Magnitude of arithmetic scalar and matrix categories

Steve Huntsman

STR

`steve.huntsman@str.us`

We develop tools for explicitly constructing categories enriched over generating data and that compose via ordinary scalar and matrix arithmetic arithmetic operations. We characterize meaningful size maps, weightings, and magnitude that reveal features analogous to outliers that these same notions have previously been shown to reveal in the context of metric spaces. Throughout, we provide examples of such "outlier detection" relevant to the analysis of computer programs, neural networks, cyber-physical systems, and networks of communications channels.

1 Introduction

We consider a fairly general model $\mathbf{C}$ of a finite state machine or similar structure in which matrix data associated to states and/or transitions compose coherently, compatibly, and scalably with ordinary scalar and matrix arithmetic (e.g., Jacobians in numerical programs; filters in signal processing systems, etc.). That is, $\mathbf{C}$ is a category enriched [21, 15] over a suitable category $\mathbf{M}$ of matrices, with the underlying scalars as a special (and as it turns out, universal) case. For a scalar example, see Figure 1.

As a motivating conceptual example that is mathematically archetypal, consider a system of networked components with various external inputs and outputs. It might be that some co-located inputs must take different internal paths through the system because of engineering considerations, yet still necessary to instantiate behavior that is completely independent of the path taken. In the linear time-invariant setting [2, 26, 38, 4], this essentially means multiplying different matrices along different paths, yet producing the same result at any point where paths meet. [1] This is the case of taking matrix multiplication (or perhaps componentwise multiplication in a Fourier basis) as monoidal product [27, 15] in $\mathbf{M}$.

We aim to probe the geometry of $\mathbf{C}$ by analogy with subsets of Euclidean space, where the *weightings* recalled in §2 provide an excellent scale-dependent boundary/outlier detection mechanism [42, 6, 19]. This boundary-detecting behavior is related to the potential-theoretical notion of Bessel capacities [29].

The following paragraphs give a bit more specificity. For a finite digraph $D = (V, A)$, the *category* $\langle D \rangle$ *determined by* D is a digraph with objects/vertices $V(D)$ and morphisms/arcs $A(\langle D \rangle)$ given by self-loops on all vertices along with any arcs in the transitive closure of D. We want to use "generating" data in such a way as to obtain a $\mathbf{M}$-category $\mathbf{C}$ with underlying category $\langle D \rangle$. Such a construction can model memoryless systems in which series of transitions produce effects that only depend on the initial and final states. A similar construction in which nondeterministic automata are endowed with a scalar "cost function" on inputs was briefly considered in §5.3 of [8].

Taking $\mathbf{M}$ to be the discrete monoidal category $M_n(R)$ of $n \times n$ matrices over a commutative ring R with product given by matrix multiplication informs switched linear systems [26]. Taking $\mathbf{M}$ to be Arr(**FinVect**) or **FinStoch** with product given by the ordinary (Kronecker) tensor product informs quantum circuits [31] and the information theory of networks of discrete memoryless channels [37]. We

---

[1] In situations where there might be path dependence, a workaround is to consider finite sub-polytrees of the universal cover of D [18]. This is akin to loop unrolling for a compiler [10]. A violation of path-independent compositionality would indicate nonzero curvature (that physicists call a "field strength tensor") in a principal bundle over a discrete directed base [13, 14, 28].

treat these examples in turn, illustrating how to find "outliers" in structures directly relevant to computer program analysis, neural networks, cyber-physical systems, and networks of communications channels.

We can do this economically because the semiring S involved in the constructions below is universal in the following sense: there exists a unique S-category $\mathbf{C}$ such that $Z_{jk} = \mathbf{C}(j,k)$, and subsequent calculations only rely on ordinary matrix arithmetic. [2] In the cases we will deal with $\mathbf{M}$ is parameterized by a dimension, with the dimension one case essentially recovering S, so that the *size map* introduced in §2 is essentially a projection. In this sense, considering the *magnitude* of $\mathbf{M}$-categories *versus* S-categories *per se* is mostly uninteresting apart from the identification of a size map, and it factors through a universal case. However, the *construction* of $\mathbf{M}$-categories introduces a bit more nuance than the construction of S-categories and also gives important context for applications.

The paper is organized as follows. §2 reviews the category-theoretic formulation of magnitude. §3 discusses the existence, construction, and analysis of categories enriched over ordinary scalar arithmetic, with examples relevant to the analysis of computer programs and structures informing neural networks. §4 extends these results to the context of categories enriched over ordinary matrix arithmetic, with an example relevant to networks of communications channels. Finally, §5 makes some speculative remarks on combining arithmetic and process-oriented data.

2 Magnitude

Let $\mathbf{M} = (\mathbf{M}, \otimes, 1)$ be a monoidal category [27, 15] and $\mathbf{C}$ a finite $\mathbf{M}$-category. Recall that this means that $\mathbf{C}$ is specified by a finite set $\mathrm{Ob}(\mathbf{C})$; hom-objects $\mathbf{C}(j,k) \in \mathbf{M}$ for all $j,k \in \mathrm{Ob}(\mathbf{C})$; identity morphisms $1 \to \mathbf{C}(j,j)$ for all $j \in \mathrm{Ob}(\mathbf{C})$; and composition morphisms $\mathbf{C}(j,k) \otimes \mathbf{C}(k,\ell) \to \mathbf{C}(j,\ell)$ for all $j,k,\ell \in \mathrm{Ob}(\mathbf{C})$, all satisfying associativity and unitality properties [21, 15].

The theory of magnitude [24, 23] takes two principal inputs. The first input is a $\mathbf{M}$-category $\mathbf{C}$. The second input is a *size map* $\sigma : \mathrm{Ob}(\mathbf{M}) \to S$ where S is a semiring. [3] The size map is required to be constant on isomorphism classes and to satisfy $\sigma(1) = 1$ and $\sigma(X \otimes Y) = \sigma(X) \odot \sigma(Y)$, where the semiring unit and multiplication are indicated on the right-hand sides. If $n := |\mathrm{Ob}(\mathbf{C})| < \infty$ then its *similarity matrix* $Z \in M_n(S)$ is given by $Z_{jk} := \sigma(\mathbf{C}(j,k))$. Introducing the (common) notation

$$(f[X])_{jk} := f(X_{jk})$$

as a shorthand where X is a matrix over S and f is a function on S, we have $Z := \sigma[\mathbf{C}]$.

A *weighting* is a column vector w satisfying $Zw = 1$, where the semiring matrix multiplication and column vector of ones are indicated. A *coweighting* is the transpose of a weighting for Z^T. If Z has both a weighting and a coweighting, its *magnitude* is the sum of the components (both sums coincide).

Presently, examples and applications of magnitude are focused almost entirely on *Lawvere metric spaces*, i.e., categories enriched over $([0,\infty], \geq)$ where the monoidal product is ordinary addition. These are also known as *extended quasipseudometric spaces* since they generalize metric spaces by allowing distances that are infinite (extended), asymmetric (quasi-), or zero (pseudo-). As far as we are aware, the only exceptions to this focus at present are this paper and [9, 18].

---

[2] Recall that a semiring S is a discrete monoidal category with auxiliary additive structure.

[3] In our context, when $\mathbf{M}$ is a semiring like $M_n(R)$ it is possible–and may be useful–to take σ to be the identity so $S = \mathbf{M}$.

3 Scalar categories

Let $S \in \{\mathbb{R}, \mathbb{C}\}$ and consider the system

$$Z_{jk}Z_{k\ell} = Z_{j\ell} \tag{1}$$

for $(j,k), (k,\ell), (j,\ell) \in A(\langle D \rangle)$. [4] This system is satisfied iff Z defines a S-category $\mathbf{C}$ such that $Z_{jk} = \mathbf{C}(j,k)$. A class of solutions to (1) is

$$Z_{jk} := p_j^{-1} p_k \tag{2}$$

for $p : V(D) \to S_\times$. We will show that (2) turns out to be the general nondegenerate case. For reasons that will become apparent in §4.2, it will be helpful to establish when integral solutions exist.

Let $\tilde{A}(D)$ denote the arcs in D that are not loops. Let $\Gamma_2(D) := \{(j,k,\ell) : (j,k), (k,\ell) \in \tilde{A}(\langle D \rangle); j \neq \ell\}$ be the set of nondegenerate paths in $\langle D \rangle$ of length two. If $|\Gamma_2(D)| = 0$, then we can trivially obtain all solutions of (1), so assume w.l.o.g. $|\Gamma_2(D)| > 0$. Lemma 1 gives a general solution to (1) (see Figure 1).

Lemma 1. *Define a $|\Gamma_2(D)| \times |\tilde{A}(\langle D \rangle)|$ matrix M to have entries that are zero except for $M_{(j,k,\ell),(j,k)} = 1$, $M_{(j,k,\ell),(k,\ell)} = 1$, and $M_{(j,k,\ell),(j,\ell)} = -1$. Then $m := \dim \ker M > 0$ and any solution to (1) is of the form*

$$Z_{jk} = \prod_{i=1}^{m} c_i^{y_{(j,k)}^{(i)}} \tag{3}$$

where $\{y^{(i)}\}_{i=1}^{m}$ is a basis for $\ker M$ and $(c_i)_{i=1}^{m} \in S^m$.

Proof. W.l.o.g., assume that $\langle D \rangle$ is weakly connected. If $(j,k,\ell) \in \Gamma_2$, then by transitivity of composition $(j,\ell) \in \tilde{A}(\langle D \rangle)$ and similarly $|\Gamma_2(D)| < |\tilde{A}(\langle D \rangle)|$. $My = 0$ iff $y_{(j,k)} + y_{(k,\ell)} - y_{(j,\ell)} = 0$, so we can take $y_{(j,k)} = \log Z_{jk}$. Since $|\Gamma_2(D)| < |\tilde{A}(\langle D \rangle)|$, we have $m := \dim \ker M > 0$. $\square$

By normalizing the reduced row echelon form, we get $y^{(i)} \in \mathbb{Z}^{|\tilde{A}(\langle D \rangle)|}$. If $(c_i)_{i=1}^{m} \in \mathbb{Z}_+^m$, then $Z_{jk} \in \mathbb{Q}_+$.

Proposition 1. *If D (or equivalently, $\langle D \rangle$) is a directed acyclic graph (DAG), the system (1) admits nontrivial solutions over $\mathbb{Z}_+$. More generally, $\prod_{(j,k) \in \gamma} Z_{jk} = 1$ for any cycle $\gamma \in \langle D \rangle$, so any solution over $\mathbb{Z}_+$ must be unity on any arcs that are in a cycle: in particular, if (j,k) is in a cycle, then $Z_{jk} = Z_{kj}^{-1}$.*

We now aim at closed forms for generating data on arcs and on vertices. Let U be the functor from quivers to undirected graphs that forgets arc orientations and multiplicities. For $(i,i') \in V(D)^2$, write

$$\varepsilon(i,i') := \begin{cases} 1 & \text{if } (i,i') \in A(D) \\ -1 & \text{if } (i',i) \in A(D) \text{ and } (i,i') \notin A(D) \, . \\ 0 & \text{otherwise} \end{cases}$$

Theorem 1. *Let D be weak (= weakly connected), T be a spanning tree of $U(D)$, and $W : E(T) \to S$ nondegenerate, where here as usual $E(\cdot)$ indicates the edges of an undirected graph. The assignment*

$$Z_{jk} = \prod_{(i,i') \in T[j,k]} W(i,i')^{\varepsilon(i,i')}, \tag{4}$$

with the product over edges in the path $T[j,k]$ in T from j to k, is well-defined and satisfies (1). If D is a DAG, then $W : E(T) \to \mathbb{Z}_+$ yields $Z_{jk} \in \mathbb{Z}_+$. Finally, any nondegenerate solution of (1) has the form (4).

---

[4]If we write $d := \tau \log[Z]$ for a suitable scalar τ, then (1) takes the form of the triangle equality $d_{jk} + d_{k\ell} = d_{j\ell}$, which highlights a similarity with "vanilla" magnitude of Lawvere metric spaces that obey a criterion similar to Menger convexity [25]. There are other related notions: a weighted undirected graph is called *graph-geodetic* (respectively, *cutpoint-additive*) when the preceding triangle inequality holds if (respectively, iff) every path from j to ℓ passes through k [22, 12, 7].

Proof. The assignment (4) is well-defined since $T[j,k]$ exists; it is unique since D is weak. The assignment satisfies (1) because the concatenation of $T[j,k]$ and $T[k,\ell]$ is $T[j,\ell]$. If D is a DAG, then $\varepsilon \equiv 1$. Finally, by transitivity, any nondegenerate solution of (1) has the form (4).   □

Theorem 2. *If D is weak, any nondegenerate solution of* (1) *has the form $Z_{jk} = p_j^{-1} p_k$ for some p.* [5]

Proof. Let Z satisfy (1) and let P be a spanning polytree of D, i.e., a spanning digraph such that $T := U(P)$ is a tree. Pick $i \in V(D)$ and set $p_i = 1$, then extend p to $V(P) = V(D)$ by traversing P and solving $Z_{jk} = p_j^{-1} p_k$ on $A(P)$. Finally, take $W(j,k) := Z_{jk}^{\varepsilon(j,k)}$ on T and apply Theorem 1 to recover Z.   □

Example 1. *Figure 1 shows an example that is essentially generic in light of the structural characterization of transitive digraphs in Proposition 2.3.1 of [5]. However, its magnitude is undefined.*
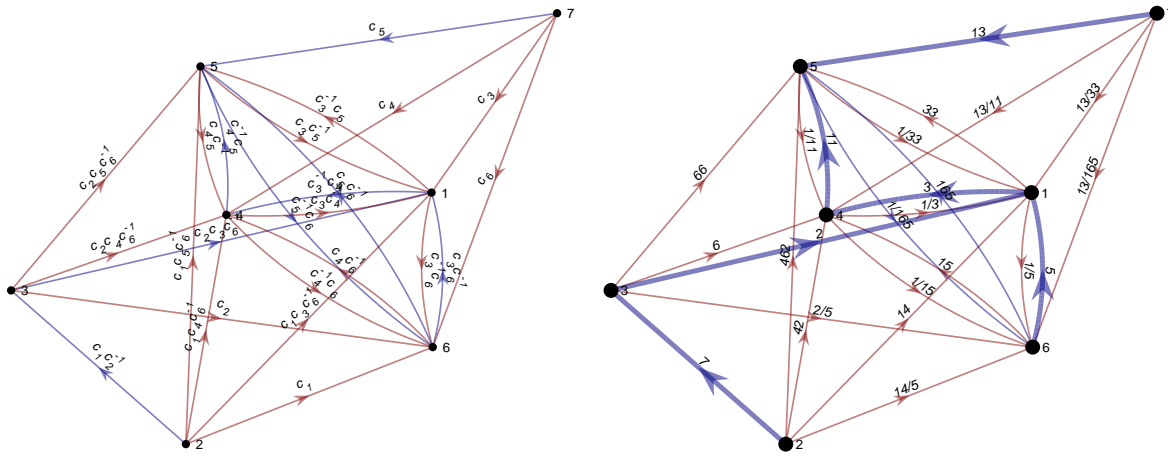

Figure 1: (Left) A symbolic solution to (1) of the form (3) on a transitive digraph (red) generated from an underlying digraph (blue). (Right) A specific solution to (1) generated using the assignment (4) for the **highlighted spanning (poly)tree**. Taking $p \propto (14, 1, 7, 42, 462, 14/5, 462/13)$ yields $Z_{jk} = p_j^{-1} p_k$.

The simplest solutions to (1) fail to give interesting structure.

Lemma 2. *If $Z_{jk} = Z_{kj}^{-1}$ for all (j,k), then* **C** *does not have well-defined magnitude unless $Z_{jk} \equiv 1$.*

Proof. By hypothesis $Z_{j\ell} = Z_{jk} Z_{k\ell} = Z_{kj}^{-1} Z_{k\ell}$ for all ℓ. Since the jth and kth rows of Z are constant multiples of each other, the equation $Zw = 1$ only has a solution if $Z_{jk} = 1$.   □

Similar considerations also show that if D has a cycle (that is not a loop) and **C** has magnitude, the magnitude must be unity. However, the space of weightings still encodes nontrivial information about **C**.

Example 2. *Consider a toy program that is constructed as follows. We generate a program "skeleton" using productions from the probabilistic context free grammar [40]*

$$S \rightarrow S;\ S \,|\, if\ b;\ S;\ fi \,|\, while\ b;\ S;\ end$$

---

[5]If B is the incidence matrix of D, then $\log[\text{vec}\,Z] = B^T \log[p]$, where vec stacks matrix columns and p is taken as a vector.

Table 1: Control flow graph arc: $[\cdot] :=$ line number of matching token.

source at line j	target($\top$)	target($\bot$)
if b	$j+1$	[fi]+1
while b	$j+1$	[end]+1
end	[while]	$\cdot$
fi or S	$j+1$	$\cdot$

where S is shorthand for a line separator: the production probabilities are respectively 0.6, 0.1, *and* 0.3. *The tokens S and b respectively represent statements/subroutines and Boolean predicates.*

Next, we form the resulting control flow graph *[10] by associating vertices with lines in the skeleton and edges according to Table 1. We also prepend a START line/vertex/arc and append a HALT line/vertex/arc to both the program and the control flow graph. We can explicitly instantiate an executable program by i) replacing a token S on line j of the program by an explicit statement and ii) replacing a token b on line k of the program by an explicit predicate.*

Consider the following assignment of scalar data to arcs of the control flow graph:

- *An arc of the form* $(S_j, \cdot_{j+1})$ *is assigned* $\delta_j \in S$;

- *All other arcs not of the form* $(if, \cdot_{[fi]+1})$ *or* $(end, while_{[while]})$ *are assigned* $1 \in S$;

- *An arc of the form* $(if, \cdot_{[fi]+1})$ *is assigned the (ordered) product of data assigned to any other path between its source and target;*

- *An arc of the form* $(end, while_{[while]})$ *is assigned the inverse of the (ordered) product of data assigned to any other path between its target and source.*

Proposition 2. *The assignment above is well defined and uniquely corresponds to an S-category.* □

Consider the control flow graph obtained with 20 *context free grammar productions shown in Figure 2 and the assignment* $\delta_j = 2$ *for all j. The resulting similarity matrix and its kernel are respectively shown in the left and right panels of Figure 3. The space of weightings is obtained by adding the vector* $(1, 0, \ldots, 0)^T$ *to the kernel: this vector corresponds to the START vertex. The space of coweightings is similar (not shown). The basis vectors in the kernel all sum to zero, so the magnitude is always unity.*
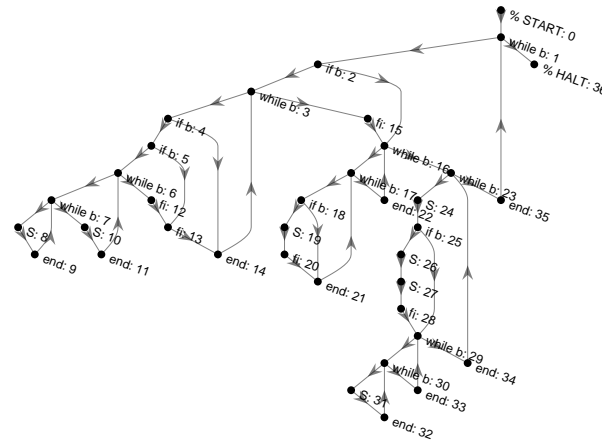

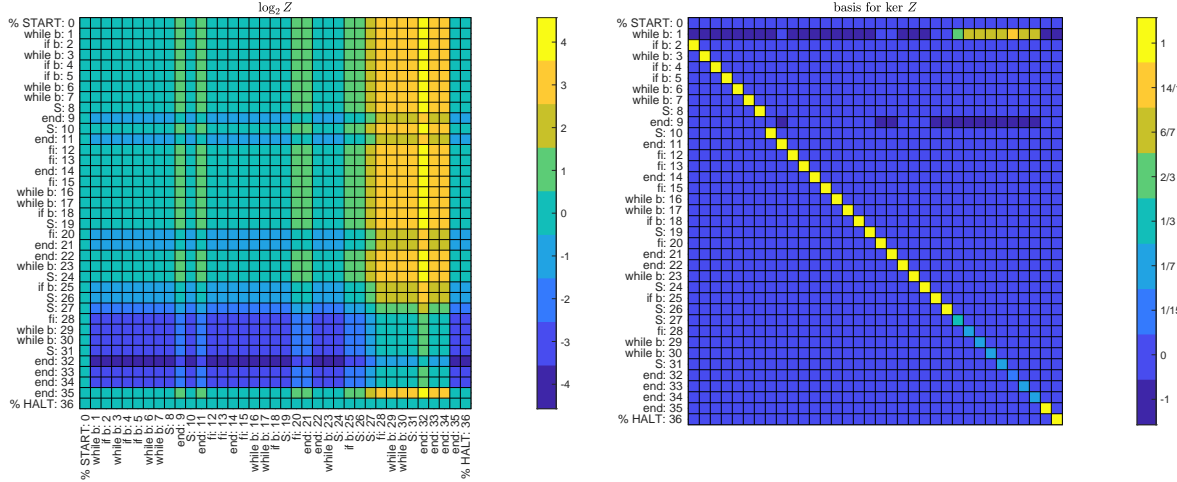
Figure 2: A simple control flow graph.

Figure 3: (Left) $\log_2$ of the similarity matrix Z arising from the assignment $\delta_j = 2$ for all j to the program with control flow graph in Figure 2. (Right) A rational basis for the kernel of Z, with maximum absolute value of vector entries normalized to unity. Note that the color scheme is quantized/nonlinear.

Suppose now that we change one of the δ_j to equal 8 instead of 2. The (maxima of the rows of the) resulting kernels are shown in the left panels of Figure 4; the right panels are similar but with $\delta_j = 1/4$. Although the magnitude is always unity, the space of weightings encodes globally contextualized information about the program "geometry."

If the sizes are positive, taking logarithms yields a dissimilarity that further essentially reduces the situation to a case of ordinary Lawvere metric magnitude for arc-weighted DAGs. In fact we can exploit this to gain some intuition for more general digraphs. Suppose all of the sizes on a spanning (poly)tree satisfy $\sigma = \exp(-1)$: then up to degeneracies related to reachability, $Z = \exp[-d]$ where d is the digraph distance. By analogy with Euclidean distance matrices, we thus expect the resulting space of weightings to indicate vertices that are "large" by virtue of being "peripheral" in a way that sometimes but not always explicitly correlates to degree.

Example 3. *Let $K^{\rightarrow}_{n_1,\dots,n_L}$ denote the DAG with $V(K^{\rightarrow}_{n_1,\dots,n_L}) := \bigsqcup_{\ell=1}^{L} K_\ell$ for $K_\ell := [n_\ell]$ and $A(K^{\rightarrow}_{n_1,\dots,n_L}) := \{(v,v') : v \in K_\ell, v' \in K_{\ell+1}; \ell \in [L-1]\}$. This DAG corresponds to the architecture of a fully connected multilayer perceptron (MLP) with L layers of widths $n_1,\dots,n_L$ [17].*

Now consider a sub-DAG $D \subset K^{\rightarrow}_{n_1,\dots,n_L}$ with arc weights $\sim U([0,1])$ corresponding to a sparsely connected MLP. Fixing D and retaining the arc weights corresponding to a random spanning tree of $U(D)$ [6] defines Z and hence w as a random variable, as shown for $N = 500$ realizations on the right of Figure 5 corresponding to the weighted DAG on the left.

w is statistically well-behaved: in our experiments, individual components of w all satisfy the hypothesis of being sampled from a normal distribution according to the Anderson-Darling test [34] with the best significance levels that are provided for in a standard computational implementation.[7] None of the components of w are trivial except those corresponding to vertices with outdegree zero, which have

---

[6]We produce a random spanning tree by taking a minimal spanning tree of $U(D)$ augmented with temporary edge weights $\sim U([0,1])$. This spanning tree is generally not *uniformly* random, but such trees can be produced [1].

[7]Multivariate normality tests along the lines of [41] are computationally prohibitive and cannot be nearly as conclusive in our context because of the high dimension.
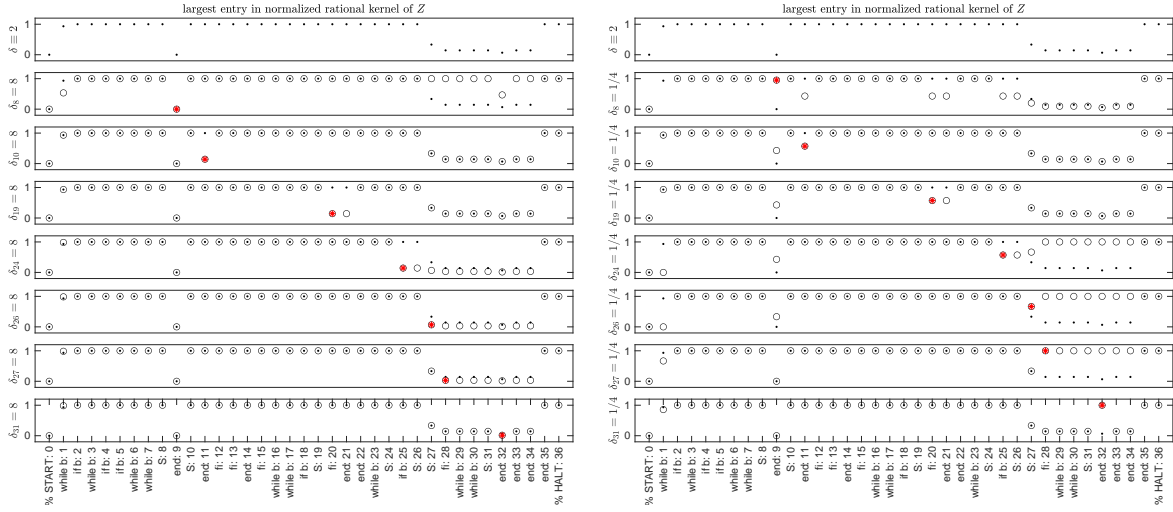
Figure 4: (Top left panel) Maxima of rows of the kernel of Z from figure 3. (Lower left panels) As in the top panel, but also showing (with ○) maxima for $\delta_j = 8$ with $j = 8, 10, 19, 24, 26, 27, 31$, respectively. The maximum for $j+1$ is shown with a red *. (Right panels) As in the left panels, but taking $\delta_j = 1/4$.
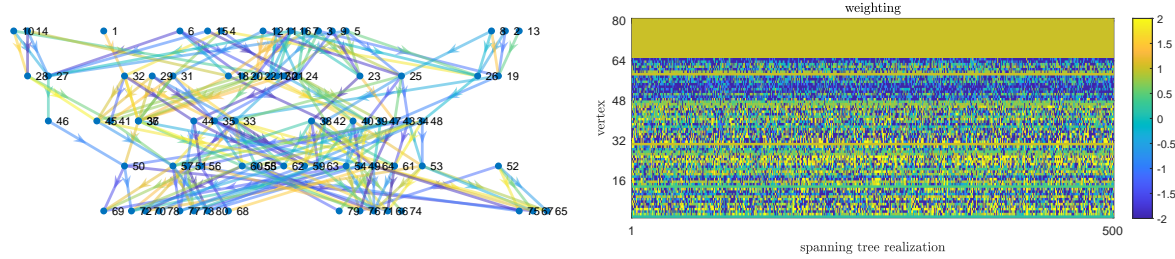

Figure 5: (Left) A weighted sub-DAG of $K^{\rightarrow}_{16,16,16,16,16}$ with arc weights $\sim U([0,1])$ according to the same (rescaled) color scheme on the right. (Right) $N = 500$ realizations of the weighting of the $\mathbb{R}$-category obtained from generating data on a random spanning tree. The color axis has been truncated for clarity. The statistical regularity of w is visible as horizontal striping.

unit values; components for vertices with indegree zero each have fixed values, and components for other vertices are normally distributed. Neighboring vertices tend to have weighting components of opposite signs, consistent with the general intuition in Euclidean space from [42, 6, 19] that negative weighting components tend to occur "just behind a local boundary" with large positive weighting components.

In larger networks, presumptive "near-outlier" vertices with the statistically least and greatest weighting components tend to be densely connected. Specifically, consider a measure of central tendency $\mathbb{T}$ (e.g., mean, median, etc.) and $t_- < t_+$ such that just a few vertices are in each of the sets $T_- := \{j : \mathbb{T}(w_j) < t_-\}$ and $T_+ := \{j : \mathbb{T}(w_j) > t_+\}$. Now writing $T := T_- \cup T_+$, consider the set

$$X := T \cup \left(\{i : (i,j) \in A(D); j \in T\} \cap \{k : (j',k) \in A(D); j' \in T\} \right).$$

The sub-DAG induced by X is generally densely connected, though this sub-DAG itself can be noisy. Figure 6 shows an example building on Figure 5; we have observed this behavior across other examples.

Taken as a whole, these results suggest that enforcing such compositionality of weights in a regularization and/or pruning strategy for neural networks might be useful in a way akin to dropout [17].
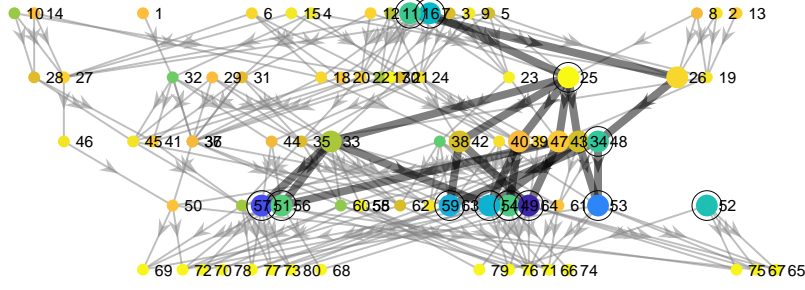
Figure 6:   The subgraph of the DAG from Figure 5 induced by X for $\mathbb{T} = $ median and $(t_-, t_+) = (0.14, 0.99)$. Vertices are colored by median weighting; X and T are respectively larger and circled.

The behavior described above (except for many disconnected components in the salient sub-polytree, due to obvious and otherwise irrelevant structure) manifests unambiguously when D is a polytree. Figure 7 shows the weighting on a binary polytree with a realization of arc weights $\sim U(\{1,2\})$. [8]
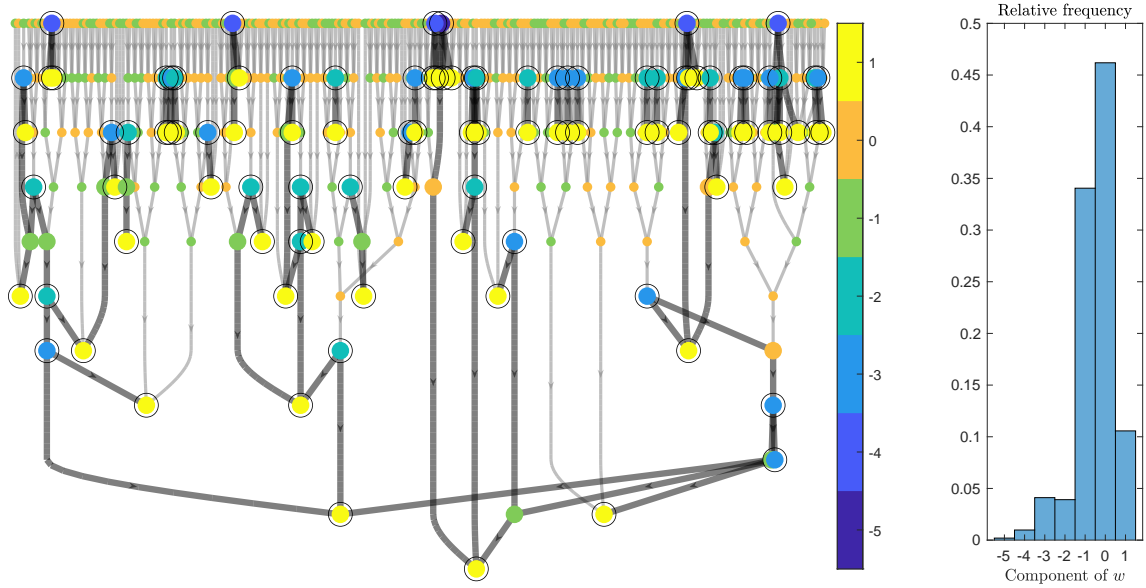

Figure 7:   (Left) As in Figure 6 for a polytree with arc weights $\sim U(\{1,2\})$ and $\{j : w_j < -1\}$ and $\{j : w_j > 0\}$ as respective analogues of T_- and T_+. (Right) Relative frequencies of weighting components.

4   Matrix categories

A *matrix category* is an enriched category whose hom-sets are matrices. In particular, a matrix category is a representation of a digraph *qua* quiver [36] that satisfies a compositional coherence condition. [9]

---

[8]It turns out that taking arc weights $\sim U([n])$ gives approximately the same result up to affine scaling for any $n \geq 2$ so long as the weights are obtained by quantizing the output of the same pseudorandom number generator.

[9]The category of quiver representations has a subcategory of digraph representations, and there is in turn a category of matrix categories of a digraph. The representation theory of these objects is probably interesting in its own right.

There are several inequivalent flavors of this construction, though from the perspective of magnitude these factor through the universal construction of §3.

4.1   Matrix multiplication as monoidal product

We first consider a model of a finite state machine in which linear maps are associated to the states and/or transitions. This model is applicable to many situations in control theory and/or the analysis of cyber-physical systems, e.g., switched linear systems [26].

For $n < \infty$, we can treat the monoid $M_n(R)$ of $n \times n$ matrices over a commutative ring R as a monoidal category with objects $M_n(R)$, only identity morphisms, and with ordinary matrix multiplication as the monoidal product. Suitably reinterpreted (in particular, taking ordered products), the assignment (4) defines a $M_n(R)$-category, and any finite $M_n(R)$-category can be realized in this way: similarly, we can write $\mathbf{C}(j,k) = p_j^{-1} p_k$ for suitable p. Figure 8 shows examples for $SL_2(\mathbb{Z})$: these are convenient to write down, though trivial from the perspective of magnitude.
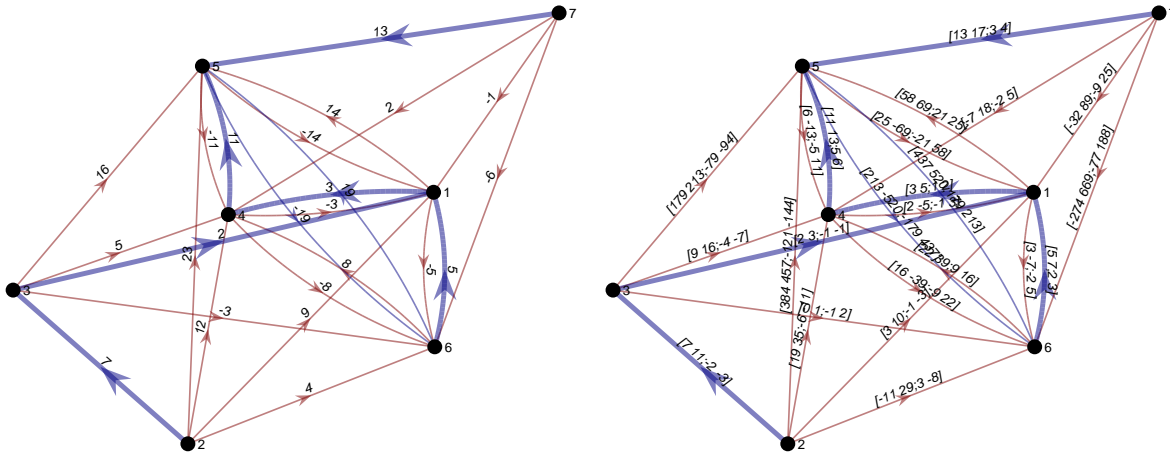

Figure 8:   (Left) As in the right panel of Figure 1 but labeled with the $(1,2)$ entries of matrices of the form $u(x) := \left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right)$. Note that this essentially uses scalar addition as monoidal product, and the matrices involved here all have unit determinant. Taking $p = u^{\times 7}(9, 0, 7, 12, 23, 4, 10)$ gives $\mathbf{C}(j,k) = p_j^{-1} p_k$.   (Right) A more involved example over $SL_2(\mathbb{Z})$: here taking $p = \left(\left(\begin{smallmatrix} 3 & 10 \\ -1 & -3 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 7 & 11 \\ -2 & -3 \end{smallmatrix}\right), \left(\begin{smallmatrix} 19 & 35 \\ -6 & -11 \end{smallmatrix}\right), \left(\begin{smallmatrix} 384 & 457 \\ -121 & -144 \end{smallmatrix}\right), \left(\begin{smallmatrix} -11 & 29 \\ 3 & -8 \end{smallmatrix}\right), \left(\begin{smallmatrix} 165 & -587 \\ -52 & 185 \end{smallmatrix}\right)\right)$ gives $\mathbf{C}(j,k) = p_j^{-1} p_k$.

Meanwhile, the determinant furnishes a suitable size map and so we can apply the constructions of §2. [10] As far as similarity matrices, weightings, and magnitude are concerned, we can replace matrices with their determinants and apply the results from §3 without explicitly forming all of the data for a $M_n(R)$-category. That is, from this perspective we do not really gain anything by considering matrices *versus* scalars: the case of $n > 1$ factors through the case $n = 1$.

---

[10] More generally, any well behaved size map factors through the determinant. Note also that determinants similarly inform enrichment over Arr(**Field**) or Arr(**Ring**) in a way that borrows from the present section and §4.2: in these cases, a size map is given by the norm of a field extension or an ideal [20].

4.2   Kronecker product as monoidal product

There is also a variant of §4.1 with the usual (i.e., Kronecker tensor) monoidal product. The most obvious possible application is to quantum circuits [31] [11] although as we shall see considerations of magnitude once again factor through to the scalar case of §3 and we postpone an example to §4.3.

Recall that the *arrow category* $\mathrm{Arr}(\mathbf{X})$ of a category $\mathbf{X}$ has as objects the morphisms of $\mathbf{X}$; morphisms given by commutative squares in $\mathbf{X}$; and composition given by concatenation of these commutative squares. If $\mathbf{X}$ is monoidal, then so is $\mathrm{Arr}(\mathbf{X})$. [12] In particular, for $S \in \{\mathbb{R}, \mathbb{C}\}$, the objects of $\mathrm{Arr}(\mathbf{FinVect}_S)$ are (linear maps that can be represented as) matrices over S, [13] and $\mathrm{Arr}(\mathbf{FinVect}_S)$ is monoidal with respect to the usual tensor product $\otimes$, i.e., the enriched composition law is

$$\mathbf{C}(j,k) \otimes \mathbf{C}(k,\ell) = \mathbf{C}(j,\ell). \tag{5}$$

We proceed to construct the space of $\mathrm{Arr}(\mathbf{FinVect}_S)$-categories $\mathbf{C}$ with underlying category determined by a finite digraph $D = (V, A)$. As in §4.1, the hom-objects of $\mathbf{C}$ are linear maps, but since here the monoidal product is the usual tensor product, the sizes of these matrices vary unless they are all 1×1. Write $s_{jk} := \dim \mathrm{dom}\, \mathbf{C}(j,k)$ and $t_{jk} := \dim \mathrm{cod}\, \mathbf{C}(j,k)$. Then s and t must satisfy (1).

If (j,k) and (k,ℓ) are part of a cycle, then so is (j,ℓ), which yields a simple proposition.

Proposition 3. *The system* (1) *does not admit solutions over $\mathbb{Z}_+$ that are nontrivial on arcs that can reach a strong component of D.* $\square$

Note that unless D is a DAG, it has nontrivial strong components. Except in degenerate cases, we can perform "Kronecker division" of appropriately sized matrices, which leads to the following corollary.

Corollary 1. *If D is a finite DAG, P is a spanning polytree (i.e., a digraph whose image under U is a tree) of D, and s, t satisfy* (1) *over $\mathbb{Z}_+$, an $\mathrm{Arr}(\mathbf{FinVect}_S)$-category is specified by a nondegenerate element of $\prod_{(j,j') \in A(P)} M_{(s_{(j,j')}, t_{(j,j')})}(S)$, and every $\mathrm{Arr}(\mathbf{FinVect}_S)$-category with underlying category determined by D arises in this way.*

In the present setting over the base field $S = \mathbb{C}$, the size maps that are also norms are commonly called *cross norms*. The permutation-invariant size maps for matrices over $\mathbb{C}$ are precisely the Schatten p-norms $\|\cdot\|_p$, i.e., the ℓ_p norms of the vector whose entries are the singular values of a matrix [3]. [14]

As in §4.1, replacing the arc matrices with sizes factors through to the scalar case of §3 all over again.

4.3   Stochastic matrices

The constructions of §4.1 and §4.2 trivially specialize to the case where the matrices involved are (row)-stochastic. [15] There is not an information-theoretical size map that can take the place of the determinant

---

[11]Although from the perspective of magnitude our constructions trivialize for unitaries, we can consider the Hermitian matrices obtained via matrix logarithms. This requires considering instead the monoidal product defined by the so-called *Kronecker sum* $H \boxplus H' := H \otimes I_{H'} + I_H \otimes H'$ that satisfies $\exp(H) \otimes \exp(H') = \exp(H \boxplus H')$, where we emphasize here that the matrix exponential $\exp(\cdot)$ is indicated instead of the componentwise exponential $\exp[\cdot]$.

[12]See, e.g., exercise 4 on p. 165 of [27] for a more general result.

[13]Some of our development applies to the category of semimodules over a semiring (= rig), but essential parts require working over $\mathbf{FinVect}_S$ for $S \in \{\mathbb{R}, \mathbb{C}\}$.

[14]The norms for $p = 1, 2, \infty$ respectively are called trace/nuclear; Frobenius/Hilbert-Schmidt; and operator/spectral. It is the case that $\|S\|_\infty \leq \|S\|_2 \leq \|S\|_1$. The relevant theory also owes much to both Grothendieck [33] and von Neumann [35].

[15]Note however that the inverse of a nonnegative stochastic matrix must contain negative entries, so in either case only DAGs are obvious candidates for applications. The category $\mathbf{FinStoch}$ whose objects and morphisms are respectively finite sets and stochastic maps represented as row-stochastic matrices is discussed in [16].

for an analogue of §4.1, as footnote 10 points out. However, there is a meaningful information-theoretical size map that produces a nontrivial analogue of §4.2, viz. $\mathbf{C}(j,k) \mapsto \exp C(\mathbf{C}(j,k))$, where C indicates the *channel capacity* [37]. [16]

Example 4. *A discrete memoryless channel (DMC) is specified by a matrix of conditional probabilities $W_{jk} = \mathbb{P}(y_k|x_j)$, where x_j and y_k respectively denote input and output symbols corresponding to realizations of random variables X and Y. W_{jk} is the probability that if Alice transmits x_j, then Bob receives y_k. An input distribution $p_j = \mathbb{P}(x_j)$ yields joint probabilities $V_{jk} := \mathbb{P}(x_j, y_k) = p_j W_{jk}$. The channel coding theorem states that reliable communication is possible iff the ratio of informative to transmitted bits is less than the channel capacity $C := \sup_p I(X;Y)$, where $I(X;Y)$ is the mutual information.*

Let D be given with arc data as in Figure 9, and where $W^{(i)}$ indicates the matrix for a DMC with corresponding capacity $\log c_i$. The resulting similarity matrix is

$$Z = \begin{pmatrix} 1 & 0 & c_1 c_3 & c_1 & c_1 c_3 c_4 & c_1 c_3 c_5 \\ 0 & 1 & c_2 c_3 & c_2 & c_2 c_3 c_4 & c_2 c_3 c_5 \\ 0 & 0 & 1 & 0 & c_4 & c_5 \\ 0 & 0 & 0 & 1 & c_3 c_4 & c_3 c_5 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and the equation $Zw = 1$ can be solved by hand to yield

$$w = (1 - [1-c_4]c_1 c_3 - [1-c_3 c_5]c_1, 1 - [1-c_5]c_2 c_3 - [1-c_3 c_4]c_2, 1 - c_4 - c_5, 1 - c_3 c_4 - c_3 c_5, 1, 1)^T.$$

There is an asymmetry in form between $w_3 = 1 - c_4 - c_5$ and $w_4 = 1 - c_3 c_4 - c_3 c_5$. This reflects the fact that even though the end-to-end capacity is path-independent, the capacity of channels to node 3 differs from the capacity of channels to node 4. Note that if we switch the channels from vertex 2 with each other and the channels to vertex 6 with each other that this asymmetry disappears.
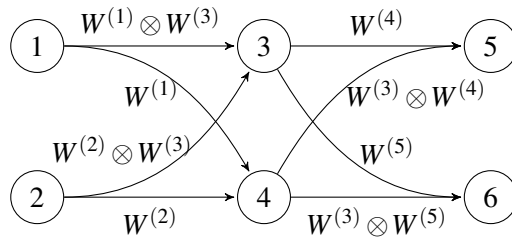

Figure 9: A **FinStoch**-category (with nontrivial compositions suppressed for clarity).

---

[16] For background on information theory, see [11]. B. Fong and D. I. Spivak showed in a personal communication (2019) that channel capacity is a strict monoidal lax functor $(\mathbf{FinStoch}, \otimes) \to ([0,\infty], (\infty, \min), \geq, (+, 0))$, where we indicate the lax monoidal po-monoid with one object $*$, with $\text{Hom}(*, *) = [0, \infty]$, with identity ∞ and composition given by min, with local partial order (po) structure given by the usual $\geq$, and with monoidal structure given by $+$ and 0. This result also holds when replacing both i) the Kronecker/tensor $\otimes$ monoidal structure on **FinStoch** with the direct sum $\oplus$ monoidal structure and ii) channel capacity C with its exponent $\exp C$.

4.3.1 Sidebar on channel capacity and coweightings

In the spirit of understanding and applying magnitude in atypical contexts, it can be instructive to consider linear equations $Zw = 1$ and/or $vZ = 1^T$ where the dissimilarity matrix Z cannot be written in the form $\exp[-td]$ for any Lawvere metric d. For instance, the capacity-achieving input distribution of a DMC with invertible channel matrix W can be thought of as such a normalized coweighting. The key to realizing this is a classical formula due to Muroga [30].

Let W be an invertible channel matrix with $M := W^{-1}$ and define $H_j := -\sum_k W_{jk} \log W_{jk}$. Furthermore, define $v_j := \sum_i M_{ij} \exp\left(-\sum_k M_{ik} H_k\right)$. The Muroga formula states that if $v > 0$, then $e^C = \sum_j \exp\left(-\sum_k M_{jk} H_k\right)$, and $p := e^{-C} v$ is a capacity-achieving distribution. Now $v = \exp(-MH)M = 1^T \Delta(\exp(-MH))M$, so writing $Z := W\Delta(\exp(MH))$, we obtain $vZ = 1^T$. However, the putative distance $-t^{-1}[\log W_{jk} - (MH)_k]$ is generically negative on the diagonal, though also approximately equal to a constant times $1 - \delta_{jk}$ for a "good channel" $W \approx I$. The better the DMC is, the closer its induced "metric geometry" (where the negative diagonal entails the quotation marks) is to that of a regular simplex.

5 Remarks on "flow-like" graphs augmented with data

An interesting if still speculative possibility for deeper applications to processes, computer programs, etc. is to consider categories of the form $\mathbf{F} \times \mathbf{M}$ where $\mathbf{F}$ is a suitable category of "flow-like" digraphs that have single inputs and outputs and $\mathbf{M}$ is a suitable category of data (perhaps scalar if not in the spirit of a nondegenerate matrix category). The monoidal structure on $\mathbf{F}$ is simply gluing the output of one 'flow-like' digraph to the input of another along the lines of [18]. That is, each "flow-like" graph has associated data that compose nicely under series composition. A specific instantiation of $\mathbf{F}$ that is algorithmically and categorically well-behaved (in particular, semicartesian) is likely obtainable via minor technical modifications to the notion of a *two-terminal graph* [39].

In order to have a useful notion of magnitude in this context, it is necessary to have a size function (or for magnitude homology [25], a strong symmetric monoidal functor from a semicartesian symmetric monoidal category to a suitable abelian symmetric monoidal category) that consolidates the graphical and matrix data. A good candidate for constructing something in this vein appears to be the zeta function of a finite Markov chain [32] that encodes (e.g.) graph traversal probabilities.

Acknowledgement

References

[1] David J Aldous (1990): *The random walk construction of uniform spanning trees and uniform labelled trees.* SIAM Journal on Discrete Mathematics 3(4), pp. 450–465.

[2] Panos J Antsaklis & Anthony N Michel (1997): *Linear Systems.* Springer.

[3] Guillaume Aubrun & Ion Nechita (2011): *The multiplicative property characterizes ℓ_p and L_p norms.* Confluentes Mathematici 3(04), pp. 637–647.

[4] Georgios Bakirtzis, Cody H Fleming & Christina Vasilakopoulou (2021): *Categorical semantics of cyber-physical systems theory*. ACM Transactions on Cyber-Physical Systems 5(3), pp. 1–32.

[5] Jørgen Bang-Jensen & Gregory Z Gutin (2008): *Digraphs: Theory, Algorithms and Applications*. Springer.

[6] Eric Bunch et al. (2020): *Practical applications of metric space magnitude and weighting vectors*. Available at `https://arxiv.org/abs/2006.14063`.

[7] Pavel Chebotarev & Elena Deza (2020): *Hitting time quasi-metric and its forest representation*. Optimization Letters 14(2), pp. 291–307.

[8] Simon Cho (2019): *Quantales, persistence, and magnitude homology*. arXiv preprint arXiv:1910.02905.

[9] Joseph Chuang, Alastair King & Tom Leinster (2016): *On the magnitude of a finite dimensional algebra*. Theory and Applications of Categories 31(3), pp. 63–72.

[10] Keith D Cooper & Linda Torczon (2011): *Engineering a Compiler*. Elsevier.

[11] Thomas M Cover & Joy A Thomas (2006): *Elements of Information Theory*. Wiley-Interscience.

[12] Michel Marie Deza & Elena Deza (2009): *Encyclopedia of Distances*. Springer.

[13] Aristophanes Dimakis & Folkert Muller-Hoissen (1994): *Differential calculus and gauge theory on finite sets*. Journal of Physics A: Mathematical and General 27(9), p. 3159.

[14] Aristophanes Dimakis & Folkert Müller-Hoissen (1994): *Discrete differential calculus: Graphs, topologies, and gauge theory*. Journal of Mathematical Physics 35(12), pp. 6703–6735.

[15] Brendan Fong & David I Spivak (2019): *An Invitation to Applied Category Theory: Seven Sketches in Compositionality*. Cambridge.

[16] Tobias Fritz (2020): *A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics*. Advances in Mathematics 370, p. 107239.

[17] Ian Goodfellow, Yoshua Bengio & Aaron Courville (2016): *Deep Learning*. MIT.

[18] Steve Huntsman (2022): *Magnitude and topological entropy of digraphs*. In: Proceedings of Applied Category Theory.

[19] Steve Huntsman (2023): *Diversity enhancement via magnitude*. In: Proceedings of Evolutionary Multi-Criterion Optimization.

[20] Gerald J Janusz (1996): *Algebraic Number Fields*. AMS.

[21] Gregory Maxwell Kelly (1982): *Basic Concepts of Enriched Category Theory*. Cambridge.

[22] Douglas J Klein & H-Y Zhu (1998): *Distances and volumina for graphs*. Journal of Mathematical Chemistry 23(1-2), pp. 179–195.

[23] Tom Leinster (2021): *Entropy and Diversity: the Axiomatic Approach*. Cambridge.

[24] Tom Leinster & Mark W Meckes (2017): *The magnitude of a metric space: from category theory to geometric measure theory*. In Nicola Gigli, editor: Measure Theory in Non-Smooth Spaces, De Gruyter, doi:10.1515/9783110550832-005.

[25] Tom Leinster & Michael Shulman (2021): *Magnitude homology of enriched categories and metric spaces*. Algebraic & Geometric Topology 21(5), pp. 2175–2221.

[26] Daniel Liberzon (2003): *Switching in Systems and Control*. Birkhäuser.

[27] Saunders Mac Lane (2013): *Categories for the Working Mathematician*. Springer.

[28] Juan Maldacena (2015): *The symmetry and simplicity of the laws of physics and the Higgs boson*. European Journal of Physics 37(1), p. 015802.

[29] Mark W Meckes (2015): *Magnitude, diversity, capacities, and dimensions of metric spaces*. Potential Analysis 42(2), pp. 549–572.

[30] Saburo Muroga (1953): *On the capacity of a discrete channel. I*. Journal of the Physical Society of Japan 8(4), pp. 484–494.

[31] Michael A Nielsen & Isaac L Chuang (2010): *Quantum Computation and Quantum Information*.

[32] William Parry & Robert F Williams (1977): *Block coding and a zeta function for finite Markov chains*. Proceedings of the London Mathematical Society 3(3), pp. 483–495.

[33] Gilles Pisier (2012): *Grothendieck's theorem, past and present*. Bulletin of the American Mathematical Society 49(2), pp. 237–323.

[34] Nornadiah Mohd Razali, Yap Bee Wah et al. (2011): *Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests*. Journal of Statistical Modeling and Analytics 2(1), pp. 21–33.

[35] Robert Schatten & John von Neumann (1948): *The cross-space of linear transformations. III*. Annals of Mathematics, pp. 557–582.

[36] Ralf Schiffler (2014): *Quiver Representations*. Springer.

[37] Claude E Shannon (1957): *Certain results in coding theory for noisy channels*. Information and Control 1(1), pp. 6–25.

[38] David I Spivak (2015): *The steady states of coupled dynamical systems compose according to matrix arithmetic*. arXiv preprint arXiv:1512.00802.

[39] Jussi Vanhatalo, Hagen Völzer & Jana Koehler (2008): *The refined process structure tree*. In: Business Process Management: 6th International Conference, BPM 2008, Milan, Italy, September 2-4, 2008. Proceedings 6, Springer, pp. 100–115.

[40] Nikita Visnevski, Vikram Krishnamurthy, Alex Wang & Simon Haykin (2007): *Syntactic modeling and signal processing of multifunction radars: A stochastic context-free grammar approach*. Proceedings of the IEEE 95(5), pp. 1000–1025.

[41] Chun-Chao Wang (2015): *A MATLAB package for multivariate normality test*. Journal of Statistical Computation and Simulation 85(1), pp. 166–188.

[42] Simon Willerton (2009): *Heuristic and computer calculations for the magnitude of metric spaces*. arXiv preprint arXiv:0910.5500.