplified and that a computation over a general hypergraph essentially boils down to the contraction of a tensor network associated with a closed graph (Section 3.2). Then, we show that the notion of HWM extends several models defined on particular families of graphs: this is the case for the classical notion of recognizable series on strings and trees, as well as for the more recent model of recognizable picture series (Section 3.3). We then investigate to what extent HWMs inherit fundamental properties that are satisfied by the classical notions of recognizable series on strings and trees. We will see that some of these properties are satisfied by HWMs in general, while others are not. When faced which such a property that is not satisfied in general, we will try to identify smaller families of graphs for which this property holds. For example, we show that HWMs satisfy two important closure properties: if r and s are two recognizable series defined on a family $\mathcal{H}$ of connected hypergraphs, then $r + s$ and $r \cdot s$, respectively defined for every graph $G \in \mathcal{H}$ by $(r + s)(G) = r(G) + s(G)$ and $(r \cdot s)(G) = r(G)s(G)$ (the Hadamard product[1]), are HWM-recognizable. However, HWM-recognizable series defined on general families of hypergraphs are not closed under scalar multiplication.

[1] The Hadamard product is also often denoted by $\odot$ but this symbol will have a different meaning in this paper.

Nonetheless, we show that HWM-recognizable series are closed under scalar multiplication for any family of rooted graphs (Section 3.4).

Then, we develop some examples (Section 4) that show how HWMs can be used to compute series on non classical, albeit natural, graphs families such as rooted circular strings or pictures. In particular, we show that the fact that HWM can deal with complex syntactical structures may entail sparse description of computation models: d-dimensional HWM can compute d^2-dimensional string rational series.

Recognizable series on strings and trees include polynomials (or finite support series), i.e. series that take a non zero value on only a finite number of hypergraphs. This is not always the case for recognizable series defined on more general families of hypergraphs. For example, we show that finite support series are not recognizable on the family of circular strings. The example generalizes to the case of any families of connected graphs that contain cycles. The main reason is that if a recognizable series is not null on some hypergraph G, it must be also different from zero on some *coverings* of G, i.e. connected hypergraphs made of copies of G (a notion that is close to the classical notion of *covering* for unlabeled graphs). We show that if a graph family is covering-free, then finite support series are recognizable (Section 5). Strings and trees, as any family of rooted hypergraphs, are covering-free.

The last section is devoted to give some insight on the issues that have motivated the present study and to describe some of our perspectives. In machine learning, a classical problem consists in inferring an unknown recognizable series f from examples $(x, f(x))$ of this function. The systematical use of algebraic representations for rational series on strings and trees has entailed the development of several successful methods, such as the so-called spectral methods, which try first to estimate several algebraic characteristics (spectrum, eigenspaces, singular vectors, etc) of the underlying operators from learning data, from which the target can be reconstructed. We expect that similar learning schemes for HWM-recognizable graph series can be developed. As an illustration of this general research program, we present preliminary results on the learnability of HWMs defined on the family of circular strings.

Related work. String recognizable series and weighted automata have their roots in automata theory [22, 48] and their study can be found in [9, 20, 32, 45, 46]. The extension of rational/recognizable series and weighted automaton to trees is presented in [8, 20]. A model of recognizable series on

2-dimensional words has been introduced in [10]. Computational models used to parse and generate graphs have been proposed using the formalism of grammars (see e.g. [44] and references therein). More generally, while several unweighted automata models for graphs have been proposed (see e.g. [15, 24, 50]) the quantitative setting has received less attention. The extension of weighted automata to graphs by mean of weighted logic has been considered in [19] where the authors propose a quantitative version of Thomas' unweighted model of graph acceptors [50] and show that this model is expressively equivalent to some suited monadic second order logic. The definability of graph functions in monadic second order logic has also been investigated in the work of Makovsky et al. (see e.g. [35]). Investigating the connections between HWMs and the models proposed in [19] and [35] would be of particular interest and may entail characterizations of HWM-recognizable series in terms of second-order logic; the simple observation that HMWs generalize weighted picture automata (see Section 3.3) and that both a "Nivat theorem" and the equivalence with a weighted MSO logic has been proved for this computational model [3] constitutes a first step in this direction. Tensor network diagrams have been introduced in [40], they have been extensively used in quantum theory (see e.g. [11, 47, 39]), and the interest for tensor networks has recently been growing in other fields (e.g. data mining and machine learning [13, 49, 38]). Spectral methods for inference of stochastic languages of strings/trees have been developed upon the notion of linear representation of a rational series (see [4, 29, 5, 6, 17] for example).

We recall notions on tensors and hypergraphs in Section 2, we introduce the Hypergraph Weighted Model and present some of its properties in section 3, we study some examples in Section 4, we introduce the notion of coverings and we study the recognizability of finite support series in Section 5, we present a learning scheme for HWMs on circular strings in Section 6, and we then propose a short conclusion.

2. Preliminaries

2.1. Recognizable Series on Strings and Trees

We refer to [8, 9, 14, 20, 45] for notions about recognizable series on strings and trees and we briefly recall below some basic definitions.

Let Σ be a finite *alphabet*, let Σ^* be the set of strings on Σ, and let ε denote the empty word. A *series* on Σ^* is a mapping $r : \Sigma^* \to \mathbb{K}$ where $\mathbb{K}$ is a semiring. A series r is *recognizable* if there exists a linear representation

$\langle V, \boldsymbol{\iota}, \{\mathbf{M}_x\}_{x\in\Sigma}, \boldsymbol{\tau}\rangle$ where $V = \mathbb{K}^d$ for some integer $d \geq 1$, $\boldsymbol{\iota}, \boldsymbol{\tau} \in V$ and $\mathbf{M}_x \in \mathbb{K}^{d\times d}$ for each symbol $x \in \Sigma$, such that for any $u = u_1 \dots u_n \in \Sigma^*$, $r(u) = \boldsymbol{\iota}^\top \mathbf{M}_u \boldsymbol{\tau}$, where $\mathbf{M}_u = \mathbf{M}_{u_1} \dots \mathbf{M}_{u_n}$ and $\mathbf{M}_\varepsilon = \mathbf{I}_d$ is the $d \times d$ identity matrix. The integer d is called the *dimension* of the linear representation. The *rank* of a recognizable series is the smallest d such that there exists a d-dimensional linear representation.

A *ranked* alphabet $\mathcal{F}$ is a tuple $(\Sigma, \sharp)$ where Σ is a finite alphabet and where $\sharp$ maps each symbol x of Σ to a natural number $\sharp x$ called its *arity*; for any $k \in \mathbb{N}$, let us denote $\mathcal{F}_k = \sharp^{-1}(\{k\})$. A ranked alphabet is *positive* if $\sharp$ takes its values in $\mathbb{N}_+$.

The set of trees over a ranked alphabet $\mathcal{F}$ is denoted by $T(\mathcal{F})$. A *tree series* on $T(\mathcal{F})$ is a mapping $r : T(\mathcal{F}) \to \mathbb{K}$. A tree series r is *recognizable* if there exists a tuple $\langle V, \mu, \boldsymbol{\lambda}\rangle$, where $V = \mathbb{K}^d$ for some integer $d \geq 1$, μ maps each $f \in \mathcal{F}_p$ to a p-multilinear mapping $\mu(f) \in \mathcal{L}(V^p; V)$ for each $p \geq 0$ and $\boldsymbol{\lambda} \in V$, such that $r(t) = \boldsymbol{\lambda}^\top \mu(t)$ for all t in $T(\mathcal{F})$, where $\mu(t) \in V$ is inductively defined by $\mu(f(t_1, \dots, t_p)) = \mu(f)(\mu(t_1), \dots, \mu(t_p))$.

2.2. Tensors

Let $d \geq 1$ be an integer, $V = \mathbb{K}^d$ where $\mathbb{K}$ is either[2] $\mathbb{R}$ or $\mathbb{C}$, and let $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ be the canonical basis of V. A tensor $\boldsymbol{\mathcal{T}} \in \bigotimes^k V = V \otimes \cdots \otimes V$ (k times) can uniquely be expressed as a linear combination

$$\boldsymbol{\mathcal{T}} = \sum_{i_1,\dots,i_k \in [d]} \boldsymbol{\mathcal{T}}_{i_1\dots i_k} \mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_k}$$

(where $[d] = \{1, \cdots, d\}$) of *rank-one tensors* $\mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_k}$ which form a basis of $\bigotimes^k V$ [28]. Hence, the tensor $\boldsymbol{\mathcal{T}}$ can be represented as the multi-array $(\boldsymbol{\mathcal{T}}_{i_1\dots i_k})$. The integer k is the *order* of the tensor and each axis $1, 2, \cdots, k$ of a k-th order tensor is called a *mode*. A tensor is called symmetric if it is invariant under permutation of its modes, that is $\boldsymbol{\mathcal{T}}_{i_1,\dots,i_k} = \boldsymbol{\mathcal{T}}_{\sigma(i_1),\dots,\sigma(i_k)}$ for any permutation σ over $[k]$. Throughout the paper, we will use bold lower case letters to denote vectors (e.g. $\mathbf{v}$), bold upper case letters for matrices (e.g. $\mathbf{M}$), and bold calligraphic letters for higher order tensors (e.g. $\boldsymbol{\mathcal{T}}$); the

[2]It would be interesting to generalize the results presented in this paper to the case where $\mathbb{K}$ is an arbitrary commutative semiring. However, since we are mainly interested in numerical applications of graph and hypergraph series we focus on the fields of real and complex numbers.

$d \times d$ identity matrix will be denoted by $\mathbf{I}_d$, or simply by $\mathbf{I}$ if the dimension is clear from context.

Definition 2.1. *The* tensor product *of* $\boldsymbol{\mathcal{T}} \in \bigotimes^p V$ *and* $\boldsymbol{\mathcal{U}} \in \bigotimes^q V$ *is the tensor* $\boldsymbol{\mathcal{T}} \otimes \boldsymbol{\mathcal{U}} \in \bigotimes^{p+q} V$ *defined by*

$$(\boldsymbol{\mathcal{T}} \otimes \boldsymbol{\mathcal{U}})_{i_1 \cdots i_p j_1 \cdots j_q} = \boldsymbol{\mathcal{T}}_{i_1 \cdots i_p} \boldsymbol{\mathcal{U}}_{j_1 \cdots j_q}.$$

Let $\odot : V \times V \to V$ be an associative and symmetric bilinear mapping: $\forall u, v, w \in V, u \odot v = v \odot u$ and $u \odot (v \odot w) = (u \odot v) \odot w$. The mapping $\odot$ is called a *product*.

Example 1. Let $\mathbf{1} = (1, \ldots, 1)^\top$ and let $\odot_{id}$ be defined by $\mathbf{e}_i \odot_{id} \mathbf{e}_j = \delta_{ij} \mathbf{e}_i$, where δ is the Kronecker symbol: $\odot_{id}$ is called the *identity product*.

The operation of applying the linear form $\mathbf{v} \mapsto \mathbf{1}^\top \mathbf{v}$ to the identity product $\mathbf{a} \odot_{id} \mathbf{b}$ of two vectors is related to the notions of *generalized trace* and *contraction*: if $\boldsymbol{\mathcal{A}} = \sum_{ij \in [d]} \boldsymbol{\mathcal{A}}_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ is a 2-order tensor over $\mathbb{K}^d$ (i.e. a square matrix), $\mathbf{v} = \sum_{ij \in [d]} \boldsymbol{\mathcal{A}}_{ij} \mathbf{e}_i \odot_{id} \mathbf{e}_j$ is the diagonal vector of $\boldsymbol{\mathcal{A}}$ and $\mathbf{1}^\top \mathbf{v}$ is its trace. Furthermore, if $\boldsymbol{\mathcal{A}} = \sum_{i,j \in [d]} \boldsymbol{\mathcal{A}}_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ and $\boldsymbol{\mathcal{B}} = \sum_{i,j \in [d]} \boldsymbol{\mathcal{B}}_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ are 2-order tensors over $\mathbb{K}^d$, then

$$\sum_{ijkl} \boldsymbol{\mathcal{A}}_{ij} \boldsymbol{\mathcal{B}}_{kl} \mathbf{e}_i \otimes \mathbf{1}^\top (\mathbf{e}_j \odot_{id} \mathbf{e}_k) \otimes \mathbf{e}_l = \sum_{ijl} \boldsymbol{\mathcal{A}}_{ij} \boldsymbol{\mathcal{B}}_{jl} \mathbf{e}_i \otimes \mathbf{e}_l$$

is the tensor form of the matrix product $\boldsymbol{\mathcal{A}} \cdot \boldsymbol{\mathcal{B}}$ (i.e. the contraction of the tensor $\boldsymbol{\mathcal{A}} \otimes \boldsymbol{\mathcal{B}}$ along its 2nd and 3rd modes).

2.3. Hypergraphs

Definition 2.2. *A hypergraph* $G = (V, E, \ell)$ *over a positive ranked alphabet* $(\Sigma, \sharp)$ *is given by a non empty finite set* V*, a mapping* $l : V \to \Sigma$ *and a partition* E*of* $P_G = \{(v, j) : v \in V, 1 \leq j \leq \sharp v\}$ *where* $\sharp v = \sharp \ell(v)$*.*

V is the set of *vertices*, P_G is the set of *ports*, and E is the set of *hyperedges* of G. The arity of a symbol x is equal to the number of ports of any vertex labeled by x. We will sometimes use the notation $v^{(i)}$ for the port $(v, i) \in P_G$. A hypergraph G can be represented as a bipartite graph where vertices from one partite set represent the vertices of G and vertices from the other represent its hyperedges (see Figure 1). A hypergraph is *connected* if for any partition $V = V_1 \cup V_2$, there exists a hyperedge $h \in E$ and ports $v_1^{(i)}, v_2^{(j)} \in h$ s.t. $v_1 \in V_1$ and $v_2 \in V_2$. A hypergraph is a *graph* if $|h| \leq 2$ for all $h \in E$ and a hypergraph is *closed* if $|h| \geq 2$ for all $h \in E$.

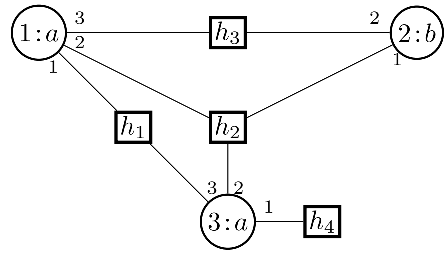


Figure 1: The hypergraph G from Example 2.

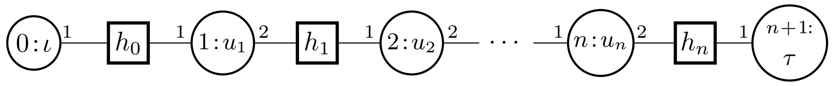


Figure 2: Graph associated with a string $u = u_1 \cdots u_n$ (where the notation $i : x$ means that $\ell(i) = x$)

Example 2. Over the ranked alphabet $\{(a, 3), (b, 2)\}$, let $V = \{v_1, v_2, v_3\}$, $\ell(v_1) = \ell(v_3) = a$, $\ell(v_2) = b$, $E = \{h_1, h_2, h_3, h_4\}$ where $h_1 = \{v_1^{(1)}, v_3^{(3)}\}$, $h_2 = \{v_1^{(2)}, v_2^{(1)}, v_3^{(2)}\}$, $h_3 = \{v_1^{(3)}, v_2^{(2)}\}$, and $h_4 = \{v_3^{(1)}\}$ (see Figure 1).

Example 3. A string $u = u_1 \ldots u_n$ over an alphabet Σ can be seen as a (hyper)graph over the ranked alphabet $(\Sigma \cup \{\iota, \tau\}, \sharp)$ where ι and τ are new symbols, $\sharp x = 2$ for any $x \in \Sigma$ and $\sharp\iota = \sharp\tau = 1$. Let $V = \{0, \cdots, n + 1\}$, $\ell(0) = \iota$, $\ell(n + 1) = \tau$, and $\ell(i) = u_i$ for $1 \leq i \leq n$. Let $E = \{h_0, h_1, \ldots, h_n\}$ where $h_0 = \{(0, 1), (1, 1)\}$ and $h_i = \{(i, 2), (i + 1, 1)\}$ for $1 \leq i \leq n$ (see Figure 2). The set of strings Σ^* gives rise to a family of graphs.

Example 4. Similarly, we can associate any tree t over a ranked alphabet $(\Sigma, \sharp)$ with a graph G_t on the ranked alphabet $(\Sigma \cup \{\lambda\}, \sharp')$ where $\sharp'(f) = \sharp f + 1$ for any $f \in \Sigma$ and where the new symbol λ of arity 1 is connected to the free port of the vertex corresponding to the root of t.

Formally, let $\mathcal{F} = (\Sigma, \sharp)$ be a ranked alphabet. A tree t over $\mathcal{F}$ can be defined as a mapping from a finite non-empty prefix-closed set $Pos(t) \subseteq \mathbb{N}^*$ to $\mathcal{F}$, satisfying the following condition: $\forall p \in Pos(t)$, if $t(p) \in \mathcal{F}_n$ then $\{j : p \cdot j \in Pos(t)\} = \{1, ..., n\}$ (using the convention $\{1, \cdots, 0\} = \emptyset$).

A tree t over $\mathcal{F}$ can be seen as a hypergraph over the ranked alphabet $(\Sigma \cup \{\lambda\}, \sharp')$ where $\sharp'(\lambda) = 1$ and $\sharp'(f) = \sharp f + 1$ for any $f \in \Sigma$. Let $V = Pos(t) \cup \{0\}$, $\ell(0) = \lambda$, and $\ell(p) = t(p)$ for any $p \in Pos(t)$. Let $E = \{\{(0, 1), (\varepsilon, 1)\}\} \cup \bigcup_{p.j \in Pos(t)} \{\{(p, j + 1), (p.j, 1)\}\}$. The set of trees $T(\mathcal{F})$ gives rise to a family of hypergraphs. The graph associated with the tree

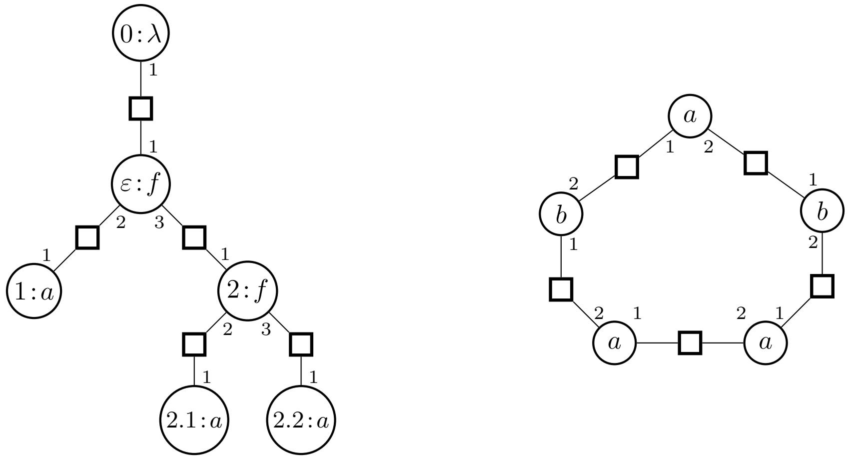


Figure 3: (left) Hypergraph G_t associated with the tree $t = f(a, f(a, a))$. (right) Example of circular string on the alphabet $\{a, b\}$

$t = f(a, f(a, a))$ is shown as an example in Figure 3.

Definition 2.3. *Given a finite alphabet Σ, let $\mathcal{F} = (\Sigma, \sharp)$ be the ranked alphabet where $\sharp x = 2$ for each $x \in \Sigma$. We say that a hypergraph $G = (V, E)$ on $\mathcal{F}$ is a* circular string *(over Σ) if and only if G is connected and every hyperedge $h \in E$ is of the form $h = \{(v, 2), (w, 1)\}$ for $v, w \in V$ (see Figure 3).*

Example 5. Another interesting extension of strings (naturally modeled by graphs) is the set of 2D-words (or pictures) $w \in \Sigma^{M \times N}$ on a finite alphabet Σ, see Section 3.3 and 4.2 for details.

3. Hypergraph Weighted Models

3.1. Definition

In this section, we give the formal definition of a Hypergraph Weighted Model. We then explain how to compute its value for a given hypergraph.

Definition 3.1. *A Hypergraph Weighted Model (HWM) on a ranked alphabet $(\Sigma, \sharp)$ is a tuple $M = \langle \mathbb{K}^d, \{\boldsymbol{\mathcal{T}}^x\}_{x\in\Sigma}, \odot, \boldsymbol{\alpha} \rangle$ where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, $\odot$ is a product on $\mathbb{K}^d$, $\boldsymbol{\alpha} \in \mathbb{K}^d$, and $\{\boldsymbol{\mathcal{T}}^x\}_{x\in\Sigma}$ is a family of tensors where each $\boldsymbol{\mathcal{T}}^x \in \bigotimes^{\sharp x} \mathbb{K}^d$. The integer d is called the* dimension *of the HWM.*

Let $G = (V, E, \ell)$ be a hypergraph and let $\Gamma = [d]^{P_G}$ be the set of mappings from P_G to $[d]$. The series *r_M computed by the HWM M is defined by*

$$r_M(G) = \sum_{\gamma\in\Gamma} \boldsymbol{\mathcal{T}}_\gamma \prod_{h\in E} \boldsymbol{\alpha}^\top \bigodot_{i\in\gamma(h)} \mathbf{e}_i$$

where $\boldsymbol{\mathcal{T}}_\gamma = \prod_{v\in V} \boldsymbol{\mathcal{T}}^v_{\gamma(v^{(1)})\dots\gamma(v^{(\sharp v)})}$ (using the notation $\boldsymbol{\mathcal{T}}^v = \boldsymbol{\mathcal{T}}^{\ell(v)}$).

Let $V = \{v_1, \cdots, v_n\}$. The tensor $\boldsymbol{\mathcal{T}}^{v_1} \otimes \boldsymbol{\mathcal{T}}^{v_2} \otimes \cdots \otimes \boldsymbol{\mathcal{T}}^{v_n}$ is of order $|P_G|$ and any element $\gamma \in \Gamma$ is a labeling of the ports of G with states in $[d]$ which can be seen as a multi-index of $[d]^{|P_G|}$. Thus, $\boldsymbol{\mathcal{T}}_\gamma$ is the $\Big(\gamma(v_1^{(1)}), \cdots, \gamma(v_1^{(\sharp v_1)}), \cdots, \gamma(v_n^{(1)}), \cdots, \gamma(v_n^{(\sharp v_n)})\Big)$-coordinate of the tensor $\bigotimes_{i=1}^n \boldsymbol{\mathcal{T}}^{v_i}$, which can be thought of as a weight for the corresponding labeling of the ports of G.

Example 6. Consider the hypergraph G from Example 2. We have

$$r_M(G) = \sum_{i_1,\cdots,i_8} \boldsymbol{\mathcal{T}}^a_{i_1i_2i_3}\boldsymbol{\mathcal{T}}^b_{i_4i_5}\boldsymbol{\mathcal{T}}^a_{i_6i_7i_8}\boldsymbol{\alpha}^\top(\mathbf{e}_{i_1}\odot\mathbf{e}_{i_8})\boldsymbol{\alpha}^\top(\mathbf{e}_{i_2}\odot\mathbf{e}_{i_4}\odot\mathbf{e}_{i_7})\boldsymbol{\alpha}^\top(\mathbf{e}_{i_3}\odot\mathbf{e}_{i_5})\boldsymbol{\alpha}^\top\mathbf{e}_{i_6}.$$

Example 7. If $\boldsymbol{\alpha} = \mathbf{1}$ and if $\odot = \odot_{id}$ (cf. Example 1), then $r_M(G) = \sum_{\gamma\in\Gamma_{Id}} \boldsymbol{\mathcal{T}}_\gamma$ where $\Gamma_{Id} = \{\gamma \in \Gamma : \forall h \in E, p, q \in h \Rightarrow \gamma(p) = \gamma(q)\}$. For the hypergraph G from Example 2, this would lead to the following contractions of the tensor $\boldsymbol{\mathcal{T}}^a \otimes \boldsymbol{\mathcal{T}}^b \otimes \boldsymbol{\mathcal{T}}^a$:

$$r_M(G) = \sum_{i_1,i_2,i_3,i_6} \boldsymbol{\mathcal{T}}^a_{i_1i_2i_3}\boldsymbol{\mathcal{T}}^b_{i_2i_3}\boldsymbol{\mathcal{T}}^a_{i_6i_2i_1} \ .$$

Remark 1. Let Σ be a finite alphabet, let $\mathbf{M}_\sigma \in \mathbb{K}^{d\times d}$ for $\sigma \in \Sigma$, and let $A = \langle \mathbb{K}^d, \{\mathbf{M}_\sigma\}_{\sigma\in\Sigma}, \odot_{id}, \mathbf{1}\rangle$ be a HWM. For any non empty word $w = w_1 \cdots w_n \in \Sigma^*$ and its corresponding circular string G_w, it follows from the definition of $\odot_{id}$ (see Example 1) that $r_A(G_w) = \mathrm{Tr}(\mathbf{M}_{w_1} \cdots \mathbf{M}_{w_n})$ (where $\mathrm{Tr}(\mathbf{M})$ is the trace of the matrix $\mathbf{M}$).

The case where $\boldsymbol{\alpha} = \mathbf{1}$ and $\odot = \odot_{id}$ will play a particular role further on. We will show that HWMs can compute traditional recognizable series on strings and trees using this choice of $\boldsymbol{\alpha}$ and $\odot$ (see Section 3.3). We will also show that when $\mathbb{K} = \mathbb{C}$ any HWM-recognizable function defined over a family of closed graphs can be computed by a HWM for which $\boldsymbol{\alpha} = \mathbf{1}$ and $\odot = \odot_{id}$ (see Proposition 3.4).

When $\boldsymbol{\alpha} = \mathbf{1}$ and $\odot = \odot_{id}$, labeling the ports of G is equivalent to labeling its edges which can help us gain some intuition on the computations of a HWM. Consider the HWM $M = \langle \mathbb{K}^d, \{\boldsymbol{\mathcal{T}}^x\}_{x\in\Sigma}, \odot_{id}, \mathbf{1}\rangle$. Since Γ_{Id} is isomorphic to $[d]^E$, the computation of M on a hypergraph $G = (V, E, \ell)$ can be interpreted in the following way:

- Each component $\boldsymbol{\mathcal{T}}^x_{i_1,\cdots,i_{\sharp x}}$ of a tensor $\boldsymbol{\mathcal{T}}^x$ represents the *weight* of a vertex labeled by x when its first port is in state i_1, its second port in state i_2...
- Each *configuration* in $[d]^E$ (i.e. a labeling of the hyperedges of G with states in $[d]$) assigns a state to each port of the hypergraph, thus a configuration assigns a weight to each vertex of G (using the tensors $\boldsymbol{\mathcal{T}}^x$). The product of these weights represents the weight of a configuration.
- The value computed by M is the sum of the weights of all possible configurations in $[d]^E$.

We conclude this section by two simple remarks and by formally defining our notion of recognizability for hypergraph series.

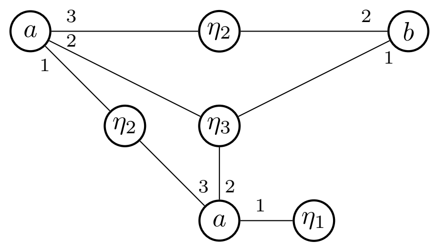
Remark 2. Let $A = \langle \mathbb{R}^d, \{\boldsymbol{\mathcal{A}}^x\}_{x\in\Sigma}, \odot, \boldsymbol{\alpha}\rangle$ be a HWM. Each tensor $\boldsymbol{\mathcal{A}}^x$ can be decomposed as a sum of rank one tensors $\boldsymbol{\mathcal{A}}^x = \sum_{r=1}^{R} \mathbf{a}_r^{(x,1)} \otimes \cdots \otimes \mathbf{a}_r^{(x,\sharp x)}$ where R is the maximum rank of the tensors $\boldsymbol{\mathcal{T}}^x$ for $x \in \Sigma$. The computation of the HWM A on $G = (V, E, \ell)$ can then be written as $r(G) = \prod_{h\in E} \boldsymbol{\alpha}^\top \left[\bigodot_{(v,i)\in h} \left(\sum_{r=1}^{R} \mathbf{a}_r^{(\ell(v),i)}\right)\right]$.

Remark 3. If G is a hypergraph with two connected components G_1 and G_2, we have $r_M(G) = r_M(G_1) \cdot r_M(G_2)$ for any HWM M.

Definition 3.2. *Let $\mathcal{H}$ be a family of hypergraphs on a ranked alphabet $(\Sigma, \sharp)$. We say that a hypergraph series $r : \mathcal{H} \to \mathbb{K}$ is* recognizable *if and only if there exists a HWM M such that $r_M(G) = r(G)$ for all $G \in \mathcal{H}$.*

As in the previous definition, we will often present results or properties that hold for series defined on a particular family of hypergraphs $\mathcal{H}$. By this we mean that hypergraphs outside of $\mathcal{H}$ are disregarded, i.e. the series may behave arbitrarily on them.

In the following sections, we first present some simplifications of the model that can be assumed in particular situations. We then show that HWMs satisfy some basic properties which are desirable for a model extending the notion of recognizable series to hypergraphs.



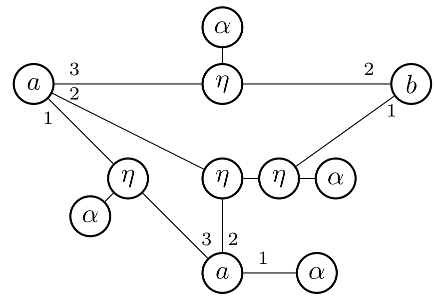


Figure 4: Two possible mappings from the hypergraph in Example 2 to a graph. (left) Each arity-j hyperedge is replaced with a vertex labeled by the new symbol η_j. (right) Each arity-j hyperedge is replaced by a chain of $j-1$ vertices labeled by the arity-3 symbol η connected to one vertex labeled α. Note that any numbering of the ports of the vertices labeled by the symbol η or η_j can be envisaged (this follows from the symmetry and associativity of $\odot$).

3.2. HWMs over closed graphs

We first show how any HWM-recognizable series on hypergraphs can be computed by a HWM defined over closed graphs using a specific mapping from hypergraphs to closed graphs. This mapping consists in replacing each hyperedge h in the hypergraph by a vertex labeled with a new symbol η_j of arity j, where $j = |h|$ (see Figure 4, left).

Formally, let $G = (V, E, \ell)$ be a hypergraph on the ranked alphabet $(\Sigma, \sharp)$ and let $\mathcal{F}' = (\Sigma \cup \{\eta_j\}_{j\geq 1}, \sharp')$ where $\sharp' x = \sharp x$ for any $x \in \Sigma$ and $\sharp' \eta_j = j$ for any $j \geq 1$. Let P_G be the set of ports of G and let $P_E = \{(h, k) : h \in E,\ 1 \leq k \leq |h|\}$. Finally, let $f : P_G \to P_E$ be any bijection satisfying $f(p) = (h, k) \Rightarrow p \in h$.

We define the closed graph $\texttt{graph}(G) = (V', E', \ell')$ on the ranked alphabet $\mathcal{F}'$ by[3]

- $V' = V \cup E$ (thus the set of ports of $\texttt{graph}(G)$ is $P_G \cup P_E$),
- $\ell'(v) = \ell(v)$ for any $v \in V$ and $\ell'(h) = \eta_{|h|}$ for any $h \in E$,
- $E' = \big\{\{p, f(p)\} : p \in P_G\big\}$.

[3]To avoid cluttering the notations we omit the dependency on the bijection f in the notation $\texttt{graph}(G)$. This is without consequences since the value computed by any HWM on $\texttt{graph}(G)$ does not depend on the particular choice of f.

Proposition 3.3. *Let $A = \langle \mathbb{K}^d, \{\boldsymbol{\mathcal{T}}^x\}_{x\in\Sigma}, \odot, \boldsymbol{\alpha}\rangle$ be a HWM on the ranked alphabet $(\Sigma, \sharp)$. The HWM $B = \langle \mathbb{K}^d, \{\boldsymbol{\mathcal{T}}^x\}_{x\in\Sigma} \cup \{\boldsymbol{\mathcal{T}}^{\eta_j}\}_{j\geq 1}, \odot_{id}, \mathbf{1}\rangle$, where $\boldsymbol{\mathcal{T}}^{\eta_j}$ is defined by*

$$\boldsymbol{\mathcal{T}}^{\eta_j}_{i_1,\cdots,i_j} = \boldsymbol{\alpha}^\top(\mathbf{e}_{i_1} \odot \cdots \odot \mathbf{e}_{i_j}) \text{ for all } i_1, \cdots, i_j \in [d],$$

satisfies $r_A(G) = r_B(graph(G))$ for any hypergraph G.

Proof. Let $G = (V, E, \ell)$ be a hypergraph and let $\texttt{graph}(G) = (V', E', \ell')$ be the closed graph defined previously. Let P be the set of ports of G and P' be the set of ports of $\texttt{graph}(G)$. Finally let $\Gamma = [d]^P$ and $\Gamma' = [d]^{P'}$. The key ingredient of the proof consists in exhibiting the isomorphism between Γ and Γ'_{Id}. For any $\gamma \in \Gamma$, let $\gamma' \in \Gamma'$ be defined by

$$\gamma'(p) = \gamma'(f(p)) = \gamma(p) \text{ for all } p \in P .$$

Since any edge in E' is of the form $\{p, f(p)\}$ we have $\gamma' \in \Gamma'_{Id} = \{\gamma' \in \Gamma' : \forall h \in E', p, q \in h \Rightarrow \gamma'(p) = \gamma'(q)\}$ (by definition of γ'). Furthermore, we have $|\Gamma| = |\Gamma'_{Id}|$ (since Γ'_{Id} is isomorphic to $[d]^{E'}$) and the mapping $g : \gamma \mapsto \gamma'$ is injective, thus g is a bijection between Γ and Γ'_{Id}.

It then follows from the definition of $\odot_{id}$ (cf. Example 7) that

$$r_B(\texttt{graph}(G)) = \sum_{\gamma'\in\Gamma'_{Id}} \prod_{v'\in V'} \boldsymbol{\mathcal{T}}^{\ell'(v')}_{\gamma'(v')} = \sum_{\gamma'\in\Gamma'_{Id}} \prod_{v\in V} \boldsymbol{\mathcal{T}}^{\ell'(v)}_{\gamma'(v)} \prod_{h\in E} \boldsymbol{\mathcal{T}}^{\ell'(h)}_{\gamma'(h)},$$

where we used the notation $\boldsymbol{\mathcal{T}}^{\ell'(v')}_{\gamma'(v')} = \boldsymbol{\mathcal{T}}^{\ell'(v')}_{\gamma'(v',1),\cdots,\gamma'(v',\sharp'v'))}$. For any $h = \{p_1, \cdots, p_n\} \in E$ and any $\gamma \in \Gamma$, let $\boldsymbol{\mathcal{T}}^{\eta_n}_{\gamma(h)} = \boldsymbol{\mathcal{T}}^{\eta_n}_{\gamma(p_1),\cdots,\gamma(p_n)}$ (note that the ordering of the ports in h is not relevant since the tensor $\boldsymbol{\mathcal{T}}^{\eta_n}$ is symmetric). We then have $\boldsymbol{\mathcal{T}}^{\ell'(h)}_{\gamma'(h)} = \boldsymbol{\mathcal{T}}^{\eta_{|h|}}_{\gamma(h)}$ for any $h \in E$ and $\gamma' \in \Gamma'_{Id}$. Finally, using the isomorphism between Γ and Γ'_{Id} we obtain

$$r_B(\texttt{graph}(G)) = \sum_{\gamma\in\Gamma} \prod_{v\in V} \boldsymbol{\mathcal{T}}^{\ell(v)}_{\gamma(v)} \prod_{h\in E} \boldsymbol{\mathcal{T}}^{\eta_{|h|}}_{\gamma(h)} = \sum_{\gamma\in\Gamma} \prod_{v\in V} \boldsymbol{\mathcal{T}}^{\ell(v)}_{\gamma(v)} \prod_{h\in E} \boldsymbol{\alpha}^\top \bigodot_{i\in\gamma(h)} \mathbf{e}_i = r_A(G) .$$

□

The main drawback of the mapping from hypergraphs to closed graphs used in the previous proposition is that it needs to introduce new symbols with unbounded arity. The next remark shows that it is possible to circumvent this issue by exploiting the structure of the tensors $\boldsymbol{\mathcal{T}}^{\eta_j}$.

Remark 4. Let η be a new symbol of arity 3 and define its associated tensor by $\mathcal{T}^{\eta}_{i_1,i_2,i_3} = \mathbf{e}_{i_1}^\top(\mathbf{e}_{i_2} \odot \mathbf{e}_{i_3})$. Observing that $\mathbf{e}_{i_2} \odot \mathbf{e}_{i_3} = \sum_{i_1} \mathcal{T}^{\eta}_{i_1,i_2,i_3}\mathbf{e}_{i_1}$ for any $i_2, i_3 \in [d]$, one can check that the tensors $\mathcal{T}^{\eta_j}$ from the previous proposition satisfy the following recurrence relation:

$$\mathcal{T}^{\eta_1} = \boldsymbol{\alpha} \text{ and } \mathcal{T}^{\eta_j}_{i_1,\cdots,i_j} = \sum_k \mathcal{T}^{\eta}_{k,i_1,i_2}\mathcal{T}^{\eta_{j-1}}_{k,i_3,\cdots,i_j} \ .$$

Thus, each vertex labeled by a symbol η_j in $\texttt{graph}(G)$ can be replaced by a chain of $j-1$ vertices labeled by the new symbol η (this chain has $j+1$ free ports) connected to one vertex labeled by the new symbol α (see Figure 4, right). It is then easy to build a new HWM using the tensors $\mathcal{T}^{\eta}$ and $\mathcal{T}^{\alpha} = \boldsymbol{\alpha}$ on the *finite* alphabet $\Sigma \cup \{\eta, \alpha\}$ that will compute the same series. Note that the number of new vertices added to G is unbounded.

The following proposition shows that any recognizable $\mathbb{C}$-valued series on closed graphs can be computed by a HWM with coefficients in $\mathbb{C}$ using the identity product $\odot_{id}$ and the vector $\mathbf{1}$.

Proposition 3.4. *Let $A = \langle \mathbb{C}^d, \{\mathcal{A}^x\}_{x\in\Sigma}, \odot_A, \boldsymbol{\alpha}\rangle$ be a HWM. There exists a HWM $B = \langle \mathbb{C}^d, \{\mathcal{B}^x\}_{x\in\Sigma}, \odot_{id}, \mathbf{1}\rangle$ such that $r_B(G) = r_A(G)$ for any closed graph G.*

Proof. We consider the decomposition $\mathcal{A}^x = \sum_{r=1}^R \mathbf{a}_r^{(x,1)} \otimes \cdots \otimes \mathbf{a}_r^{(x,\sharp x)}$ for each $x \in \Sigma$ (see Remark 2). Let $\mathbf{M} \in \mathbb{R}^{d\times d}$ be the matrix defined by $\mathbf{M}_{ij} = \boldsymbol{\alpha}^\top(\mathbf{e}_i \odot_A \mathbf{e}_j)$ and check that $\mathbf{u}^\top\mathbf{M}\mathbf{v} = \boldsymbol{\alpha}^\top(\mathbf{u}\odot_A\mathbf{v})$ and $\mathbf{1}^\top(\mathbf{u}\odot_{id}\mathbf{v}) = \mathbf{u}^\top\mathbf{v}$ for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$. Let $\mathbf{Q} \in \mathbb{C}^{d\times d}$ be such that $\mathbf{M} = \mathbf{Q}^\top\mathbf{Q}$ (such a decomposition exists since $\mathbf{M}$ is symmetric) and let $\mathcal{B}^x = \sum_{r=1}^R(\mathbf{Q}\mathbf{a}_r^{(x,1)}) \otimes \cdots \otimes (\mathbf{Q}\mathbf{a}_r^{(x,\sharp x)})$. For any closed graph $G = (V, E, \ell)$, it then follows that

$$\begin{aligned}
r_B(G) &= \prod_{\substack{h\in E,\\ h=\{(v,i),(w,j)\}}} \mathbf{1}^\top\left[\left(\sum_{r=1}^R \mathbf{Q}\mathbf{a}_r^{(\ell(v),i)}\right)\odot_{id}\left(\sum_{r=1}^R \mathbf{Q}\mathbf{a}_r^{(\ell(w),j)}\right)\right] \\
&= \prod_{\substack{h\in E,\\ h=\{(v,i),(w,j)\}}} \left(\sum_{r=1}^R \mathbf{a}_r^{(\ell(v),i)}\right)^\top \mathbf{Q}^\top\mathbf{Q}\left(\sum_{r=1}^R \mathbf{a}_r^{(\ell(w),j)}\right) \\
&= \prod_{\substack{h\in E,\\ h=\{(v,i),(w,j)\}}} \boldsymbol{\alpha}^\top\left[\left(\sum_{r=1}^R \mathbf{a}_r^{(\ell(v),i)}\right)\odot_A\left(\sum_{r=1}^R \mathbf{a}_r^{(\ell(w),j)}\right)\right] \\
&= r_A(G).
\end{aligned}$$

□

Remark 5. We exhibit two counter-examples to give an intuition on why the previous proposition only holds *over* $\mathbb{C}$ and for *closed* graphs.

- Consider the one-dimensional HWM $A = \langle \mathbb{R}, \{\boldsymbol{\mathcal{A}}^a = \mathbf{e}_1\}, \odot, \boldsymbol{\alpha} = \mathbf{e}_1\rangle$ on the one-letter alphabet $\{a(\cdot)\}$ where $\odot$ is defined by $\mathbf{e}_1 \odot \mathbf{e}_1 = -\mathbf{e}_1$. We have $r_A(a\!-\!\square\!-\!a) = -1$ and it is easy to check that any one-dimensional HWM with real coefficients using the product $\odot_{id}$ and the vector $\mathbf{1}$ will assign a positive value to the graph $a\!-\!\square\!-\!a$.

- Consider the one-dimensional HWM

  $$B = \langle \mathbb{C}, \{\boldsymbol{\mathcal{B}}^a = \mathbf{e}_1, \boldsymbol{\mathcal{B}}^b = \mathbf{e}_1 \otimes \mathbf{e}_1\}, \odot_{id}, \boldsymbol{\alpha} = \frac{1}{2}\mathbf{e}_1\rangle$$

  on the ranked alphabet $\{a(\cdot), b(\cdot,\cdot)\}$.

  For any HWM $\widetilde{B} = \langle \mathbb{C}, \{\widetilde{\boldsymbol{\mathcal{B}}}^a, \widetilde{\boldsymbol{\mathcal{B}}}^b\}, \odot_{id}, \mathbf{1}\rangle$ we have $\widetilde{\boldsymbol{\mathcal{B}}}^a = \beta_a \mathbf{e}_1$ and $\widetilde{\boldsymbol{\mathcal{B}}}^b = \beta_b \mathbf{e}_1 \otimes \mathbf{e}_1$ for some $\beta_a, \beta_b \in \mathbb{C}$. Suppose $r_B = r_{\widetilde{B}}$, then since $r_B(a\!-\!\square) = \frac{1}{2}$ and $r_B(\square\!-\!b\!-\!\square) = \frac{1}{4}$ we must have $\beta_a = \frac{1}{2}$ and $\beta_b = \frac{1}{4}$, but then $r_B(a\!-\!\square\!-\!b\!-\!\square) = \frac{1}{4}$ and $r_{\widetilde{B}}(a\!-\!\square\!-\!b\!-\!\square) = \frac{1}{8}$, a contradiction.

The results presented in this section show that when $\mathbb{K} = \mathbb{C}$, we can without any loss of generality restrict our attention to HWMs on closed graphs (instead of hypergraphs) using the product $\odot_{id}$ and the vector $\mathbf{1}$. Consequently, we will sometimes omit the product $\odot$ and the vector $\boldsymbol{\alpha}$ in the definition of a HWM, implying that $\odot = \odot_{id}$ and $\boldsymbol{\alpha} = \mathbf{1}$. Nonetheless, the flexibility offered by the choice of $\odot$ and $\boldsymbol{\alpha}$ will allow us to simplify some of the proofs and we will always make it clear that the results we present hold for hypergraphs.

HWMs and tensor networks. When $\boldsymbol{\alpha} = \mathbf{1}$ and $\odot = \odot_{id}$ the computation of a HWM on a given hypergraph H is equivalent to the contraction of the tensor network naturally associated with the closed graph $\texttt{graph}(H)$: each node in the graph represents the tensor associated with its label and the contraction of this tensor network (along the edges of the graph) results in a scalar equal to the value computed by the HWM on H. Considering graphs rather than hypergraphs for sake of simplicity, one can pursue this analogy between HWMs and tensor networks and define the computations of a HWM $M = \langle \mathbb{K}^d, \{\boldsymbol{\mathcal{T}}^x\}_{x\in\Sigma}, \odot_{id}, \mathbf{1}\rangle$ in an inductive way by mapping any graph G with

k free ports to the k-th order tensor $\boldsymbol{\mathfrak{T}}^G$ naturally obtained by interpreting G as a tensor network:

- the singleton graph containing only one vertex labeled by the symbol x (which has $\sharp x$ free ports) is mapped to the tensor $\boldsymbol{\mathfrak{T}}^x$;

- the union of two graphs G_1 (with k_1 free ports) and G_2 (with k_2 free ports) obtained by juxtaposing G_1 and G_2 is mapped to the tensor product of the tensors $\boldsymbol{\mathfrak{T}}^{G_1}$ and $\boldsymbol{\mathfrak{T}}^{G_2}$ (which is of order $k_1 + k_2$);

- the graph obtained by connecting two free ports p_1 and p_2 in a graph G with $k \geq 2$ free ports with a new edge is mapped to the tensor obtained by contracting the modes corresponding to p_1 and p_2 in $\boldsymbol{\mathfrak{T}}^G$ (resulting in a tensor of order $k - 2$).

The value computed by M is then obtained by summing the components of $\boldsymbol{\mathfrak{T}}^G$ (which corresponds to plugging the vector $\boldsymbol{\alpha} = \mathbf{1}$ in the free ports of the tensor network $\boldsymbol{\mathfrak{T}}^G$). We refer the reader to [42, Chapter 2] for more details about the connections between HWMs and tensor networks.

The connection between computations of weighted automata and tensor networks has been previously noticed in e.g. [16] where the authors show the close relationship between hidden Markov models (i.e. probabilistic automata) and matrix product states tensor networks. The extension of strings and trees weighted automata to hypergraphs we propose here can be seen as a natural prolongation of this analogy to the more general setting of hypergraphs. We again refer the reader to [42, Chapter 2] for a more detailed discussion about the connection between classical weighted automata, tensor networks, and HWMs.

From a practical perspective, it is well known that computational problems on tensor networks can be hard. For example, deciding if the scalar computed by a given tensor network is positive is NP-hard [31] and so is the problem of finding an optimal sequence of contractions [12]. Nonetheless, developing efficient heuristics to achieve fast computations in tensor networks has been an active line of research in the physics community (see e.g. [41]) and the methods and results obtained in this community will certainly prove useful for the practical implementation of HWM computations.

3.3. Strings, Trees, and Pictures

String and Trees. The following propositions show that the proposed model naturally generalizes the notion of linear representation of recognizable series on strings and trees.

Proposition 3.5. *Let $r = \langle \mathbb{K}^d, \boldsymbol{\iota}, \{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\tau}\rangle$ be a recognizable series on Σ^*. For any word $w \in \Sigma^*$, let G_w be the associated hypergraph on the ranked alphabet $(\Sigma \cup \{\iota, \tau\}, \sharp)$, whose construction is described in Example 3. Consider the HWM $M = \langle \mathbb{K}^d, \{\boldsymbol{\mathfrak{T}}^x\}_{x\in\Sigma\cup\{\iota,\tau\}}, \odot_{id}, \mathbf{1}\rangle$ where $\boldsymbol{\mathfrak{T}}^\tau = \boldsymbol{\tau}$, $\boldsymbol{\mathfrak{T}}^\iota = \boldsymbol{\iota}$ and $\boldsymbol{\mathfrak{T}}^\sigma = \mathbf{M}^\sigma$ for all $\sigma \in \Sigma$.*

Then, $r(w) = r_M(G_w)$ for all strings $w \in \Sigma^*$.

Proof. Let $w = w_1 \cdots w_n$. We have

$$r(w) = \boldsymbol{\iota}^\top \mathbf{M}^{w_1} \cdots \mathbf{M}^{w_n} \boldsymbol{\tau} = \sum_{i_0,\dots,i_n} \boldsymbol{\iota}_{i_0} \mathbf{M}^{w_1}_{i_0,i_1} \dots \mathbf{M}^{w_1}_{i_{n-1},i_n} \boldsymbol{\tau}_{i_n}$$
$$= \sum_{i_0,\dots,i_n} \boldsymbol{\mathfrak{T}}^\iota_{i_0} \boldsymbol{\mathfrak{T}}^{w_1}_{i_0,i_1} \dots \boldsymbol{\mathfrak{T}}^{w_n}_{i_{n-1},i_n} \boldsymbol{\mathfrak{T}}^\tau_{i_n} = r_M(G_w).$$

□

In the previous proposition, the vectors $\boldsymbol{\iota}$ and $\boldsymbol{\tau}$ of a linear representation were directly encoded in the structure of the graph representation of a string w on Σ using the new symbols ι and τ. The next proposition shows that it is possible to encode these linear forms in the vector $\boldsymbol{\alpha}$ of a HWM with complex coefficients, using a graph representation of strings without new symbols: for any string $w = w_1 \cdots w_n$ over Σ, we consider the graph $H_w = (V, E, \ell)$ on $(\Sigma, \sharp)$ where $V = [n]$, $\ell(i) = w_i$, and the set of hyperedges is composed of $\{(1, 1)\}$, $\{(n, 2)\}$, and $\{(i, 2), (i + 1, 1)\}$ for $i \in [n - 1]$ (note that the graph representation of a string is different from the graph representation of its mirror because of the identification of the ports).

Proposition 3.6. *Let $r = \langle \mathbb{R}^d, \boldsymbol{\iota}, \{\mathbf{M}_\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\tau}\rangle$ be a real-valued recognizable string series on Σ^*. There exists a HWM $M = \langle \mathbb{C}^d, \{\boldsymbol{\mathfrak{T}}^\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\alpha}, \odot\rangle$ such that $r_M(H_w) = r(w)$ for all $w \in \Sigma^*$.*

Proof. We first show that given a recognizable string series $r = \langle \mathbb{R}^d, \boldsymbol{\iota}, \{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\tau}\rangle$ there exists a recognizable series $s = \langle \mathbb{C}^d, \boldsymbol{\alpha}, \{\mathbf{N}^\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\alpha}\rangle$, such that $s(w) = r(w)$ for all $w \in \Sigma^*$. Indeed, let $(\mathbf{e}_1, \cdots, \mathbf{e}_d)$ be a basis of $\mathbb{R}^d$ such that $\mathbf{e}_i^\top \boldsymbol{\tau} \neq 0$ and $\mathbf{e}_i^\top \boldsymbol{\iota} \neq 0$ for all $i \in [d]$. Let $\mathbf{D} \in \mathbb{C}^{d\times d}$ be the diagonal matrix defined by $\mathbf{D}_{ii} = (\mathbf{e}_i^\top \boldsymbol{\tau})^{1/2} / (\mathbf{e}_i^\top \boldsymbol{\iota})^{1/2}$ (where $x^{1/2} = \mathrm{i}|x|^{1/2}$ if $x < 0$). We have $\mathbf{D}\boldsymbol{\iota} = \mathbf{D}^{-1}\boldsymbol{\tau}$ and the series $s : \langle \mathbb{C}^d, \mathbf{D}\boldsymbol{\iota}, \mathbf{D}^{-1}\boldsymbol{\tau}, \{\mathbf{D}^{-1}\mathbf{M}^\sigma\mathbf{D}\}_{\sigma\in\Sigma}\rangle$ is such that $s(w) = r(w)$ for all $w \in \Sigma^*$.

Then, the HWM $M = \langle \mathbb{C}^d, \{\mathfrak{T}^x\}_{x\in\Sigma}, \boldsymbol{\alpha}, \odot\rangle$, where $\mathfrak{T}^\sigma = \mathbf{M}^\sigma$, $\boldsymbol{\alpha} = \mathbf{D}\boldsymbol{\iota} = \mathbf{D}^{-1}\boldsymbol{\tau}$, and $\odot$ is defined by $\mathbf{e}_i \odot \mathbf{e}_j = \delta_{ij}\frac{1}{\boldsymbol{\alpha}_i}\mathbf{e}_i$, is such that $r_M(H_w) = r(w)$ for all $w \in \Sigma^*$. □

Proposition 3.7. *Let $r = \langle V, \mu, \lambda\rangle$ be a recognizable series on trees on the ranked alphabet $\mathcal{F} = (\Sigma, \sharp)$. For any tree t over $\mathcal{F}$, let $G^t = (V_t, E_t)$ be the associated hypergraph on the ranked alphabet $(\Sigma \cup \{\lambda\}, \sharp')$, whose construction is described in Example 4. Consider the HWM $M : \langle V, \{\mathfrak{T}^x\}_{x\in\Sigma\cup\{\lambda\}}, \odot_{id}, \mathbf{1}\rangle$ where $\mathfrak{T}^\lambda = \lambda$ and $\mathfrak{T}^f$ is defined by $\mathfrak{T}^f_{i_0\dots i_k} = \mathbf{e}_{i_0}^\top \mu(f)(\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_k})$ for all k and $f \in \mathcal{F}_k$.*

Then, $r(t) = r_M(G^t)$ for every tree t over $\mathcal{F}$.

Proof. For any $\gamma \in \Gamma_{Id}$, let $\mathfrak{U}_\gamma = \prod_{v\in V_t\setminus\{0\}} \mathfrak{T}^{\ell(v)}_{\gamma(v,1)\dots\gamma(v,\sharp v)}$. We first prove by induction on t that $\mu(t) = \sum_{\gamma\in\Gamma_{Id}} \mathfrak{U}_\gamma \mathbf{e}_{\gamma(\varepsilon,1)}$.

If $t = a$, then $\sum_{\gamma\in\Gamma_{Id}} \mathfrak{U}_\gamma \mathbf{e}_{\gamma(\varepsilon,1)} = \sum_{i\in[d]} \mathfrak{T}^a_i \mathbf{e}_i = \mu(a)$.

If $t = f(t_1, \dots, t_k)$, first notice that $\Gamma_{Id} = \Gamma^{(0)}_{Id} \times \Gamma^{(1)}_{Id} \times \dots \times \Gamma^{(k)}_{Id}$, where $\Gamma^{(0)}_{Id}$ fixes the values of the port $(\varepsilon, 1)$ and where $\Gamma^{(j)}_{Id}$, for $j \in [k]$, fixes the values of the ports of the subgraph corresponding to the subtree t_j.

We then have

$$\begin{aligned}
\mu(f(t_1, \dots, t_k)) &= \mu(f)(\mu(t_1), \dots, \mu(t_k)) \\
&= \sum_{i_0,i_1,\dots,i_k} \mathfrak{T}^f_{i_0,i_1,\dots,i_n} \prod_{j\in[k]} \mathbf{e}_{i_j}^\top \mu(t_j)\mathbf{e}_{i_0} \\
&= \sum_{i_0,i_1,\dots,i_k} \mathfrak{T}^f_{i_0,i_1,\dots,i_n} \prod_{j\in[k]} \mathbf{e}_{i_j}^\top \left(\sum_{\gamma\in\Gamma^{(j)}_{Id}} \mathfrak{U}_\gamma \mathbf{e}_{\gamma(j,1)}\right)\mathbf{e}_{i_0} \\
&= \sum_{i_0,i_1,\dots,i_k} \mathfrak{T}^f_{i_0,i_1,\dots,i_n} \prod_{j\in[k]} \sum_{\gamma\in\Gamma^{(j)}_{Id},\gamma(j,1)=i_j} \mathfrak{U}_\gamma \mathbf{e}_{i_0} \\
&= \sum_{i_0,\gamma_1\in\Gamma^{(1)}_{Id},\dots,\gamma_k\in\Gamma^{(k)}_{Id}} \mathfrak{T}^f_{i_0,\gamma_1(1,1),\dots,\gamma_k(k,1)} \prod_{j\in[k]} \mathfrak{U}_{\gamma_j}\mathbf{e}_{i_0} = \sum_{\gamma\in\Gamma_{Id}} \mathfrak{U}_\gamma \mathbf{e}_{\gamma(\varepsilon,1)}
\end{aligned}$$

and it is easy to check that

$$\lambda(\mu(t)) = \sum_i \sum_{\gamma\in\Gamma_{Id}} \mathfrak{T}^\lambda_i \mathfrak{U}_\gamma \mathbf{e}_i^T \mathbf{e}_{\gamma(\varepsilon,1)} = \sum_{\gamma\in\Gamma_{Id}} \mathfrak{T}^\lambda_{\gamma(\varepsilon,1)} \mathfrak{U}_\gamma = \sum_{\gamma\in\Gamma_{Id}} \mathfrak{T}_\gamma = r_M(t).$$

□

Pictures. We now show that HWMs also generalize the more recent model of recognizable picture series.

A *picture* $p \in \Sigma^{++}$ over a finite alphabet Σ is defined as a non-empty rectangular array of elements of Σ, formally $\Sigma^{++} = \cup_{m,n\geq 1}\Sigma^{m\times n}$. We write $p_{i,j}$ for the component of p at position (i,j). A *picture language* is a set of pictures, while a *picture series* is a function from Σ^{++} to a commutative semiring. Regular picture languages can equivalently be described in terms of automata, set of tiles, rational operations or monadic second order logic [25, 26, 30, 34]. The extension of regular picture languages to the quantitative setting led to the definition of *recognizable picture series* whose theoretical study has been of recent interest [10, 36, 23, 3]. Recognizable picture series have been first introduced in [10] by means of weighted picture automata.

Definition 3.8. *A* weighted (quadropolic) picture automaton (WPA) *[10] on a finite alphabet Σ is a tuple $\mathcal{A} = \langle Q, R, F_w, F_n, F_e, F_s, \delta\rangle$ consisting of a finite set of states Q, a finite set of rules $R \subseteq \Sigma \times Q^4$, four poles of acceptance $F_w, F_n, F_e, F_s \subseteq Q$, and a weight function $\delta : R \to \mathbb{K}$, where $\mathbb{K}$ is a commutative semiring.*

Given a rule $(\sigma, q_w, q_n, q_e, q_s) \in R$ we denote by $\ell(r)$ its label σ and by $w(r) = q_w, n(r) = q_n, e(r) = q_e$, and $s(r) = q_s$ the states corresponding to its four poles.

A *run* c of $\mathcal{A}$ on a picture $p \in \Sigma^{m\times n}$ is an element in $R^{m\times n}$ satisfying the following compatibility properties:

$$\begin{array}{l} \forall i \leq m-1, j \leq n : s(c_{i,j}) = n(c_{i+1,j}) \\ \forall i \leq m, j \leq n-1 : e(c_{i,j}) = w(c_{i,j+1}) \end{array} \quad (1)$$

and $\ell(c_{i,j}) = p_{i,j}$ for all $i \leq m, j \leq n$. A run is *successful* if its outer pole-states are in the respective poles of acceptance, that is

$$\forall i \leq m, j \leq n : w(c_{i,1}) \in F_w, n(c_{m,j}) \in F_n, e(c_{i,n}) \in F_e, s(c_{1,j}) \in F_s.$$

We denote by $R(p)$ the set of all successful runs on a picture p.

We extend the weight function δ to runs by setting $\delta(c) = \prod_{i,j}\delta(c_{i,j})$. The *weight* of a picture p is the sum of the weights of all successful runs on p. It defines a picture series $r_{\mathcal{A}} : \Sigma^{++} \to \mathbb{K}$ with $r_{\mathcal{A}}(p) = \sum_{c\in R(p)}\delta(c)$. If there are no successful run on p then $r_{\mathcal{A}}(p) = 0$.

With each picture $p \in \Sigma^{++}$ of size $m \times n$ we associate a closed graph $G_p = (V, E, \ell)$ on the ranked alphabet $\mathcal{F}_\Sigma = (\Sigma \cup \{w, n, e, s\}, \sharp)$ where $\sharp\sigma = 4$

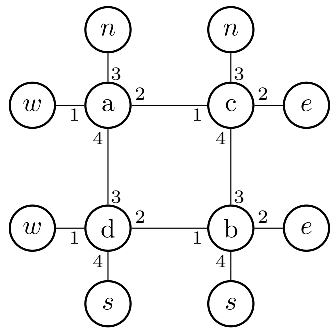


Figure 5: Graph G_p associated with a 2×2 picture p on the alphabet $\Sigma = \{a, b, c, d\}$.

for all $\sigma \in \Sigma$ and $\sharp w = \sharp n = \sharp e = \sharp s = 1$. The graph G_p is constructed in a straightforward way by translating p into a graph and adding nodes to the border of the picture (see Figure 5): nodes on the west border will be labeled by the symbol w, on the east border by e...

Let $\mathcal{A} = \langle Q, R, F_w, F_n, F_e, F_s, \delta \rangle$ be a WPA with d states $(q_1, \cdots, q_d)$ on the alphabet Σ. A HWM $\mathcal{M} : \langle \mathbb{K}^d, \{\boldsymbol{\mathfrak{T}}^f\}_{f \in \mathcal{F}}, \odot_{id}, \mathbf{1} \rangle$ can be associated with $\mathcal{A}$ by letting

$$\boldsymbol{\mathfrak{T}}^x_i = \begin{cases} 1 & \text{if } q_i \in F_x \\ 0 & \text{otherwise} \end{cases} \text{ and } \boldsymbol{\mathfrak{T}}^\sigma_{i_1 i_2 i_3 i_4} = \begin{cases} \delta(r) & \text{if } r = (\sigma, q_{i_1}, q_{i_2}, q_{i_3}, q_{i_4}) \in R \\ 0 & \text{otherwise} \end{cases}$$

for any $x \in \{w, s, e, n\}$ and $\sigma \in \Sigma$.

One can show that $r_{\mathcal{M}}(G_p) = r_{\mathcal{A}}(p)$ for all pictures $p \in \Sigma^{++}$ (see [42, Proposition 4] for a detailed proof).

Expressiveness of HWMs. We showed that HWMs are a generalization of weighted automata over strings, trees, and pictures. The converse question naturally arises: e.g. restricted to the class of strings, can every HWM be simulated by a weighted string automaton? The answer is yes for the three cases (strings, trees and pictures). We show this result for strings in the following proposition (the case of trees and pictures can be treated similarly).

Proposition 3.9. *Any HWM-recognizable series on the family of strings on a finite alphabet Σ (as defined in Example 3) can be computed by a weighted string automaton.*

Proof. Let ι and τ be the two new symbols used to denote the beginning and end of the graph representation of a string (cf. Example 3). Let $M = \langle \mathbb{K}^d, \{\boldsymbol{\mathfrak{T}}^x\}_{x\in\Sigma\cup\{\iota,\tau\}}, \odot, \boldsymbol{\alpha}\rangle$ be a HWM and let $\mathbf{A}^\sigma = \boldsymbol{\mathfrak{T}}^\sigma$ for all $\sigma \in \Sigma$, $\boldsymbol{\iota} = \boldsymbol{\mathfrak{T}}^\iota$, and $\boldsymbol{\tau} = \boldsymbol{\mathfrak{T}}^\tau$. If $\odot = \odot_{id}$ and $\boldsymbol{\alpha} = \mathbf{1}$ then the proof of Proposition 3.5 shows that the WA $A = \langle \mathbb{K}^d, \boldsymbol{\iota}, \{\mathbf{A}^\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\tau}\rangle$ satisfies $r_A(w) = r_M(G_w)$ for all $w \in \Sigma^*$. If this is not the case (i.e. either $\odot \neq \odot_{id}$ or $\boldsymbol{\alpha} \neq \mathbf{1}$), let $\mathbf{O} \in \mathbb{R}^{d\times d}$ be the matrix defined by $\mathbf{O}_{i,j} = \boldsymbol{\alpha}^\top(\mathbf{e}_i \odot \mathbf{e}_j)$ for all $i, j \in [d]$ where $\mathbf{e}_1, \cdots, \mathbf{e}_d$ is the canonical basis of $\mathbb{R}^d$. We claim that the WA $\tilde{A} = \langle \mathbb{K}^d, \boldsymbol{\iota}, \{\mathbf{O}\mathbf{A}^\sigma\}_{\sigma\in\Sigma}, \mathbf{O}\boldsymbol{\tau}\rangle$ satisfies $r_{\tilde{A}}(w) = r_M(G_w)$ for all $w \in \Sigma^*$. Indeed, let $w \in \Sigma^*$ and $n = |w|$. We have

$$\begin{aligned}
r_M(G_w) &= \sum_{\gamma\in\Gamma} \boldsymbol{\mathfrak{T}}_\gamma \prod_{h\in E} \boldsymbol{\alpha}^\top \bigodot_{i\in\gamma(h)} \mathbf{e}_i \\
&= \sum_{i_0,\cdots,i_{2n+1}} \boldsymbol{\mathfrak{T}}^\iota_{i_0} \boldsymbol{\mathfrak{T}}^{w_1}_{i_1,i_2} \cdots \boldsymbol{\mathfrak{T}}^{w_n}_{i_{2n-1},i_{2n}} \boldsymbol{\mathfrak{T}}^\tau_{i_{2n+1}} \boldsymbol{\alpha}^\top(\mathbf{e}_{i_0} \odot \mathbf{e}_{i_1}) \cdots \boldsymbol{\alpha}^\top(\mathbf{e}_{i_{2n}} \odot \mathbf{e}_{i_{2n+1}}) \\
&= \sum_{i_0,\cdots,i_{2n+1}} \boldsymbol{\iota}_{i_0} \mathbf{A}^{w_1}_{i_1,i_2} \cdots \mathbf{A}^{w_n}_{i_{2n-1},i_{2n}} \boldsymbol{\tau}_{i_{2n+1}} \mathbf{O}_{i_0,i_1} \mathbf{O}_{i_1,i_2} \cdots \mathbf{O}_{i_{2n},i_{2n+1}} \\
&= \boldsymbol{\iota}^\tau \mathbf{O}\mathbf{A}^{w_1} \mathbf{O}\mathbf{A}^{w_2} \cdots \mathbf{O}\mathbf{A}^{w_n} \mathbf{O}\boldsymbol{\tau} = r_{\tilde{A}}(w).
\end{aligned}$$

□

3.4. Closure Properties

The following propositions show that the set of HWMs is closed under addition for HWMs defined over families of connected hypergraphs and closed under Hadamard product for HWMs defined over arbitrary families of hypergraphs.

Proposition 3.10. *Let $A = \langle \mathbb{K}^m, \{\boldsymbol{\mathcal{A}}^x\}_{x\in\Sigma}, \odot_A, \boldsymbol{\alpha}\rangle$ and $B = \langle \mathbb{K}^n, \{\boldsymbol{\mathcal{B}}^x\}_{x\in\Sigma}, \odot_B, \boldsymbol{\beta}\rangle$ be two HWMs. Let r_A (resp. r_B) be the series computed by A (resp. by B). Define the HWM $C = \langle \mathbb{K}^{m+n}, \{\boldsymbol{\mathfrak{C}}^x\}_{x\in\Sigma}, \odot, \boldsymbol{\tau}\rangle$ by*

- $\boldsymbol{\mathfrak{C}}^x_{i_1\dots i_{\sharp x}} = \begin{cases} \boldsymbol{\mathcal{A}}^x_{i_1\dots i_{\sharp x}} & \textit{if } 1 \leq i_1, \dots, i_{\sharp x} \leq m \\ \boldsymbol{\mathcal{B}}^x_{j_1\dots j_{\sharp x}} & \textit{if } m < i_1, \dots, i_{\sharp x} \leq m+n \textit{ where } j_k = i_k - m \textit{ for any } k \\ 0 & \textit{otherwise,} \end{cases}$
- $\boldsymbol{\tau}_i = \boldsymbol{\alpha}_i$ *if* $1 \leq i \leq m$ *and* $\boldsymbol{\beta}_{i-m}$ *otherwise, and* [4]

[4]For sake of readability we use the same notation $\mathbf{e}_i$ to denote the ith vector of the canonical basis of $\mathbb{K}^m$, $\mathbb{K}^n$, $\mathbb{K}^{m+n}$ and $\mathbb{K}^{mn}$ in this proposition and the following one; the actual vector space being referred to will always appear clearly from context.

- $\mathbf{e}_i \odot \mathbf{e}_j = \begin{cases} \mathbf{e}_i \odot_A \mathbf{e}_j & \text{if } 1 \leq i,j \leq m \\ t_m(\mathbf{e}_{i-m} \odot_B \mathbf{e}_{j-m}) & \text{if } m < i,j \leq n \\ 0 & \text{otherwise} \end{cases}$

where $t_m : \mathbb{K}^n \to \mathbb{K}^{m+n}$ *is the linear mapping defined by* $t_m(\mathbf{e}_k) = \mathbf{e}_{k+m}$ *for any* $1 \leq k \leq n$.

Then the HWM C computes the series r_{A+B} *defined by* $r_{A+B}(G) = r_A(G) + r_B(G)$*, for every connected hypergraph* G.

Proof. Let $G = (V, E, \ell)$, let P_G be the set of ports of G, let $\Gamma = [m+n]^{P_G}$, $\Gamma_1 = \{\gamma \in \Gamma : \gamma(P_G) \subseteq [m]\}$, and let $\Gamma_2 = \{\gamma \in \Gamma : \gamma(P_G) \subseteq \{m+1, \ldots, m+n\}\}$.

If $\gamma \notin \Gamma_1 \cup \Gamma_2$, then $\boldsymbol{\mathcal{C}}_\gamma \prod_{h \in E} \boldsymbol{\tau}^\top \bigodot_{i \in \gamma(h)} \mathbf{e}_i = 0$. Indeed, let $V_1 = \{v \in V : \exists v^{(i)} \in P_G \text{ s.t. } \gamma(v^{(i)}) \leq m\}$ and $V_2 = \{v \in V : \exists v^{(i)} \in P_G \text{ s.t. } \gamma(v^{(i)}) > m\}$. Note that V_1 and V_2 are not empty.

- If there exists $v \in V_1 \cap V_2$, then $\boldsymbol{\mathcal{C}}^v_{\gamma(v,1)\ldots\gamma(v,\sharp v)} = 0$ and therefore $\boldsymbol{\mathcal{C}}_\gamma = 0$
- If $V_1 \cap V_2 = \emptyset$, there exists a hyperedge h and ports $v_1^{(i)}, v_2^{(j)} \in h$ such that $v_1 \in V_1$ and $v_2 \in V_2$, since G is connected. Then, $\bigodot_{i \in \gamma(h)} \mathbf{e}_i = 0$.

Now,

$$\begin{aligned} r_C(G) &= \sum_{\gamma \in \Gamma} \boldsymbol{\mathcal{C}}_\gamma \prod_{h \in E} \boldsymbol{\tau}^\top \bigodot_{i \in \gamma(h)} \mathbf{e}_i \\ &= \sum_{\gamma \in \Gamma_1} \boldsymbol{\mathcal{C}}_\gamma \prod_{h \in E} \boldsymbol{\tau}^\top \bigodot_{i \in \gamma(h)} \mathbf{e}_i + \sum_{\gamma \in \Gamma_2} \boldsymbol{\mathcal{C}}_\gamma \prod_{h \in E} \boldsymbol{\tau}^\top \bigodot_{i \in \gamma(h)} \mathbf{e}_i \\ &= \sum_{\gamma \in \Gamma_A} \boldsymbol{\mathcal{A}}_\gamma \prod_{h \in E} \boldsymbol{\alpha}^\top \bigodot_{i \in \gamma(h)} \mathbf{e}_i + \sum_{\gamma \in \Gamma_B} \boldsymbol{\mathcal{B}}_\gamma \prod_{h \in E} \boldsymbol{\beta}^\top \bigodot_{i \in \gamma(h)} \mathbf{e}_i \end{aligned}$$

where $\Gamma_A = [m]^{P_G}$ and $\Gamma_B = [n]^{P_G}$. Eventually, $r_C(G) = r_A(G) + r_B(G)$. □

We will now exhibit a counter-example showing that the previous proposition only holds for connected hypergraphs. More precisely, we show that *there exist recognizable series on graphs whose sum is not recognizable.* We consider HMWs defined over closed graphs on the one letter alphabet $\mathcal{F} = (\{x\}, \sharp)$ where $\sharp x = 2$. Let G_1 be the circular string x (i.e. G_1 has one vertex labeled by x whose first and second ports are connected) and let G_2 be the graph with two connected components, each one isomorphic to G_1. Consider the two HWMs $A_1 = (\mathbb{C}^d, \{\mathbf{A}\}, \odot_{id}, \mathbf{1})$ and $A_2 = (\mathbb{C}^d, \{-\mathbf{A}\}, \odot_{id}, \mathbf{1})$, where $\mathbf{A} \in \mathbb{C}^{d\times d}$ is such that $\mathrm{Tr}(\mathbf{A}) \neq 0$, and let r_1 and r_2 be the series computed by A_1 and A_2 respectively. One can check that

$$r_1(G_1) = \mathrm{Tr}(\mathbf{A}) = -r_2(G_1) \quad \text{and} \quad r_1(G_2) = \mathrm{Tr}(\mathbf{A})^2 = r_2(G_2).$$

Suppose now that the series $r_1 + r_2$ is recognizable. Then, by Proposition 3.4 there would exist a HWM $C = (\mathbb{C}^n, \{\mathbf{C}\}, \odot_{id}, \mathbf{1})$ computing $r_1 + r_2$, whence $\mathrm{Tr}(\mathbf{C}) = r_C(G_1) = r_1(G_1) + r_2(G_1) = 0$ and $\mathrm{Tr}(\mathbf{C})^2 = r_C(G_2) = r_1(G_2) + r_2(G_2) = 2\,\mathrm{Tr}(\mathbf{A})^2 \neq 0$, a contradiction.

Proposition 3.11. *Let $A = \langle \mathbb{K}^m, \{\boldsymbol{\mathcal{A}}^x\}_{x\in\Sigma}, \odot^A, \boldsymbol{\alpha}\rangle$ and $B = \langle \mathbb{K}^n, \{\boldsymbol{\mathcal{B}}^x\}_{x\in\Sigma}, \odot^B, \boldsymbol{\beta}\rangle$ be two HWMs.*

Identifying $\mathbb{K}^m \otimes \mathbb{K}^n$ with $\mathbb{K}^{mn}$ via the mapping $\mathbf{e}_i \otimes \mathbf{e}_j \mapsto \mathbf{e}_{n(i-1)+j}$, we define the HWM $D = \langle \mathbb{K}^m \otimes \mathbb{K}^n, \{\boldsymbol{\mathfrak{D}}^x\}_{x\in\Sigma}, \odot, \boldsymbol{\delta}\rangle$ by

- *$\boldsymbol{\mathfrak{D}}^x = \boldsymbol{\mathcal{A}}^x \otimes \boldsymbol{\mathfrak{B}}^x$ for all $x \in \Sigma$,*
- *$(\mathbf{a}_1 \otimes \mathbf{b}_1) \odot (\mathbf{a}_2 \otimes \mathbf{b}_2) = (\mathbf{a}_1 \odot^A \mathbf{a}_2) \otimes (\mathbf{b}_1 \odot^B \mathbf{b}_2)$ for all $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{K}^m$ and $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{K}^n$,*
- *$\boldsymbol{\delta} = \boldsymbol{\alpha}\otimes\boldsymbol{\beta}$ (i.e. $\boldsymbol{\delta}^\top(\mathbf{a}\otimes\mathbf{b}) = (\boldsymbol{\alpha}^\top\mathbf{a})(\boldsymbol{\beta}^\top\mathbf{b})$ for any $\mathbf{a} \in \mathbb{K}^m$ and $\mathbf{b} \in \mathbb{K}^m$).*

Let r_A (resp. r_B) be the series computed by A (resp. by B). Then the HWM C computes the series r_C given by $r_C(G) = r_A(G)r_B(G)$, for every hypergraph G.

Proof. Let $G = (V, E, \ell)$ be a hypergraph and let $\Gamma_k = [k]^{P_G}$ for any integer k. We will identify $[m] \times [n]$ with $[mn]$ via the mapping $(i, j) \mapsto n(i-1)+j$ and by extension we identify Γ_{mn} with $\Gamma_{m\times n} = ([m] \times [n])^{P_G}$. For any $\gamma \in \Gamma_{m\times n}$ we will denote by (γ_1, γ_2) the only element of $\Gamma_m \times \Gamma_n$ satisfying $\gamma(\cdot) = (\gamma_1(\cdot), \gamma_2(\cdot))$.

First note that for any hyperedge $h \in E$ and $\gamma \in \Gamma_{m\times n}$, we have

$$\begin{aligned}\boldsymbol{\delta}^\top \bigodot_{(i,j)\in\gamma(h)} (\mathbf{e}_i \otimes \mathbf{e}_j) &= \boldsymbol{\delta}^\top \left[\left(\overset{A}{\bigodot_{i\in\gamma_1(h)}} \mathbf{e}_i\right) \otimes \left(\overset{B}{\bigodot_{j\in\gamma_2(h)}} \mathbf{e}_j\right)\right] \\ &= \left(\boldsymbol{\alpha}^\top \overset{A}{\bigodot_{i\in\gamma_1(h)}} \mathbf{e}_i\right)\left(\boldsymbol{\beta}^\top \overset{B}{\bigodot_{j\in\gamma_2(h)}} \mathbf{e}_j\right).\end{aligned}$$

Then, note that for any $\gamma \in \Gamma_{m\times n}$, we have

$$\boldsymbol{\mathcal{D}}_\gamma = \prod_{v\in V} \boldsymbol{\mathcal{D}}^{\ell(v)}_{\gamma(v,1)\cdots\gamma(v,\sharp v)} = \prod_{v\in V} \boldsymbol{\mathcal{A}}^{\ell(v)}_{\gamma_1(v,1),\cdots,\gamma_1(v,\sharp v)} \boldsymbol{\mathcal{B}}^{\ell(v)}_{\gamma_2(v,1),\cdots,\gamma_2(v,\sharp v)} = \boldsymbol{\mathcal{A}}_{\gamma_1}\boldsymbol{\mathcal{B}}_{\gamma_2}$$

Finally, we have

$$\begin{aligned} r_D(G) &= \sum_{\gamma\in\Gamma} \boldsymbol{\mathcal{D}}_\gamma \prod_{h\in E} \boldsymbol{\delta}^\top \bigodot_{(i,j)\in\gamma(h)} (\mathbf{e}_i \otimes \mathbf{e}_j) \\ &= \sum_{\gamma_1\in\Gamma_m}\sum_{\gamma_2\in\Gamma_n} \boldsymbol{\mathcal{A}}_{\gamma_1}\boldsymbol{\mathcal{B}}_{\gamma_2} \prod_{h\in E} \left(\boldsymbol{\alpha}^\top \bigodot^{A}_{i\in\gamma_1(h)} \mathbf{e}_i\right)\left(\boldsymbol{\beta}^\top \bigodot^{B}_{j\in\gamma_2(h)} \mathbf{e}_j\right) \\ &= r_A(G) r_B(G). \end{aligned}$$

□

Interestingly, the set of HWM-recognizable series is not closed under scalar multiplication in general. To see this, consider the HWM-recognizable series on circular strings on the alphabet $\Sigma = \{a(\cdot,\cdot)\}$ defined by $r(a^n) = \kappa^n$ for some $\kappa \in \mathbb{R}\setminus\{0,1\}$. Then, for any real number $\alpha \notin \mathbb{N}$ the series $r' : a^n \mapsto \alpha r(a^n)$ is not HWM-recognizable. Indeed, this would imply that there exists some matrix $\mathbf{M}$ such that $\mathrm{Tr}(\mathbf{M}^n) = \alpha\kappa^n$ for all n, but it follows from the following lemma that if $\mathrm{Tr}((\frac{\mathbf{M}}{\kappa})^n) = \alpha$ for all n, then $\alpha \in \mathbb{N}$.

Lemma 3.12. *Let $\mathbf{M} \in \mathbb{R}^{n\times n}$. If there exists an integer k such that $\mathrm{Tr}(\mathbf{M}^k) = \cdots = \mathrm{Tr}(\mathbf{M}^{k+n})$, then $\forall m \in \mathbb{N}, \mathrm{Tr}(\mathbf{M}^m) = \mathrm{Tr}(\mathbf{M}) \in \mathbb{N}$.*

Proof. Let $\lambda_1,\dots,\lambda_p \in \mathbb{C}$ be the distinct non zero eigenvalues of $\mathbf{M}$, with multiplicities $n_1,\dots,n_p$ respectively. If $p = 0$, the spectrum of $\mathbf{M}$ is reduced to 0 and $\forall m \in \mathbb{N}, \mathrm{Tr}(\mathbf{M}^m) = \mathrm{Tr}(\mathbf{M}) = 0$.

Suppose that $p > 0$. Let $\mathbf{N} \in \mathbb{R}^{p\times p}$ be the square matrix defined by $\mathbf{N}[i,j] = \lambda_j^{i-1}$. The matrix $\mathbf{N}$ is full rank and its determinant is equal to $\prod_{i<j}(\lambda_j - \lambda_i)$. For any $k \in \mathbb{N}$, let $\mathbf{u}_k = (\lambda_1^k n_1,\dots,\lambda_p^k n_p)^\top$. We have $\mathbf{N}\mathbf{u}_k = (\mathrm{Tr}(\mathbf{M}^k),\dots,\mathrm{Tr}(\mathbf{M})^{k+p-1})$. Now, suppose that there exists an integer k such that $\mathrm{Tr}(\mathbf{M}^k) = \cdots = \mathrm{Tr}(\mathbf{M})^{k+p} = \alpha$. Then, $\mathbf{N}\mathbf{u}_k = \mathbf{N}\mathbf{u}_{k+1}$ and since $\mathbf{N}$ is invertible, $\mathbf{u}_k = \mathbf{u}_{k+1}$. Hence, $\lambda_1 = \cdots = \lambda_p = 1$ and $p = 1$. Therefore, $\forall m \in \mathbb{N}, \mathrm{Tr}(\mathbf{M}^m) = \mathrm{Tr}(\mathbf{M}) = n_1$.

□

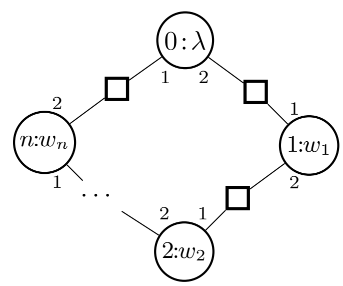


Figure 6: A rooted circular string.

Nonetheless, there are several families of hypergraphs on which HWMs are closed under scalar multiplication. An example of such a family is the family of *rooted hypergraphs* (hypergraphs on a ranked alphabet $(\Sigma \cup \{\lambda\}, \sharp)$ where the special *root* symbol λ appears exactly once): scalar multiplication by a real number γ is achieved by multiplying the tensor $\mathfrak{T}^{\lambda}$ of the original HWM by γ. The same idea applies to any family of hypergraphs with a fixed number k of edges where scalar multiplication by a real number γ is achieved by multiplying the vector $\boldsymbol{\alpha}$ of the original HWM by $\gamma^{1/k}$.

4. Examples

In this section, we present some examples of recognizable hypergraph series in order to give some insight on the expressiveness of HWMs and on how their computation relates to the usual notion of recognizable series on strings and trees.

4.1. Rooted Circular Strings

Instead of the construction described in Example 3, we can map each string w on a finite alphabet Σ to a *rooted circular string*. Let $w = w_1 \cdots w_n \in \Sigma^*$, we will consider the circular string G_w on the ranked alphabet $(\Sigma' = \Sigma \cup \{\lambda\}, \sharp)$ where λ is a new symbol, $\sharp x = 2$ for any $x \in \Sigma \cup \{\lambda\}$. G_w has vertices $V = \{0, \cdots, n\}$, labels $\ell(0) = \lambda$ and $\ell(i) = w_i$ for $i \in [n]$, and edges $\{(n, 2), (0, 1)\}$ and $\{(i, 2), (i + 1, 1)\}$ for $i \in \{0, \cdots, n - 1\}$ (see Figure 6).

Let $r = \langle \mathbb{K}^d, \boldsymbol{\iota}, \{\mathbf{M}_\sigma\}_{\sigma \in \Sigma}, \boldsymbol{\tau} \rangle$ be a rational series on Σ^*. We define the HWM $A = \langle \mathbb{K}^d, \{\boldsymbol{\mathcal{A}}^x\}_{x \in \Sigma'}, \odot_{id}, \mathbf{1} \rangle$ where $\boldsymbol{\mathcal{A}}^\sigma = \mathbf{M}_\sigma$ for all $\sigma \in \Sigma$ and $\boldsymbol{\mathcal{A}}^\lambda =$

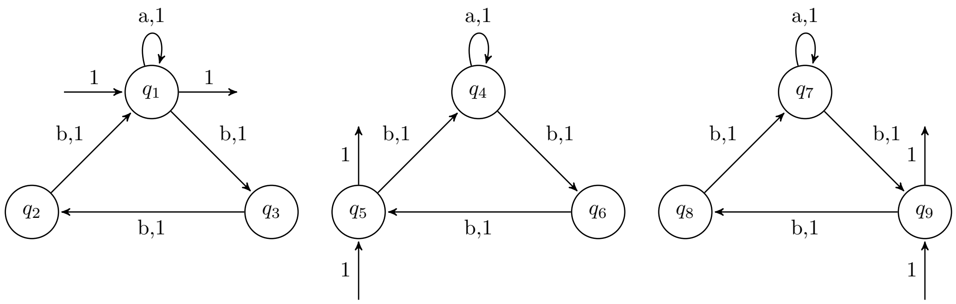


Figure 7: A minimal weighted automaton with 9 states computing a series that can be computed by a 3-dimensional HWM on rooted circular strings.

$\boldsymbol{\tau}\boldsymbol{\iota}^\top$. It is easy to check that $r_A(G_w) = r(w)$ for every word $w \in \Sigma^*$. Indeed, it follows from the fact that the trace operator is invariant under cyclic permutations that $r(w) = \boldsymbol{\iota}^\top \mathbf{M}_w \boldsymbol{\tau} = \mathrm{Tr}(\boldsymbol{\iota}^\top \mathbf{M}_w \boldsymbol{\tau}) = \mathrm{Tr}(\boldsymbol{\mathcal{A}}^\lambda \mathbf{M}_w) = r_A(G_w)$.

Now consider m rational string series $r_1, \dots, r_m$ whose linear representations $\langle \mathbb{K}^d, \boldsymbol{\iota}_i, \{\mathbf{M}_\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\tau}_i \rangle$ for $i \in [m]$ share the same transition matrices and let $r = r_1 + \cdots + r_m$. It can easily be checked that $r(w) = \mathrm{Tr}(\boldsymbol{\mathcal{A}}^\lambda \mathbf{M}_w)$, where $\boldsymbol{\mathcal{A}}^\lambda = \sum_{i=1}^m \boldsymbol{\tau}_i \boldsymbol{\iota}_i^\top$. Hence, the d-dimensional HWM $A = \langle \mathbb{K}^d, \{\boldsymbol{\mathcal{A}}^x\}_{x\in\Sigma'}, \odot_{id}, \mathbf{1}\rangle$, where $\boldsymbol{\mathcal{A}}^\sigma = \mathbf{M}_\sigma$ for all $\sigma \in \Sigma$, is such that $r_A(G_w) = r(w)$ for all $w \in \Sigma^*$.

It can easily be shown, by decomposing $\boldsymbol{\mathcal{A}}^\lambda$ as a sum of at most d rank-one matrices, that the rank of r is at most d^2 (while A is d-dimensional). The following proposition shows that this upper bound can be achieved (see also Figure 7).

Proposition 4.1. *There exists a recognizable string series of rank d^2 that can be computed by a d-dimensional HWM on rooted circular strings.*

Proof. Let $(\mathbf{E}^{i,j})_{1\leq i,j\leq n}$ be the canonical basis of $\mathbb{R}^{d\times d}$ (i.e. $\mathbf{E}^{i,j} = \mathbf{e}_i \mathbf{e}_j^\top$). Let $\Sigma = \{a, b\}$ and let $\mathbf{M}_a, \mathbf{M}_b \in \mathbb{R}^{d\times d}$ be defined by

$$\mathbf{M}_a = \mathbf{E}_{1,1} \text{ and } \mathbf{M}_b = \mathbf{E}_{2,1} + \mathbf{E}_{3,2} + \cdots + \mathbf{E}_{1,d}$$

where $\mathbf{E}_{i,j} \in \mathbb{R}^{d\times d}$ satisfies $\mathbf{E}_{i,j}[k,l] = \delta_{ik}\delta_{jl}$ for every $1 \leq i,j,k,l \leq d$. For any $i \in [d]$, let r_i be the recognizable series with linear representation $\langle \mathbb{R}^d, \boldsymbol{e}_i, \{\mathbf{M}_a, \mathbf{M}_b\}, \boldsymbol{e}_i \rangle$ and let $r = r_1 + \cdots + r_d$. We have $r(w) = \mathrm{Tr}(\mathbf{M}_w)$ for any $w \in \Sigma^*$, hence r can be computed by a d-dimensional HWM on rooted circular strings (with $\boldsymbol{\mathcal{A}}^\lambda = \mathbf{I}$). Moreover, the rank of the string series r is d^2. Indeed, it can be shown that the $\mathbb{R}$-algebra spanned by $\mathbf{M}_a$ and $\mathbf{M}_b$ is equal to $\mathbb{R}^{d\times d}$. Let $w_1,\dots,w_{d^2} \in \Sigma^*$ be such that the matrices $\mathbf{M}_{w_1},\dots,\mathbf{M}_{w_{d^2}}$ are linearly independent and let $\mathbf{H} \in \mathbb{R}^{d^2\times d^2}$ be the so-called Hankel matrix defined by $\mathbf{H}[i,j] = \mathrm{Tr}(\mathbf{M}_{w_i w_j})$. The rank of $\mathbf{H}$ is d^2 since

$$
\begin{aligned}
\forall j \in [d^2], \sum_{i=1}^{d^2} \alpha_i \,\mathrm{Tr}(\mathbf{M}_{w_i w_j}) = 0 &\Leftrightarrow \forall k,l \in [d], \sum_{i=1}^{d^2} \alpha_i \,\mathrm{Tr}(\mathbf{M}_{w_i}\mathbf{E}_{k,l}) = 0 \\
&\Leftrightarrow \forall k,l \in [d], \sum_{i=1}^{d^2} \alpha_i \mathbf{M}_{w_i}[l,k] = 0 \Leftrightarrow \sum_{i=1}^{d^2} \alpha_i \mathbf{M}_{w_i} = 0 \\
&\Leftrightarrow \alpha_1 = \cdots = \alpha_{d^2} = 0
\end{aligned}
$$

where we used the fact that both $(\mathbf{E}_{k,l})_{k,l}$ and $(\mathbf{M}_{w_j})_j$ are basis of $\mathbb{R}^{d\times d}$ for the first equivalence. From a fundamental theorem on recognizable string series, the rank of $\mathbf{H}$ is a lower bound of the rank of r, which entails the result. □

4.2. Crosswords on Pictures

In this section, we show that if r_1 and r_2 are two recognizable series on Σ^*, the series on 2D-words computing the product of the values of r_1 applied to each row and r_2 applied to each column is HWM-recognizable.

Instead of the construction presented in Section 3.3, we associate here a picture in Σ^{++} to a graph where the borders are left with free ports. Formally, let $(\Sigma, \sharp)$ be the ranked alphabet where all symbols in Σ have arity 4. The graph $G_w = (V, E, \ell)$ associated to a 2D-word $w \in \Sigma^{M\times N}$ is the graph with vertices $V = [M]\times[N]$, $\ell(m,n) = w_{mn}$, and edges $E = E_H \cup E_V$, where the ports are labeled by W, E, N, S, in this order, and where

$$E_H = \bigcup_{m\in[M],n\in[N-1]} \left\{\{(m,n)^E, (m,n+1)^W\}\right\} \bigcup_{m\in[M]} \left\{\{(m,1)^W\},\{(m,N)^E\}\right\}$$

and

$$E_V = \bigcup_{n\in[N],m\in[M-1]} \left\{\{(m,n)^S, (m+1,n)^N\}\right\} \bigcup_{n\in[N]} \left\{\{(1,n)^N\},\{(M,n)^S\}\right\}.$$

An example of such a graph is shown in Figure 8. Now, let $\#_1$ be a new

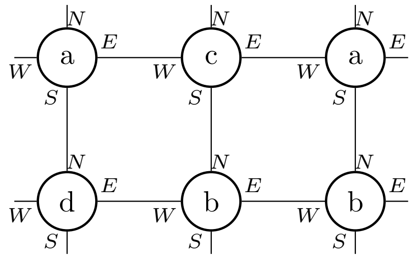


Figure 8: Graph associated to the 2D-word $\substack{aca\\dbb}$.

arity function, such that all symbols of Σ have arity 2. The graph G_w can be decomposed in two graphs over $(\Sigma, \#_1)$: $G_w^H = (V, E_H, \ell)$ where the ports are labeled W, E in this order and $G_w^V = (V, E_V, \ell)$ where the ports are labeled N, S in this order.

Proposition 4.2. *Given* $\mathcal{A} = \langle \mathbb{K}^{d_1}, \{\boldsymbol{\mathcal{A}}^\sigma\}_{\sigma\in\Sigma}, \odot_1, \boldsymbol{\beta}_1\rangle$ *and* $\mathcal{B} = \langle \mathbb{K}^{d_2}, \{\boldsymbol{\mathcal{B}}^\sigma\}_{\sigma\in\Sigma}, \odot_2, \boldsymbol{\beta}_2\rangle$, *two HWMs over* $(\Sigma, \#_1)$, *there exists a HWM* $\mathcal{C} = \langle \mathbb{K}^{d_1+d_2}, \{\boldsymbol{\mathfrak{C}}^\sigma\}_{\sigma\in\Sigma}, \odot, \boldsymbol{\beta}\rangle$ *over* $(\Sigma, \sharp)$ *such that*

$$r_{\mathcal{C}}(G_w) = r_{\mathcal{A}}(G_w^H) \times r_{\mathcal{B}}(G_w^V)$$

for any (M, N)*-crossword* w*.*

Proof. Let $(\mathbf{e}_1, \ldots, \mathbf{e}_{d_1}, \mathbf{f}_1, \ldots, \mathbf{f}_{d_2})$ be the canonical basis of $\mathbb{K}^{d_1+d_2}$. The HWM C is defined by

- $\boldsymbol{\mathfrak{C}}^\sigma = \boldsymbol{\mathcal{A}}^\sigma \otimes \boldsymbol{\mathcal{B}}^\sigma = \left(\sum_{i_1,\ldots,i_{\sharp\sigma}\in[d_1]} \boldsymbol{\mathcal{A}}_{i_1,\ldots,i_{\sharp\sigma}} \mathbf{e}_{i_1} \otimes \ldots \mathbf{e}_{i_{\sharp\sigma}}\right) \otimes \left(\sum_{i_1,\ldots,i_{\sharp\sigma}\in[d_2]} \boldsymbol{\mathcal{B}}_{i_1,\ldots,i_{\sharp\sigma}} \mathbf{f}_{i_1} \otimes \ldots \mathbf{f}_{i_{\sharp\sigma}}\right)$ for any $\sigma \in \Sigma$
- $\mathbf{e}_i \odot \mathbf{e}_j = \mathbf{e}_i \odot_1 \mathbf{e}_j, \mathbf{f}_i \odot \mathbf{f}_j = \mathbf{f}_i \odot_2 \mathbf{f}_j$, and $\mathbf{e}_i \odot \mathbf{f}_j = \mathbf{f}_j \odot \mathbf{e}_i = 0$ for any indices i, j
- $\boldsymbol{\beta} = \begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}$.

By definition, we have

$$r_{\mathcal{C}}(G_w) = \sum_{\gamma\in\Gamma} \boldsymbol{\mathfrak{C}}_\gamma \prod_{h\in E} \boldsymbol{\beta}^\top \bigodot_{i\in\gamma(h)} \mathbf{g}_i$$

where $\mathbf{g}_i = \mathbf{e}_i$ if $i \in [d_1]$ and $\mathbf{f}_i$ otherwise. Let $\Gamma_H = [d_1]^{P_{G_w^H}}$ and $\Gamma_V = [d_2]^{P_{G_w^V}}$. It is easy to check that any $\gamma \in \Gamma$ for which $\boldsymbol{\mathfrak{C}}_\gamma \neq 0$ can be associated with a tuple $(\gamma_H, \gamma_V) \in \Gamma_H \times \Gamma_V$ satisfying $\boldsymbol{\mathfrak{C}}_\gamma = \boldsymbol{\mathcal{A}}_{\gamma_H} \boldsymbol{\mathcal{B}}_{\gamma_V}$, we have

$$\begin{aligned} r_{\mathcal{C}}(G_w) &= \sum_{\gamma_H \in \Gamma_H} \sum_{\gamma_V \in \Gamma_V} \boldsymbol{\mathcal{A}}_{\gamma_H} \boldsymbol{\mathcal{B}}_{\gamma_V} \prod_{h \in E_H} \boldsymbol{\beta}_1^\top \left(\bigodot_{i \in \gamma_H(h)}^{1} \mathbf{e}_i \right) \times \prod_{h \in E_V} \boldsymbol{\beta}_2^\top \left(\bigodot_{i \in \gamma_V(h)}^{2} \mathbf{f}_i \right) \\ &= r_{\mathcal{A}}(G_w^H) \times r_{\mathcal{B}}(G_w^V). \end{aligned}$$

□

Given a (M, N)-crossword w, we denote by $w_{m:}$ the m-th row of w and by $w_{:n}$ its n-th column.

Corollary 4.3. *Let $\mathcal{A} = \langle \mathbb{R}^{d_1}, \{\mathbf{A}^\sigma\}_{\sigma \in \Sigma}, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_\infty \rangle$ and $\mathcal{B} = \langle \mathbb{R}^{d_2}, \{\mathbf{B}^\sigma\}_{\sigma \in \Sigma}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_\infty \rangle$ be two rational string series on Σ^*.*

There exists a HWM $\mathcal{C} = \langle \mathbb{C}^{d_1+d_2}, \{\boldsymbol{\mathfrak{C}}^\sigma\}_{\sigma \in \Sigma}, \odot, \boldsymbol{\gamma} \rangle$ such that

$$r_{\mathcal{C}}(G_w) = \prod_{m \in [M]} r_{\mathcal{A}}(w_{m:}) \prod_{n \in [N]} r_{\mathcal{B}}(w_{:n})$$

for any (M, N)-crossword w.

Proof. The result directly follows from Proposition 3.6 and Proposition 4.2 and from remarking that the HWM $\mathcal{M}_H$ (resp. $\mathcal{M}_V$) that computes $r_{\mathcal{A}}$ (resp. $r_{\mathcal{B}}$) satisfies $\mathcal{M}_H(G_w^H) = \prod_{m \in [M]} r_{\mathcal{A}}(w_{m:})$ (resp. $\mathcal{M}_V(G_w^V) = \prod_{n \in [N]} r_{\mathcal{B}}(w_{:n})$) since the graph G_w^H (resp. G_w^V) has M (resp. N) connected components. □

5. Recognizability of Finite Support Series

In this section, we show that finite support series (or *polynomials*: series for which the set of hypergraphs with non-zero value is finite) are not recognizable in general, but we exhibit a wide class of families of hypergraphs for which they are. In this section, *we only consider HWM defined over families of connected graphs.*

First, we show on a simple example why polynomials are not recognizable for all families of hypergraphs. Consider the family of circular strings over a one letter alphabet $\Sigma = \{a\}$ introduced in Definition 2.3 and Remark 1. The series r, defined by $r(G_a) = 1$ and $r(G_{a^k}) = 0$ for all integer $k > 1$ is not recognizable. Indeed, r would be such that $r(G_{a^k}) = \mathrm{Tr}(\mathbf{M}_a^k) = 0$ for all $k \geq 2$, but it then follows from Lemma 3.12 that $r(G_a) = \mathrm{Tr}(\mathbf{M}_a) = 0$.

This example illustrates the fact that the computation of a HWM on a hypergraph G is done independently on each hyperedge of G. This implies that if two hypergraphs are not distinguishable by just looking at the ports involved in their hyperedges, the computations of a HWM on these two hypergraphs are strongly dependent. This is clear if we consider a hypergraph G' made of two copies of a hypergraph G (i.e. G' has two connected components G_1 and G_2, which are both isomorphic to G): we have $r(G') = r(G)^2$ for any HWM r (see Remark 3).

5.1. Hypergraphs Coverings

In this section, we formally introduces the notion of *covering* of a hypergraph G (which generalizes the notion of *graph covering* (or *lift*) to our definition of hypergraphs [43, 1]) and show how this relation between hypergraphs relates to the question of the recognizability of polynomials.

Intuitively, a *covering* of a hypergraph $\widehat{G}$ is a hypergraph G built on the same alphabet which is made of copies of $\widehat{G}$ (see Figure 9). More precisely,

Definition 5.1. *Let $\widehat{G} = (\widehat{V}, \widehat{E}, \widehat{\ell})$ be a hypergraph over a ranked alphabet $(\Sigma, \sharp)$. A hypergraph $G = (V, E, \ell)$ on the same alphabet $(\Sigma, \sharp)$ is a* covering *of $\widehat{G}$ if and only if there exists a mapping $f : V \to \widehat{V}$ such that*

(i) $\ell(v) = \widehat{\ell}(f(v))$ for any $v \in V$

(ii) the mapping $g : P_G \to P_{\widehat{G}}$ defined by $g(v, i) = (f(v), i)$ is such that for all $h \in E$: $g(h) \in \widehat{E}$ and the restriction $g_{|h}$ of g to h is bijective.

We call such a mapping f a covering map *from G to $\widehat{G}$.*

The following proposition shows that for a connected hypergraph, this formal definition of covering is equivalent to the intuition of a hypergraph made of copies of the original one.

Let $G = (V, E, \ell)$ be a covering of a connected hypergraph $\widehat{G} = (\widehat{V}, \widehat{E}, \widehat{\ell})$, let $\sim_{f,V}$ be the equivalence relation defined on V by $v \sim_{f,V} v'$ iff $f(v) = f(v')$, and let $\sim_{f,E}$ be the equivalence relation defined on E by $h \sim_{f,E} h'$ iff $g(h) = g(h')$ where f and g are the mappings defined above. Clearly, $v \sim_{f,V} v'$ entails that $\ell(v) = \ell(v')$ and it can easily be shown that $h \sim_{f,E} h'$ iff $\exists v^{(i)} \in h, v'^{(i)} \in h'$ such that $v \sim_{f,V} v'$. We can thus define the quotient hypergraph $\overline{G} = (V/\sim_{f,V}, E/\sim_{f,E}, \ell)$.

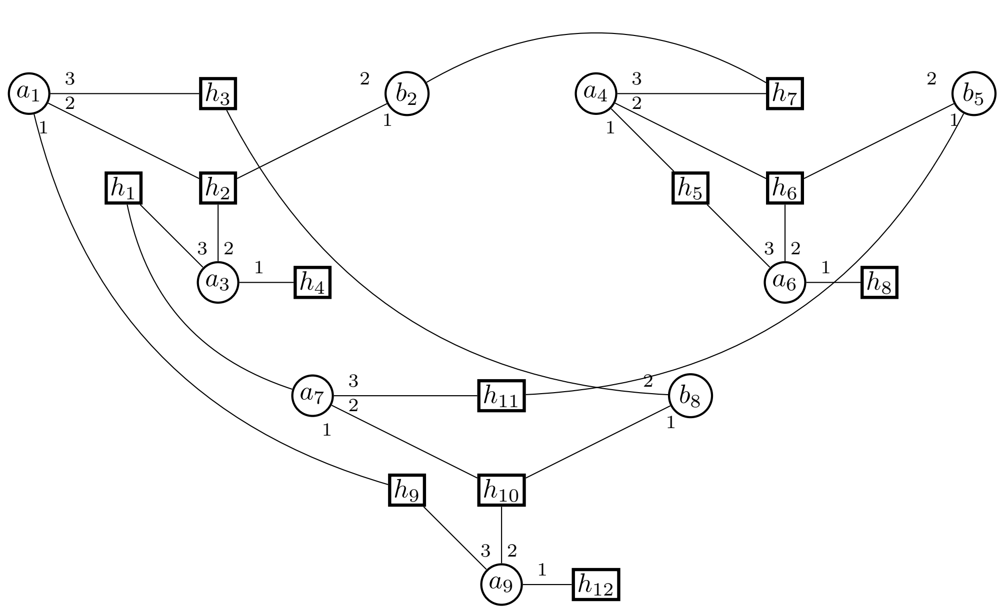


Figure 9: A covering made of three copies of the hypergraph from Example 2

Proposition 5.2. *If $G = (V, E, \ell)$ is a covering of a connected hypergraph $\widehat{G} = (\widehat{V}, \widehat{E}, \widehat{\ell})$, then $\overline{G} = (V/\sim_{f,V}, E/\sim_{f,E}, \ell)$ is isomorphic to $\widehat{G}$ and moreover, there exists a constant k such that $|f^{-1}(\{\hat{v}\})| = k$ for any $\hat{v} \in \widehat{V}$.*

Proof. We will prove the last part of the proposition, which entails the surjectivity of f. This will be enough since if f is surjective, then $\overline{G}$ is isomorphic to $\widehat{G}$.

Let m be the maximal cardinality of the sets $f^{-1}(\{\widehat{v}\})$ and suppose that they have different cardinalities. Let $V_1 = \{\widehat{v} \in \widehat{V} : Card(f^{-1}(\{\widehat{v}\})) = m\}$ and $V_2 = \widehat{V} \setminus V_1$. Since $\widehat{G}$ is connected, there exists a hyperedge $\widehat{h}$ and $\widehat{v}_1^{(i)}, \widehat{v}_2^{(j)} \in \widehat{h}$ such that $\widehat{v}_1 \in V_1$ and $\widehat{v}_2 \in V_2$. Let $f^{-1}(\{\widehat{v}_1\}) = \{v_1, \ldots, v_m\}$ and let $h_1, \ldots, h_m \in E$ be the hyperedges containing $v_1^{(i)}, \ldots, v_m^{(i)}$, respectively. Since each $g_{|h_i}$ is injective and since the vertices $v_1, \ldots, v_m$ are distinct, the hyperedges $h_1, \ldots, h_m$ are also distinct and therefore disjoint. Let $w_1^{(j)} = g_{|h_1}^{-1}(\widehat{v}_2^{(j)}), \ldots, w_m^{(j)} = g_{|h_m}^{-1}(\widehat{v}_2^{(j)})$. These ports are distinct and therefore, the vertices $w_1, \ldots, w_m$ are also distinct. Since, $f(w_1) = \cdots = f(w_m) = \widehat{v}_2$, we obtain a contradiction. □

5.2. Finite Support Series and Coverings

In the following theorem, we show that we can construct a HWM which assigns a nonzero value to a specific hypergraph over some ranked alphabet and all of its coverings, and zero to any other hypergraph on the same alphabet. This result leads to a sufficient condition on families of hypergraphs for the recognizability of finite support series.

Theorem 5.3. *Given a connected hypergraph $\widehat{G} = (\widehat{V}, \widehat{E}, \widehat{\ell})$ over $(\Sigma, \sharp)$, there exists a recognizable series $r_{\widehat{G}}$ such that for all connected hypergraphs G, $r_{\widehat{G}}(G) \neq 0$ if and only if G is a covering of $\widehat{G}$.*

Proof. Let $P_{\widehat{G}}$ be the set of ports of $\widehat{G}$. For any symbol $x \in \Sigma$, we denote by $\widehat{V}(x)$ the set of vertices in $\widehat{V}$ labelled by x.

Let $\mathcal{S} = 2^{P_{\widehat{G}}}$ be the set of subsets of $P_{\widehat{G}}$ and let $d = |\mathcal{S}|$. Instead of indexing the canonical basis of $\mathbb{K}^d$ with integers in $[d]$, we will index it with elements of $\mathcal{S}$. For example, for each port $(\hat{v}, i) \in P_{\widehat{G}}$, the singleton $\{(\hat{v}, i)\}$ is in $\mathcal{S}$, thus $\mathbf{e}_{\{(\hat{v},i)\}}$ is a basis vector (which we will note $\mathbf{e}_{(\hat{v},i)}$ for convenience).

Define the HWM $M = \langle \mathbb{K}^d, \{\boldsymbol{\mathcal{T}}^x\}_{x\in\Sigma}, \odot, \boldsymbol{\alpha} \rangle$ by

$$\boldsymbol{\mathcal{T}}^x = \begin{cases} \mathbf{e}_{\emptyset}{}^{\otimes \sharp x} & \text{if } \widehat{V}(x) = \emptyset \\ \sum_{\hat{v}\in\widehat{V}(x)} \mathbf{e}_{(\hat{v},1)} \otimes \cdots \otimes \mathbf{e}_{(\hat{v},\sharp\hat{v})} & \text{otherwise} \end{cases}$$

$$\mathbf{e}_S \odot \mathbf{e}_T = \begin{cases} \mathbf{e}_{S\cup T} & \text{if } S \neq \emptyset, T \neq \emptyset \text{ and } S \cap T = \emptyset \\ \mathbf{e}_{\emptyset} & \text{otherwise} \end{cases}$$

$$\boldsymbol{\alpha}_S = \begin{cases} 1 & \text{if } S \in \widehat{E} \quad (\text{note that } \emptyset \notin \widehat{E}) \\ 0 & \text{otherwise} \end{cases}$$

for any $x \in \Sigma$ and $S, T \in \mathcal{S}$. Let r be the series computed by M, we claim that r satisfies the property of the theorem.

For any hypergraph $G = (V, E, \ell)$ with $V = \{v_1, \cdots, v_N\}$, we have

$$r(G) = \sum_{\gamma\in\Gamma} \boldsymbol{\mathcal{T}}_\gamma \prod_{h\in E} \boldsymbol{\alpha}^\top \bigodot_{S\in\gamma(h)} e_S \tag{2}$$

where $\Gamma = \mathcal{S}^{P_G}$ and $\boldsymbol{\mathcal{T}}_\gamma = \prod_{i=1}^N \boldsymbol{\mathcal{T}}^{v_i}_{\gamma(v_i,1),\cdots,\gamma(v_i,\sharp v_i)}$.

Let $\gamma \in \Gamma$. It follows from the definition of the tensors $\boldsymbol{\mathcal{T}}^x$ that $\boldsymbol{\mathcal{T}}_\gamma$ is different from 0 if and only if for all $v \in V$, there exists $\widehat{v} \in \widehat{V}(\ell(v))$ such that $\gamma(v,i) = (\widehat{v},i)$ for all $i \in [\sharp v]$. We can thus rewrite Eq. (2) as

$$r(G) = \sum_{\widehat{v}_1 \in \widehat{V}(\ell(v_1))} \cdots \sum_{\widehat{v}_N \in \widehat{V}(\ell(v_N))} \prod_{i=1}^{N} \boldsymbol{\mathcal{T}}^{v_i}_{(\widehat{v}_i,1),\cdots,(\widehat{v}_i,\sharp\widehat{v}_i)} \prod_{h\in E} \boldsymbol{\alpha}^\top \bigodot_{(v_j,i_j)\in h} \mathbf{e}_{(\widehat{v}_j,i_j)}.$$

Furthermore, since all the summands in this expression are non-negative, it follows that $r(G) \neq 0$ if and only if there exist N vertices $\widehat{v}_i \in \widehat{V}(\ell(v_i))$ for $i \in [N]$ such that (i) $\boldsymbol{\alpha}^\top \bigodot_{(v_j,i_j)\in h} \mathbf{e}_{(\widehat{v}_j,i_j)} \neq 0$ for all $h \in E$. We claim that (i) is true if and only if G is a covering of $\widehat{G}$. Indeed, suppose that (i) holds and let $f : V \to \widehat{V}$ and $g : P_G \to P_{\widehat{G}}$ be the mappings defined by $f(v_i) = \widehat{v}_i$ and $g(v_i, j) = (\widehat{v}_i, j)$ for all $i \in [N]$. One can check that f is a covering map from G to $\widehat{G}$: it follows from the definitions of $\boldsymbol{\alpha}$ and $\odot$ that (i) is true if and only if $g(h) \in \widehat{E}$ for all $h \in E$, and there are no distinct $(v_j, i_j), (v_k, i_k)$ in a hyperedge $h \in E$ such that $(\widehat{v}_j, i_j) = (\widehat{v}_k, i_k)$, i.e. the restriction $g_{|h}$ is injective for any $h \in E$. Conversely, suppose that G is a covering of $\widehat{G}$, let $f : V \to \widehat{V}$ be a covering map from G to $\widehat{G}$ and let $g : P_G \to P_{\widehat{G}}$ be the induced mapping defined by $g(v_i, j) = (f(v_i), j)$ for all $i \in [N]$, $j \in [\sharp v_i]$. Let $h = \{p_1, \cdots, p_k\} \in E$ be any hyperedge of G (h connects the ports $p_1, \cdots, p_k \in P_G$ of G). Since the restriction $g_{|h}$ is bijective we have $\bigodot_{j\in[k]} \mathbf{e}_{g(p_j)} = \mathbf{e}_{\{g(p_1),\cdots,g(p_k)\}} = \mathbf{e}_{g(h)}$ by definition of $\odot$, and since $g(h) \in \widehat{E}$ we have that $\boldsymbol{\alpha}^\top \bigodot_{(v_j,i_j)\in h} \mathbf{e}_{(\widehat{v}_j,i_j)} \neq 0$ by definition of $\boldsymbol{\alpha}$. □

We call a family $\mathcal{H}$ of hypergraphs *covering-free* if for any $G \in \mathcal{H}$ there are no (non-trivial) covering of G in $\mathcal{H}$. For example, the family of rooted hypergraphs is covering-free but the family of circular strings is not: the circular string *abab* is a covering of *ab*.

Corollary 5.4. *For any covering-free family of connected hypergraphs $\mathcal{H}$, finite support series on $\mathcal{H}$ are recognizable.*

Proof. Let $r_{\widehat{G}}$ be the series from the previous proof and let $z = r_{\widehat{G}}(\widehat{G})$. For any scalar y, if we change the definition of $\boldsymbol{\alpha}$ to $\boldsymbol{\alpha}(\mathbf{e}_S) = (y/z)^{1/|\widehat{E}|}$ if $S \in \widehat{E}$ and 0 otherwise, we have $r_{\widehat{G}}(\widehat{G}) = y$. The corollary then directly follows from the previous theorem and Proposition 3.10. □

Remark 6. The existence of a recognizable string series invariant under cyclic shifts (i.e. $r(w_1 \cdots w_n) = r(w_n w_1 \cdots w_{n-1})$) does not imply that there exists a HWM on circular string computing the same series. Indeed, the

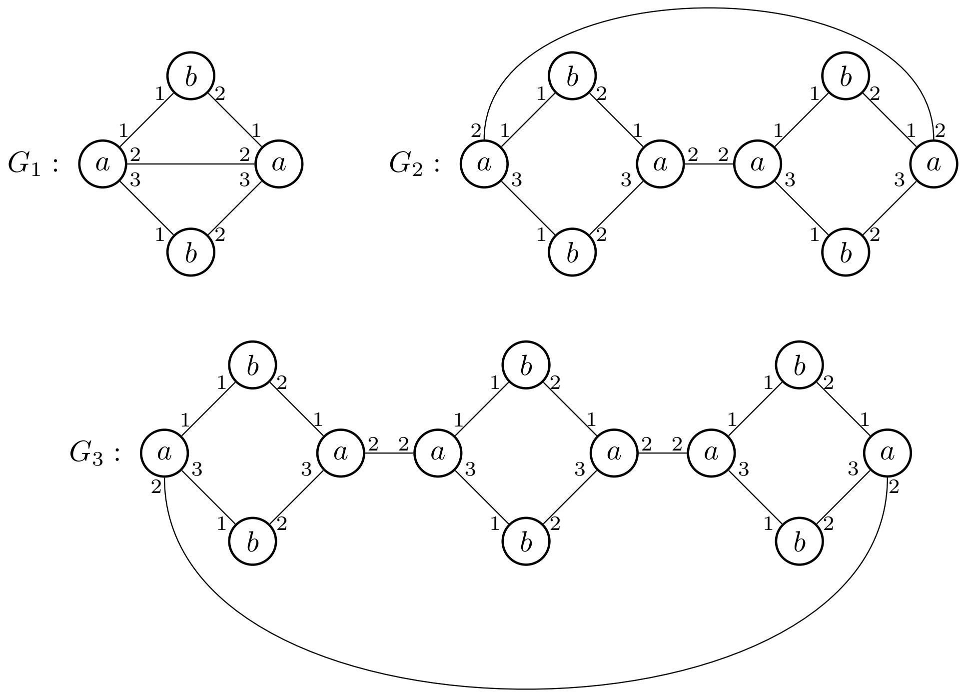


Figure 10: Each graph G_i is constructed by copying the initial graph ($G = G_1$) i times, splitting the edge between the two vertices labeled by a in each copy, and reconnecting the free ports to obtain a circular chain of copies of G.

string series $r(a^n) = 1$ if $n = 1$ and 0 otherwise is recognizable and invariant under cyclic shifts, but it follows from Lemma 3.12 that this series on circular strings cannot be computed by a HWM.

Remark 7. At the beginning of this section, we used a simple example on circular strings to show that finite support series are not recognizable in general. We now show how this simple example can be generalized to more complex families of hypergraphs. More precisely, we show that for any graph G *containing a cycle* and any recognizable series r such that $r(G) \neq 0$ there exists a *connected* hypergraph G' which is a non-trivial covering of G such that $r(G') \neq 0$.

Let r be a graph series computed by a HWM. Recall that the value computed by r on a graph G is the product of the values computed by r on the connected components of G (cf. Remark 3). Hence, if $r(G) \neq 0$ for some graph G there exists a nontrivial covering G' of G such that $r(G') \neq 0$ (consider the graph made of two copies of G). Moreover, if G contains a cycle, one can build a *connected* covering G' of G such that $r(G') \neq 0$. The construction consists in connecting copies of G by breaking the same edge in each copy of G and regrouping the freed ports into edges connecting the different copies (e.g. take two copies of the circular string a, split the edge in both copies, and reconnect the ports to obtain the circular string aa). Since G contains a cycle this can be done in such a way that G' is connected (choose the edge to split in a cycle). If we denote by G_i the graph obtained by this process from i copies of the initial graph (see Figure 10), one can show that there exists a matrix $\mathbf{M}$ such that $r(G_i) = \mathrm{Tr}(\mathbf{M}^i)$. The trace argument from Lemma 3.12 can then be used to show the result.

6. Learning HWMs: a Case Study on Circular Strings

In future works, we plan to study the problem of learning recognizable series on graphs and hypergraphs. In this section, we will tackle the problem of learning HWMs defined over the family of circular strings on a finite alphabet Σ. Circular strings are of particular interest for the study of HWMs because they can be seen as the simplest family of graphs with cycles that is not covering-free. Recall that a HWM $M = \langle \mathbb{R}^d, \{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}\rangle$ defined on circular strings computes the function

$$r_M : w \mapsto \mathrm{Tr}(\mathbf{M}^w)$$

for all $w \in \Sigma^*$ where $\mathbf{M}^w = \mathbf{M}^{\sigma_1}\mathbf{M}^{\sigma_2}\cdots\mathbf{M}^{\sigma_k}$ for any word $w = \sigma_1 \cdots \sigma_k \in \Sigma^*$.

We consider the learning paradigm of identification in the limit [27]. Let F be a class of functions from Σ^* to $\mathbb{R}$ and let $f \in F$ be the function to be learned. In this paradigm, at each time step t:

1. the learner is given a tuple $(x_t, f(x_t)) \in \Sigma^* \times \mathbb{R}$ coming from a stream of input-output examples $(x_1, f(x_1)), (x_2, f(x_2)), \ldots$ such that any string $x \in \Sigma^*$ occurs at least once in the stream, and
2. the learner makes an *hypothesis* (or guess) $h_t \in F$.

A class of functions F is said to be *identifiable in the limit* if there exists an algorithm that identifies any function f in F after examining a finite number of input-output examples, i.e. after some finite time, the hypothesis are all the same and equal to f. It is well known that the class of rational valued recognizable functions on strings (i.e. functions that can be computed by a WA) is identifiable in the limit[5] [7]. We will show in this section that the problem of learning HWMs on circular strings can be reduced (to some extent) to the problem of learning recognizable functions on strings. We will then be able to use learnability results on WAs for HWMs on circular strings. We start by investigating the relationship between HWMs on circular strings and classical weighted automata.

Equivalence between HWMs on Circular Strings and Weighted Automata. The following proposition shows that a series r computed by a HWM on circular strings can be computed by a string weighted automaton with a quadratic number of states.

Proposition 6.1. *Let* vec *be the vectorization operator that maps any matrix to the concatenation of its columns, i.e.* $\mathrm{vec}(\mathbf{A}) = (\mathbf{a}_1^\top, \cdots, \mathbf{a}_n^\top)^\top \in \mathbb{R}^{mn}$ *for any matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *with columns* $\mathbf{a}_1, \cdots, \mathbf{a}_n$.

For any HWM $M = \langle \mathbb{R}^d, \{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}\rangle$ *on circular strings on* Σ*, the recognizable string series*

$$A = \langle \mathbb{R}^{d^2}, \{\mathbf{A}^\sigma = \mathbf{I}_d \otimes \mathbf{M}^\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\iota} = \mathrm{vec}(\mathbf{I}_d), \boldsymbol{\tau} = \mathrm{vec}(\mathbf{I}_d)\rangle \tag{3}$$

satisfies $r_M(w) = r_A(w)$ *for all* $w \in \Sigma^*$.

Proof. For any $w = w_1 \cdots w_n \in \Sigma^*$ we have $r_M(w) = \mathrm{Tr}(\mathbf{M}^w) = \sum_{i\in[d]} \mathbf{M}^w_{ii} = \sum_{i\in[d]} \mathbf{e}_i^\top \mathbf{M}^w \mathbf{e}_i$ where $\mathbf{e}_i$ is the i-th vector of the canonical basis of $\mathbb{R}^d$. Since $\boldsymbol{\iota} = \boldsymbol{\tau} = (\mathbf{e}_1^\top, \cdots, \mathbf{e}_d^\top)^\top$ and $\mathbf{A}^\sigma = \mathbf{I} \otimes \mathbf{M}^\sigma$ is the block-diagonal matrix with

[5] In [7], the authors show that the class of recognizable functions is learnable in Angluin's exact learning model [2] which implies that this class is identifiable in the limit. Indeed, their algorithm will correctly identify a recognizable function f from a finite number of membership queries $(x_1, f(x_1)), ..., (x_n, f(x_n))$ and counter-examples $(z_1, f(z_1)), ..., (z_m, f(z_m))$ issued after equivalence queries. Since there exists a time T where these input-output examples will all have been presented to the learner, the WA returned by their algorithm will compute f and the hypothesis will remain unchanged after time T.

$\mathbf{M}^\sigma$ repeated d times on the diagonal, we have

$$r_A(w) = \boldsymbol{\iota}^\top \mathbf{A}^w \boldsymbol{\tau} = \begin{pmatrix}\mathbf{e}_1^\top \mathbf{e}_2^\top \cdots \mathbf{e}_d^\top\end{pmatrix} \begin{pmatrix} \mathbf{M}^w & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^w & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}^w \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_d \end{pmatrix}$$

$$= \sum_{i \in [d]} \mathbf{e}_i^\top \mathbf{M}^w \mathbf{e}_i = r_M(G_w).$$

□

It follows from the previous proposition that the computation of a d-dimensional HWM is equivalent to the sum of d string weighted automata (WA) with d states. Similarly to the rooted circular string case, the d WAs are identical except for their initial and final weights: each one has a unique initial and final state with weight one. Thus, the value of a string is the sum of the weights of all its paths in the WA starting and ending in the same state. It is the internal dynamic of this WA (and not its initial and final weights) that is relevant to the computation of the HWM on circular strings.

Going back to the learning problem, any HWM-recognizable function on circular strings can be computed by a WA and WAs are identifiable in the limit. Thus, for any HWM M on circular strings, there exists an algorithm that will return a WA $\hat{A}$ computing r_M after examining a finite number of input-output examples $\{(w_i, r_M(w_i))\}_{i=1}^N$. However, this learning result is improper in the sense that the algorithm returns a WA computing r_M rather than a HWM. It remains to show that we can construct a HWM B computing the same function as $\hat{A}$ and M; loosely speaking we need a constructive proof for the converse of Proposition 6.1. The remainder of this section will mainly be concerned with this problem which we formalize below.

Problem 6.2. *Given a minimal weighted automaton $\widehat{A} = \langle \mathbb{R}^n, \{\widehat{\mathbf{A}}^\sigma\}_{\sigma\in\Sigma}, \widehat{\boldsymbol{\iota}}, \widehat{\boldsymbol{\tau}} \rangle$ computing a series r on Σ^* that is recognizable by a HWM $M = \langle \mathbb{R}^d, \{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}\rangle$ on circular strings, can we construct a d-dimensional HWM $B = \langle \mathbb{R}^d, \{\mathbf{B}^\sigma\}_{\sigma\in\Sigma}\rangle$ such that $r_B(w) = r_{\widehat{A}}(w)$ for all $w \in \Sigma^*$?*

Before diving into the analysis of this problem we want to stress that, from a practical point of view, one could be satisfied with simply learning a HWM-recognizable function using learning algorithms for weighted automata. However, we will see that the analysis of Problem 6.2 reveals fundamental algebraic properties of HWMs that are particularly relevant to the present study. Moreover, this analysis will certainly prove useful for deriving learning algorithms for HWMs over richer families of hypergraphs than circular strings.

HWMs on circular strings and finite-dimensional algebras. We will show that Problem 6.2 can be solved when the algebra generated by the matrices $\{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}$ is simple. We conjecture that it can be solved in the general case but this is left for future work. Let us first recall some classical definitions from finite dimensional algebra (see e.g. [33]).

Let $\mathcal{M}$ be the algebra generated by the matrices $\{\mathbf{M}^\sigma\}_{\sigma\in\Sigma} \subseteq \mathbb{R}^{d\times d}$, that is the set of all (finite) linear combinations of matrices in $\{\mathbf{M}^w \mid w \in \Sigma^*\}$. The algebra $\mathcal{M}$ is said to be *simple* if it contains no two-sided ideals other than 0 and $\mathcal{M}$ itself. Similarly, a module M over $\mathcal{M}$ is said to be *simple* if $M \neq 0$ and if it does not contain any submodule other than 0 and M itself. A module M is *semi-simple* if M is the direct sum of a family of simple submodules. Similarly, $\mathcal{M}$ is *semi-simple* if it can be written as a direct sum of simple algebras, i.e.

$$\mathcal{M} = \mathcal{M}_1 \oplus \cdots \oplus \mathcal{M}_k$$

where each $\mathcal{M}_i$ is a simple sub-algebra of $\mathcal{M}$. The $\mathcal{M}$-module $\mathbb{R}^d$ can then be decomposed as $\mathbb{R}^d = E_1 \oplus \cdots \oplus E_k$ where each $E_i = \mathcal{M}_i\mathbb{R}^d$.

For example, the algebra $\mathcal{M}$ generated by the matrix $M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not semi-simple. Indeed, $M^2 - M = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ generates a proper ideal of $\mathcal{M}$ (thus $\mathcal{M}$ is not simple) and it can easily be shown that it is the only proper ideal of $\mathcal{M}$ (hence $\mathcal{M}$ is not semi-simple). However, it is easy to check that the one-letter HWM on circular strings generated by M can also be generated by the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and that the algebra generated by this matrix is semi-simple (this algebra is actually simple). We show in the following proposition that every HWM on circular strings can be generated by a semi-simple algebra of matrices.

Proposition 6.3. *Any recognizable series r on circular strings over an alphabet Σ can be computed by a HWM $N = \langle\mathbb{R}^d, \{\mathbf{N}^\sigma\}_{\sigma\in\Sigma}\rangle$ for which the algebra $\mathcal{N}$ generated by the matrices $\{\mathbf{N}^\sigma\}_{\sigma\in\Sigma}$ is semi-simple.*

Proof. Let $M = \langle \mathbb{R}^d, \{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}\rangle$ be a HWM computing r and let $\mathcal{M}$ be the algebra generated by $\{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}$. An element $\mathbf{M} \in \mathcal{M}$ is *strongly nilpotent* if $\mathbf{MX}$ is nilpotent for any $\mathbf{X} \in \mathcal{M}$ (i.e. there exists an integer k such that $(\mathbf{MX})^k = \mathbf{0}$). The proof relies on the following classical results on finite dimensional algebras:

1. the set of all strongly nilpotent elements of $\mathcal{M}$ is a two-sided ideal of $\mathcal{M}$ [21, Corollary 3.1.10], it is equal to the *radical* of $\mathcal{M}$ which we will denote by $rad(\mathcal{M})$;
2. $\mathbf{M}$ is in the radical of $\mathcal{M}$ if and only if $\mathrm{Tr}(\mathbf{MX}) = 0$ for all $\mathbf{X}$ in $\mathcal{M}$ [18, §65];
3. the quotient algebra $\mathcal{M}/rad(\mathcal{M})$ is semi-simple [21, Theorem 3.1.6];
4. there exists a subalgebra $\mathcal{N}$ of $\mathcal{M}$ that is isomorphic to $\mathcal{M}/rad(\mathcal{M})$ and such that $\mathcal{M} = \mathcal{N} \oplus rad(\mathcal{M})$ (direct sum of vector spaces) [21, Theorem 6.2.1]; furthermore, the corresponding projection $p : \mathcal{M} \to \mathcal{N}$ is an homomorphism.

The last result is known as the Wedderburn-Malcev Theorem; this theorem only holds when $\mathcal{M}/rad(\mathcal{M})$ is separable, which is always the case when the supporting field of the algebra $\mathcal{M}$ is of characteristic 0.

We can now prove the proposition. For each $\sigma \in \Sigma$, let $\mathbf{M}^\sigma = \mathbf{N}^\sigma + \bar{\mathbf{N}}^\sigma$ be the decomposition of $\mathbf{M}^\sigma$ according to the direct sum in (4.), that is $\mathbf{N}^\sigma = p(\mathbf{M}^\sigma)$ belongs to $\mathcal{N}$ and $\bar{\mathbf{N}}^\sigma$ belongs to the radical. It then follows from (3.) and (4.) that the algebra $\mathcal{N}'$ generated by $\{\mathbf{N}^\sigma\}_{\sigma\in\Sigma}$ is semi-simple. Indeed, since p is an homomorphism, the projections of a set of generator for $\mathcal{M}$ is a set of generator for $\mathcal{N}$, thus $\mathcal{N}' = \mathcal{N}$ which is semi-simple. Furthermore, using (2.) one can easily show that $\mathrm{Tr}(\mathbf{M}^w) = \mathrm{Tr}(\mathbf{N}^w)$ for all $w \in \Sigma^*$, hence the HWM $N = \langle \mathbb{R}^d, \{\mathbf{N}^\sigma\}_{\sigma\in\Sigma}\rangle$ satisfies the claim of the proposition. □

We can thus assume without loss of generality that the algebra $\mathcal{M}$ is semi-simple. We conjecture that Problem 6.2 can be solved if the algebra $\mathcal{M}$ is semi-simple and we show it here under the following additional assumption:

Assumption 6.4. *The algebra $\mathcal{M}$ generated by the matrices $\{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}$ in Problem 6.2 is simple.*

Observe that the full matrix algebra $\mathbb{R}^{d\times d}$ is simple, thus a sufficient condition for Assumption 6.4 to hold is that the matrices $\{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}$ generate the full matrix algebra[6]. One can show that this is a generic property of HWMs over circular strings on an alphabet of size at least 2: if the parameters $\{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}$ of a HWM are drawn at random from a continuous distribution over $\mathbb{R}^{d\times d}$, then the matrices $\{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}$ generate the full matrix algebra with probability one [42, Proposition 11].

Decomposition of simple algebras. We will now show how Problem 6.2 can be reduced to decomposing a simple algebra into *irreducible components*. A matrix algebra $\mathcal{T} \subseteq \mathbb{R}^{d\times d}$ is irreducible if no $\mathcal{T}$-invariant subspace other than $\mathbb{R}^d$ and $\{\mathbf{0}\}$ exists, where a $\mathcal{T}$-invariant subspace is a subspace $W \subseteq \mathbb{R}^d$ such that $\mathcal{T}W \subseteq W$. Consider for example the algebra $\mathcal{M}_{red} = \{\mathbf{I}_k \otimes \mathbf{M} \ : \ \mathbf{M} \in \mathbb{R}^{\bar{d}\times\bar{d}}\} \subseteq \mathbb{R}^{\bar{d}k\times\bar{d}k}$ where $\mathbf{I}_k$ is the $k\times k$ identity matrix. It is easy to check that $\mathcal{M}_{red}$ is simple but it is not irreducible: for example $\mathbb{R}^{\bar{d}} \oplus \{\mathbf{0}\} \oplus \cdots \oplus \{\mathbf{0}\} \subseteq \mathbb{R}^{k\bar{d}}$ is a proper $\mathcal{M}_{red}$-invariant subspace. It follows from the Wedderburn classification theorem that any simple algebra is isomorphic to a full matrix algebra (which is irreducible) [21, Corollary 2.4.6]; this implies that for any simple matrix algebra $\mathcal{A} \subseteq \mathbb{R}^{d\times d}$ there exists an invertible matrix $\mathbf{S}$ such that

$$\text{for all } \mathbf{A} \in \mathcal{A},\ \text{there exists } \bar{\mathbf{A}} \in \mathbb{R}^{\bar{d}\times\bar{d}} \ : \ \mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{I}_k \otimes \bar{\mathbf{A}}$$

where $d = k\bar{d}$, furthermore the algebra composed of the matrices $\bar{\mathbf{A}}$ is the full matrix algebra $\mathbb{R}^{\bar{d}\times\bar{d}}$.

Going back to Problem 6.2 under Assumption 6.4, since the algebra $\mathcal{M}$ generated by the matrices $\{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}$ is simple it follows from the previous discussion that there exists a set of matrices $\{\mathbf{N}^\sigma\}_{\sigma\in\Sigma} \subseteq \mathbb{R}^{\bar{d}\times\bar{d}}$ and an invertible matrix $\mathbf{S} \in \mathbb{R}^{d\times d}$ such that

$$\mathbf{S}^{-1}\mathbf{M}^\sigma\mathbf{S} = \mathbf{I}_k \otimes \mathbf{N}^\sigma \tag{4}$$

where $d = k\bar{d}$ and the algebra generated by $\{\mathbf{N}^\sigma\}_{\sigma\in\Sigma}$ is irreducible. This observation allows us to exhibit a simple minimal weighted automaton computing the function r_M.

Proposition 6.5. *Using the definition of the matrices $\{\mathbf{N}^\sigma\}_{\sigma\in\Sigma}$ in Eq.* (4)*, the weighted automaton*

$$A = \langle \mathbb{R}^{\bar{d}^2}, \{\mathbf{A}^\sigma = \mathbf{I}_{\bar{d}} \otimes \mathbf{N}^\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\iota} = k\,\mathrm{vec}(\mathbf{I}_{\bar{d}}), \boldsymbol{\tau} = \mathrm{vec}(\mathbf{I}_{\bar{d}})\rangle$$

[6] However this condition is not necessary: consider the case of a simple algebra that is not irreducible (such as the algebra $\mathcal{M}_{red}$ below).

is a minimal weighted automaton computing the same function (on circular strings) as the HWM $M = \langle \mathbb{R}^d, \{\mathbf{M}^\sigma\}_{\sigma\in\Sigma}\rangle$ from Problem 6.2.

Proof. Since every matrix $\mathbf{M}^\sigma$ is similar to $\mathbf{I}_k \otimes \mathbf{N}^\sigma$ it is easy to check that $r_M(w) = k\,\mathrm{Tr}(\mathbf{N}^w)$ for all words $w \in \Sigma^*$. Using the same argument as the one used in the proof of Proposition 6.1, it follows that the WA A computes the function r_M. It remains to show that A is minimal. Indeed, it is easy to check that for all words w,

$$\beta(w) := \mathbf{A}^w \boldsymbol{\tau} = (\mathbf{I}_{\bar{d}} \otimes \mathbf{N}^w)\,\mathrm{vec}(\mathbf{I}_{\bar{d}}) = \mathrm{vec}(\mathbf{N}^w)$$

and from a classical result from weighted automata theory the number of states of any minimal WA computing r_A is equal to the dimension of the linear space spanned by the vectors $\{\beta(w)\}_{w\in\Sigma^*}$. Since the algebra $\mathcal{N}$ is irreducible the dimension of the space spanned by the matrices $\{\mathbf{N}^w\}_{w\in\Sigma^*}$ is $\bar{d}^2$ which entails the result. □

It follows from the previous proposition that the minimal WA $\hat{A}$ from Problem 6.1 is of dimension $n = \bar{d}^2$. Moreover, $\hat{A}$ is similar to the automata A from the previous proposition. Consequently, there exists an invertible matrix $\mathbf{P}$ such that

$$\mathbf{P}^{-1}\widehat{\mathbf{A}}^\sigma\mathbf{P} = \mathbf{A}^\sigma = \mathbf{I}_{\bar{d}} \otimes \mathbf{N}^\sigma \text{ for all } \sigma \in \Sigma. \tag{5}$$

One can check that for any $w \in \Sigma^*$,

$$\mathrm{Tr}(\widehat{\mathbf{A}}^w) = \mathrm{Tr}(\mathbf{A}^w) = \bar{d}\,\mathrm{Tr}(\mathbf{N}^w) = \frac{\bar{d}}{k} r_M(w).$$

Observe that if we know the automaton A, we can retrieve the integer k (using the relation $r_A(w) = k\,\mathrm{Tr}(\mathbf{A}^w)$) and the matrices $\mathbf{N}^\sigma$. Moreover the HWM $B = \langle \mathbb{R}^d, \{\mathbf{B}^\sigma = \mathbf{I}_k \otimes \mathbf{N}^\sigma\}_{\sigma\in\Sigma}\rangle$ computes the same function as M and is a solution to Problem 6.2. Thus solving Problem 6.2 boils down to decomposing the simple[7] algebra $\widehat{\mathcal{A}}$ generated by the matrices $\{\widehat{\mathbf{A}}^\sigma\}_{\sigma\in\Sigma}$ into irreducible components.

[7] The fact that $\widehat{\mathcal{A}}$ is simple follows directly from the simplicity of $\mathcal{M}$.

Decomposing $\widehat{\mathcal{A}}$ into irreducible components.. More pragmatically, we wish to find an invertible matrix $\mathbf{P} \in \mathbb{R}^{\bar{d}^2 \times \bar{d}^2}$ such that each $\widehat{\mathbf{A}} \in \widehat{\mathcal{A}}$ satisfies

$$\mathbf{P}^{-1}\widehat{\mathbf{A}}\mathbf{P} = \mathbf{I} \otimes \mathbf{N} \tag{6}$$

where $\mathbf{I}$ is th $\bar{d} \times \bar{d}$ identity matrix and $\mathbf{N} \in \mathbb{R}^{\bar{d} \times \bar{d}}$.

We will now show how finding the matrix $\mathbf{P}$ can be done by adapting the method proposed in [37][8]. We say that a matrix $\widehat{\mathbf{A}} \in \widehat{\mathcal{A}}$ is *generic in eigenvalue structure* if the matrix $\mathbf{N}$ in decomposition (6) has distinct eigenvalues. Note that such a generic matrix $\widehat{\mathbf{A}}$ will have $\bar{d}$ distinct eigenvalues each of multiplicity $\bar{d}$. Given the WA $\widehat{A}$, it is easy to find a generic matrix in $\widehat{\mathcal{A}}$: first construct a basis of the linear space spanned by $\{\widehat{\mathbf{A}}^w\}_{w\in\Sigma^*}$ (this can be done in time polynomial in $|\Sigma|$ and $\bar{d}$) and then draw a linear combination of the basis elements at random. This is equivalent to drawing a matrix $\mathbf{N} \in \mathbb{R}^{\bar{d} \times \bar{d}}$ at random which will have distinct eigenvalues with probability 1.

The following proposition shows that the transformation $\mathbf{P}$ can be obtained through local transformations within the eigenspaces corresponding to distinct eigenvalues of a generic matrix in $\widehat{\mathcal{A}}$, followed by a global permutation of rows and columns.

Proposition 6.6. *(Adapted from [37, Proposition 3.5]) Let $\widehat{\mathbf{A}} \in \widehat{\mathcal{A}}$ be generic in eigenvalue structure and let $\mathbf{Q}$ be such that $\mathbf{Q}^{-1}\widehat{\mathbf{A}}\mathbf{Q} = \mathbf{D} \otimes \mathbf{I}$ where $\mathbf{I}$ is the $\bar{d} \times \bar{d}$ identity matrix and $\mathbf{D}$ is the diagonal matrix with the distinct eigenvalues $\lambda_1, \cdots, \lambda_{\bar{d}}$ of $\widehat{\mathbf{A}}$ on the diagonal.*

Then, the transformation $\mathbf{P}$ in (6) can be chosen in the form

$$\mathbf{P} = \mathbf{Q}diag(\mathbf{V}_1, \cdots, \mathbf{V}_{\bar{d}})\mathbf{\Pi}$$

where each $\mathbf{V}_i \in \mathbb{R}^{\bar{d} \times \bar{d}}$ is invertible and $\mathbf{\Pi}$ is a permutation matrix in $\mathbb{R}^{\bar{d}^2 \times \bar{d}^2}$.

Proof. Since $\mathbf{P}$ may be replaced by $\mathbf{P}(\mathbf{I} \otimes \mathbf{S})$ in (6) for any invertible matrix $\mathbf{S} \in \mathbb{R}^{\bar{d} \times \bar{d}}$, it may be assumed that $\mathbf{P}^{-1}\widehat{\mathbf{A}}\mathbf{P} = \mathbf{I} \otimes \mathbf{D}$.

Let $\mathbf{\Pi}$ be the $\bar{d}^2 \times \bar{d}^2$ permutation matrix satisfying $\mathbf{\Pi}(\mathbf{I} \otimes \mathbf{X})\mathbf{\Pi}^\top = \mathbf{X} \otimes \mathbf{I}$ for all $\mathbf{X} \in \mathbb{R}^{\bar{d} \times \bar{d}}$. We have $\mathbf{\Pi}\mathbf{P}^{-1}\widehat{\mathbf{A}}\mathbf{P}\mathbf{\Pi}^\top = \mathbf{D} \otimes \mathbf{I} = \mathbf{Q}^{-1}\widehat{\mathbf{A}}\mathbf{Q}$. Since the λ_j's are distinct, it follows that $\mathbf{P}\mathbf{\Pi}^\top = \mathbf{Q}diag(\mathbf{V}_1, \cdots, \mathbf{V}_{\bar{d}})$ for some invertible matrices $\mathbf{V}_1 \cdots, \mathbf{V}_{\bar{d}} \in \mathbb{R}^{\bar{d} \times \bar{d}}$ (each $\mathbf{V}_i$ acts as a change of basis on the i-th eigenspace), which concludes the proof. □

[8] In [37] the generators of the algebra $\widehat{\mathcal{A}}$ are assumed to be symmetric matrices.

Using the notations from the previous proposition, let $\widehat{\mathbf{A}}$ be a generic matrix in $\widehat{\mathcal{A}}$ and let $\mathbf{Q}$ be the matrix diagonalizing $\widehat{\mathbf{A}}$. To solve Problem 6.2, it remains to find the invertible matrices $\mathbf{V}_1, \cdots, \mathbf{V}_{\bar{d}}$. We decompose $\mathbf{Q}$ as $\mathbf{Q} = (\mathbf{Q}_1, \cdots, \mathbf{Q}_{\bar{d}})$ where each $\mathbf{Q}_i \in \mathbb{R}^{\bar{d}^2 \times \bar{d}}$ (the columns of each $\mathbf{Q}_i$ form a basis of the i-th eigenspace). We similarly decompose $\mathbf{Q}^{-1}$ as $\mathbf{Q}^{-1} = (\widetilde{\mathbf{Q}}_1^\top, \cdots, \widetilde{\mathbf{Q}}_{\bar{d}}^\top)^\top$ where each $\widetilde{\mathbf{Q}}_j \in \mathbb{R}^{\bar{d} \times \bar{d}^2}$. The method proposed in [37, Section 4.2] to find the matrices $\mathbf{V}_i$ relies on the observation that for any matrix $\widehat{\mathbf{A}}' \in \widehat{\mathcal{A}}$ we have

$$\mathbf{V}_i^{-1} \widetilde{\mathbf{Q}}_i \widehat{\mathbf{A}}' \mathbf{Q}_j \mathbf{V}_j = b_{i,j} \mathbf{I} \tag{7}$$

for some $b_{i,j} \in \mathbb{R}$ for all $i, j \in [\bar{d}]$. It follows from the previous proposition that this system of equations in $\mathbf{V}_i, b_{ij}$ for $i, j \in [\bar{d}]$ is solvable. Let $\mathbf{N}$ be the matrix satisfying $\mathbf{P}^{-1}\widehat{\mathbf{A}}\mathbf{P} = \mathbf{I} \otimes \mathbf{N}$ and let $\mathbf{B}$ be the matrix satisfying $\mathbf{P}^{-1}\widehat{\mathbf{A}}'\mathbf{P} = \mathbf{I} \otimes \mathbf{B}$. Since $\mathbf{P}$ may be replaced by $\mathbf{P}(\mathbf{I} \otimes \mathbf{S})$ in Eq. (6) for any invertible matrix $\mathbf{S} \in \mathbb{R}^{\bar{d} \times \bar{d}}$, we can choose (without loss of generality) a basis for which $\mathbf{N}_{1,1} = \lambda_1$ is the first eigenvalue of $\widehat{\mathbf{A}}$ and $\mathbf{B}_{i,i+1} = 1$ for $1 \leq i < \bar{d}$. It can easily be checked that this choice implies $\mathbf{V}_1 = \mathbf{I}$ and $\mathbf{V}_{i+1} = \widetilde{\mathbf{Q}}_{i+1} \widehat{\mathbf{A}}' \mathbf{Q}_i \mathbf{V}_i$ for $1 \leq i < \bar{d}$. Hence, given a matrix $\widehat{\mathbf{A}}' \in \widehat{\mathcal{A}}$ we can easily solve Eq. (7) for the matrices $\mathbf{V}_i$. Observe that we need to be careful with our choice of $\widehat{\mathbf{A}}'$: if we choose $\widehat{\mathbf{A}}' = \widehat{\mathbf{A}}$ it is easy to check that we would obtain $\mathbf{V}_i = \mathbf{0}$ for all $i > 1$ using the method described above. However, if $\widehat{\mathbf{A}}'$ is a generic matrix drawn at random in $\widehat{\mathcal{A}}$, one can check that each $\widetilde{\mathbf{Q}}_i \widehat{\mathbf{A}}' \mathbf{Q}_j$ is of full rank with probability one, which implies that there exists a unique solution $\mathbf{V}_1, \cdots, \mathbf{V}_{\bar{d}}$ to Eq. (7) under the constraints $\mathbf{N}_{1,1} = \lambda_1$ and $\mathbf{B}_{i,i+1} = 1$; hence the matrices $\mathbf{V}_1, \cdots, \mathbf{V}_{\bar{d}}$ obtained by the method described above will be such that $\mathbf{P} = \mathbf{Q} diag(\mathbf{V}_1, \cdots, \mathbf{V}_{\bar{d}}) \mathbf{\Pi}$ satisfies Eq. (6).

Learnability result. The overall procedure to solve Problem 6.2 under Assumption 6.4 is summarized in Algorithm 1 (whose complexity is polynomial in the size of the alphabet Σ and the dimension $n = \bar{d}^2$ of the minimal WA $\hat{A}$). The results presented in this section show that it is possible to elaborate a learning scheme for HWMs defined on circular strings by using a learning method for weighted automaton on strings: (i) learn a WA computing the series on strings and (ii) use Algorithm 1 to recover a HWM computing the same series on circular strings. This implies that under Assumption 6.4, *the class of (rational-valued) HWM-recognizable functions on circular strings is identifiable in the limit.* We conjecture that this result holds even when Assumption 6.4 is not satisfied and we plan to address this question in future works.

Algorithm 1 Solving Problem 6.2

Input: A minimal WA $\widehat{A} = \langle \mathbb{R}^{\bar{d}^2 \times \bar{d}^2}, \{\widehat{\mathbf{A}}^\sigma\}_{\sigma \in \Sigma} \rangle$ computing a function that is recognizable by a d-dimensional HWM on circular strings.
Output: A d-dimensional HWM B such that $r_{\widehat{A}}(w) = r_B(w)$ for all $w \in \Sigma^*$.

```
Let $\widehat{\mathbf{A}}$ be a generic matrix drawn at random in the algebra $\widehat{\mathcal{A}}$.
Let $\mathbf{Q}^{-1}\widehat{\mathbf{A}}\mathbf{Q} = \mathbf{D} \otimes \mathbf{I}_{\bar{d}}$ where $\mathbf{D}$ is the diagonal matrix with the distinct eigenvalues $\lambda_1, \cdots, \lambda_{\bar{d}}$ of $\widehat{\mathbf{A}}$ on the diagonal and $I_{\bar{d}}$ is the $\bar{d} \times \bar{d}$ identity matrix.
Decompose $\mathbf{Q}$ as $\mathbf{Q} = (\mathbf{Q}_1, \cdots, \mathbf{Q}_{\bar{d}})$ where each $\mathbf{Q}_i \in \mathbb{R}^{\bar{d}^2 \times \bar{d}}$, and $\mathbf{Q}^{-1}$ as $\mathbf{Q}^{-1} = (\widetilde{\mathbf{Q}}_1^\top, \cdots, \widetilde{\mathbf{Q}}_{\bar{d}}^\top)^\top$ where each $\widetilde{\mathbf{Q}}_i \in \mathbb{R}^{\bar{d} \times \bar{d}^2}$.
Let $\widehat{\mathbf{A}}' \neq \widehat{\mathbf{A}}$ be a second generic matrix drawn at random in $\widehat{\mathcal{A}}$.
Let $\mathbf{V}_1 = \mathbf{I}_{\bar{d}}$.
for $i = 2$ to $\bar{d}$ do
    Let $\mathbf{V}_i = \widetilde{\mathbf{Q}}_i \widehat{\mathbf{A}}' \mathbf{Q}_{i-1} \mathbf{V}_{i-1}$
end for
Let $\mathbf{P} = \mathbf{Q} diag(\mathbf{V}_1, \cdots, \mathbf{V}_n)\mathbf{\Pi}$ where $\mathbf{\Pi}$ is the $\bar{d}^2 \times \bar{d}^2$ permutation matrix satisfying $\mathbf{\Pi}(\mathbf{I}_{\bar{d}} \otimes \mathbf{X})\mathbf{\Pi}^\top = \mathbf{X} \otimes \mathbf{I}_{\bar{d}}$ for all $\mathbf{X} \in \mathbb{R}^{\bar{d} \times \bar{d}}$.
For each $\sigma \in \Sigma$ let $\mathbf{N}^\sigma \in \mathbb{R}^{\bar{d} \times \bar{d}}$ be the matrix satisfying
    $$\mathbf{P}^{-1}\widehat{\mathbf{A}}^\sigma \mathbf{P} = \mathbf{I}_{\bar{d}} \otimes \mathbf{N}^\sigma.$$
Let $k = r_{\widehat{A}}(\sigma)/\operatorname{Tr}(\mathbf{N}^\sigma)$ for an arbitrary $\sigma \in \Sigma$ s.t. $\operatorname{Tr}(\mathbf{N}^\sigma) \neq 0$.
return the HWM $B = \langle \mathbb{R}^d, \{\mathbf{B}^\sigma = \mathbf{I}_k \otimes \mathbf{N}^\sigma\}_{\sigma \in \Sigma} \rangle$ (where $d = k\bar{d}$).
```

From a broader perspective, the analysis provided in this section reveals a fundamental connection between functions computed by HWMs and the theory of finite dimensional algebras. This is in contrast with the theory of weighted automata (defined over fields) for which linear algebra tools are overall sufficient to tackle fundamental problems such as minimization, equivalence, and learning. For example, it is well known that two minimal WAs computing the same function are related by a simple change of basis, whereas we showed in this section that this is not the case for HWMs defined over circular strings: the 2-dimensional HWMs defined over a one letter alphabet with matrices $\mathbf{I}$ and $\mathbf{I} + \mathbf{e}_1\mathbf{e}_2^\top$ respectively, compute the same function over circular strings but these matrices are not similar. We showed that the fundamental algebraic object at play in this simple example is the radical of an algebra. It is clear that the theoretical problems that we plan to tackle for HWMs (e.g. minimality, learning, etc.) will necessitate the use of fundamental tools from algebra theory, which is both a challenging and promising perspective.

7. Conclusion

HWMs constitute a general framework to define computation models on families of graphs or hypergraphs and encompass rational or recognizable series on strings and trees. Extending the notion of finite state automata to complex structures such as graphs and hypergraphs is challenging. The notion of HWM bypasses this difficulty by focusing on the algebraic characterization of recognizable series where the finiteness is expressed by the dimension of the underlying vector space. The generative aspect of the computation is lost but the computation is still local and guided by structural components of the input (edges or hyperedges).

HWMs satisfy various expected properties, such as closure under sum and Hadamard product. Nevertheless, some others are not satisfied by the most general families of graphs or hypergraphs — closure under scalar multiplication, recognizability of finite support series — and it is necessary to restrict these families to get these properties satisfied. For example, the class of rooted graphs satisfies all the properties while widely generalizing the class of strings and trees.

Since a lot of data over a variety of fields naturally gives rise to a graph structure (images, secondary structure of RNA in bioinformatics, dependency graphs in NLP, etc.), this computational model offers a broad range of applications. It is often possible to describe graph structures by means of trees or strings but having these complex structures directly taken into account in the model may result in substantial gain, as it is illustrated by simple examples on circular strings (Proposition 4.1).

The next step will be to study how learning can be achieved within this framework, i.e. how the tensor components of a model M can be recovered or estimated from samples of the form $(G_1, \widehat{r_M(G_1)}), \ldots, (G_n, \widehat{r_M(G_n)})$. For example, learnability in Angluin's exact model will be investigated. Preliminary results on circular strings indicate that this is a promising, while not trivial, direction. General learning algorithms should rely on tensor decomposition techniques, which generalize the spectral methods used for learning rational series on strings and trees. We also plan to tackle algorithmic issues and to study how techniques and methods developed in the field of graphical models, such as message passing, variational methods, etc., could be adapted to the setting of HWMs.

References

[1] Dana Angluin. Local and global properties in networks of processors. In *Proceedings of the twelfth annual ACM symposium on Theory of computing*, pages 82–93. ACM, 1980.

[2] Dana Angluin. Queries and concept learning. *Machine learning*, 2(4):319–342, 1988.

[3] Parvaneh Babari and Manfred Droste. A Nivat theorem for weighted picture automata and weighted MSO logic. In *Language and Automata Theory and Applications*, volume 8977 of *LNCS*, pages 703–715. Springer, 2015.

[4] Raphaël Bailly, François Denis, and Liva Ralaivola. Grammatical inference as a principal component analysis problem. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 33–40. ACM, 2009.

[5] Raphaël Bailly, Amaury Habrard, and François Denis. A spectral approach for probabilistic grammatical inference on trees. In *Algorithmic Learning Theory*, pages 74–88. Springer, 2010.

[6] Borja Balle, Xavier Carreras, Franco M Luque, and Ariadna Quattoni. Spectral learning of weighted automata. *Machine learning*, 96(1-2):33–63, 2014.

[7] Amos Beimel, Francesco Bergadano, Nader H Bshouty, Eyal Kushilevitz, and Stefano Varricchio. Learning functions represented as multiplicity automata. *Journal of the ACM (JACM)*, 47(3):506–530, 2000.

[8] Jean Berstel and Christophe Reutenauer. Recognizable formal power series on trees. *Theoretical Computer Science*, 18(2):115–148, 1982.

[9] Jean Berstel and Christophe Reutenauer. *Rational series and their languages*. Springer-Verlag Berlin, 1988.

[10] Symeon Bozapalidis and Archontia Grammatikopoulou. Recognizable picture series. *Journal of Automata, Languages and Combinatorics*, 10(2/3):159–183, 2005.

[11] Jacob C Bridgeman and Christopher T Chubb. Hand-waving and interpretive dance: An introductory course on tensor networks. *arXiv preprint arXiv:1603.03039*, 2016.

[12] Lam Chi-Chung, P Sadayappan, and Rephael Wenger. On optimizing a class of multi-dimensional loops with reduction for parallel execution. *Parallel Processing Letters*, 7(02):157–168, 1997.

[13] Andrzej Cichocki. Era of big data processing: A new approach via tensor networks and tensor decompositions. *arXiv preprint arXiv:1403.2048*, 2014.

[14] Hubert Comon, Max Dauchet, Rémi Gilleron, Christof Löding, Florent Jacquemard, Denis Lugiez, Sophie Tison, and Marc Tommasi. Tree automata techniques and applications. Available on: `http://www.grappa.univ-lille3.fr/tata`, 2007.

[15] Bruno Courcelle. The monadic second-order logic of graphs. i. recognizable sets of finite graphs. *Information and computation*, 85(1):12–75, 1990.

[16] Andrew Critch. *Algebraic geometry of hidden Markov and related models*. PhD thesis, University of California, Berkeley, 2013.

[17] François Denis and Amaury Habrard. Learning rational stochastic tree languages. In *Algorithmic learning theory*, pages 242–256. Springer, 2007.

[18] Leonard Eugene Dickson. *Algebras and their arithmetics*. University of Chicago Press, 1923.

[19] Manfred Droste and Stefan Dück. Weighted automata and logics on graphs. In *Mathematical Foundations of Computer Science 2015*, pages 192–204. Springer, 2015.

[20] Manfred Droste, Werner Kuich, and Heiko Vogler. *Handbook of weighted automata*. Springer Science & Business Media, 2009.

[21] Yurj A Drozd and Vladimir V Kirichenko. *Finite dimensional algebras*. Springer Science & Business Media, 2012.

[22] Samuel Eilenberg and Bret Tilson. *Automata, languages, and machines*, volume 59. Academic press New York, 1976.

[23] Ina Fichtner. Weighted picture automata and weighted logics. *Theory of Computing Systems*, 48(1):48–78, 2011.

[24] Aniruddh Gandhi, Bakhadyr Khoussainov, and Jiamou Liu. Finite automata over structures. In *International Conference on Theory and Applications of Models of Computation*, pages 373–384. Springer, 2012.

[25] Dora Giammarresi and Antonio Restivo. Two-dimensional finite state recognizability. *Fundamenta Informaticae*, 25(3, 4):399–422, 1996.

[26] Dora Giammarresi, Antonio Restivo, Sebastian Seibert, and Wolfgang Thomas. Monadic second-order logic over rectangular pictures and recognizability by tiling systems. *Information and computation*, 125(1):32–45, 1996.

[27] E Mark Gold. Language identification in the limit. *Information and control*, 10(5):447–474, 1967.

[28] Wolfgang Hackbusch. *Tensor spaces and numerical tensor calculus*, volume 42. Springer Science & Business Media, 2012.

[29] Daniel Hsu, Sham M Kakade, and Tong Zhang. A spectral algorithm for learning hidden markov models. *Journal of Computer and System Sciences*, 78(5):1460–1480, 2012.

[30] Katsushi Inoue and Akira Nakamura. Nonclosure properties of two-dimensional on-line tessellation acceptors and one way parallel sequential array acceptors. *IEICE TRANSACTIONS (1976-1990)*, 60(9):475–476, 1977.

[31] Martin Kliesch, David Gross, and Jens Eisert. Matrix-product operators and states: Np-hardness and undecidability. *Physical review letters*, 113(16):160503, 2014.

[32] Werner Kuich and Arto Salomaa. *Semirings, automata, languages*, volume 5. Springer Science & Business Media, 2012.

[33] Serge Lang. Linear algebra. Undergraduate texts in mathematics, 1987.

[34] Michel Latteux and David Simplot. Recognizable picture languages and domino tiling. *Theoretical computer science*, 178(1):275–283, 1997.

[35] Johann A Makowsky and Tomer Kotek. Connection matrices and the definability of graph parameters. *Logical Methods in Computer Science*, 10, 2014.

[36] Ina Mäurer. Recognizable and rational picture series. In *Conference on Algebraic Informatics*, pages 141–155. Citeseer, 2005.

[37] Kazuo Murota, Yoshihiro Kanno, Masakazu Kojima, and Sadayoshi Kojima. A numerical algorithm for block-diagonal decomposition of matrix {*}-algebras with application to semidefinite programming. *Japan Journal of Industrial and Applied Mathematics*, 27(1):125–160, 2010.

[38] Alexander Novikov, Dmitrii Podoprikhin, Anton Osokin, and Dmitry P Vetrov. Tensorizing neural networks. In *Advances in Neural Information Processing Systems*, pages 442–450, 2015.

[39] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158, 2014.

[40] Roger Penrose. Applications of negative dimensional tensors. *Combinatorial mathematics and its applications*, pages 221–244, 1971.

[41] Robert NC Pfeifer, Jutho Haegeman, and Frank Verstraete. Faster identification of optimal contraction sequences for tensor networks. *Physical Review E*, 90(3):033315, 2014.

[42] Guillaume Rabusseau. *A Tensor Perspective on Weighted Automata, Low-Rank Regression and Algebraic Mixtures.* PhD thesis, Aix-Marseille Université, 2016.

[43] Kurt Reidemeister. *Einfuhrung in die kombinatorische Topologie*, volume 86. American Mathematical Soc., 1950.

[44] Grzegorz Rozenberg. *Handbook of graph grammars and computing by graph transformation*, volume 1. World Scientific, 1997.

[45] Jacques Sakarovitch. *Elements of automata theory*. Cambridge University Press, 2009.

[46] Arto Salomaa and Matti Soittola. *Automata-theoretic aspects of formal power series*. Springer Science & Business Media, 2012.

[47] Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96–192, 2011.

[48] Marcel Paul Schützenberger. On the definition of a family of automata. *Information and control*, 4(2):245–270, 1961.

[49] Edwin Stoudenmire and David J Schwab. Supervised learning with tensor networks. In *Advances in Neural Information Processing Systems*, pages 4799–4807, 2016.

[50] Wolfgang Thomas. On logics, tilings, and automata. In *International Colloquium on Automata, Languages, and Programming*, pages 441–454. Springer, 1991.