Towards a Unified Theory of Time-Varying Data

Wilmer Leal

Gatas Lab
Department of Computer Science
University of Florida, USA

Joint work with:
Benjamin Merlin Bumpus, James Fairbanks, Martti Karvonen & Frédéric Simard.

ACT 2024

$$\begin{tikzcd} G \arrow[r] \arrow[d] \arrow[dr, phantom, "\Downarrow\tau"] & A \arrow[d] \\ A \arrow[r] & S \end{tikzcd}$$

Category theory for data analysis[1]

- Object-based presentations in temporal data theories.
  Categories of temporal data.
- A temporal data theory for each kind of mathematical structure?
  Object-agnosticism.
- How are the cumulative and persistent Perspectives related?
  Formally related via an adjuntion.
- Can we systematically lift static properties to temporal ones?
  Systematic "Temporalization".
- Can we formally study connections between temporal data and dynamical systems?

[1] Based on: *Towards a Unified Theory of Time-varying Data*. BM Bumpus, J Fairbanks, M Karvonen, W Leal, F Simard. arXiv:2402.00206

Question 1: What are the data structures for temporal data?

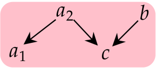
A huge, growing number of data sets generated by the underlying dynamics, e.g.:

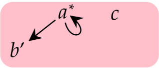
- human and animal proximity networks,
- human communication networks,
- collaboration networks,
- economic networks,
- biological, chemical networks,
- and epidemiological networks

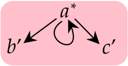
Question 1: What are the data structures for temporal data?

A huge, growing number of data sets generated by the underlying dynamics, e.g.:

- human and animal proximity networks,
- human communication networks,
- collaboration networks,
- economic networks,
- biological, chemical networks,
- and epidemiological networks





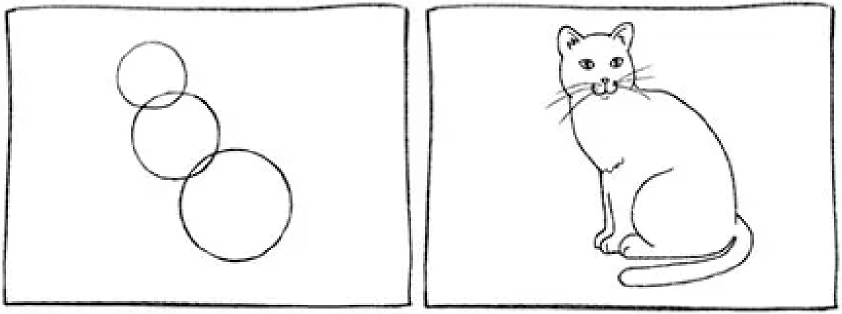


Question 1: What are the data structures for temporal data?

Let me teach you how to draw a cat!

Question 1: What are the data structures for temporal data?

Let me teach you how to draw a cat!



Easy, right?

Question 1: What are the data structures for temporal data?

We can store many kinds of data transformation using morphisms between snapshots:

$$
\begin{tikzcd}
& \{a_1, a_2, c\} \arrow[dl] \arrow[dr] & & \{a, b_2\} \arrow[dl] \arrow[dr] & \\
\{a_1, a_2, b_1, c\} & & \{a, b_2, c\} & & \{a, b_2, c'\}
\end{tikzcd}
$$

Narrative: Company a_1 and a_2 merged into a, which remained opened. Company b_1 disappeared in the second snapshot, while c remained open, etc.

Question 1: What are the data structures for temporal data?

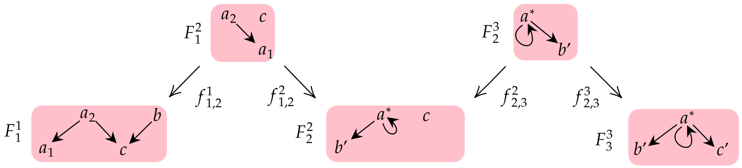


Figure: Relational data for companies as graphs

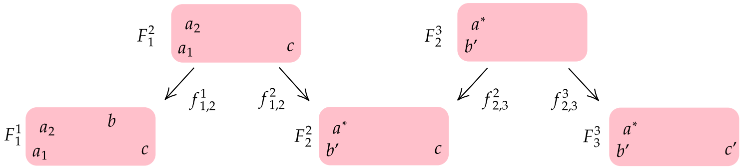


Figure: Distances between companies as metric spaces

Question 1: What are the data structures for temporal data?

We should be able to compute long-term relationships: Which companies $\{a_1, a_2, b_1, c_1\}$ persisted over the course of three years?

$$
\begin{tikzcd}
 & & F_1^3 = \{(a_1,a),(a_2,a)\} \arrow[dl] \arrow[dr] \arrow[dd, phantom, very near start, "\vee"] & & \\
 & F_1^2 = \{a_1,a_2,c_1\} \arrow[dl] \arrow[dr] & & F_2^3 = \{a,b_2\} \arrow[dl] \arrow[dr] & \\
F_1^1 := \{a_1,a_2,b_1,c_1\} & & F_2^2 := \{a,b_2,c_1\} & & F_3^3 := \{a,b_2,c_2\}
\end{tikzcd}
$$

Just compute the pullback of the sets F_1^2 and F_2^3:

$$F_1^3 := \{(x,y) \in F_1^2 \times F_2^3 \mid f_{1,2}^2(x) = f_{2,3}^2(y)\}$$

.

Question 1: What are the data structures for temporal data?

We should be able to compute long-term relationships: Data (companies) and relationships accumulated (seen) over a period of time.

$$
\begin{tikzcd}
& & F_1^3 = \{a, b_1, b_2, c_1, c_2\} & & \\
& F_1^2 = \{a, b_1, b_2, c_1\} \arrow[ur, hook] & & F_2^3 = \{a, b_2, c_1, c_2\} \arrow[ul, hook'] & \\
F_1^1 := \{a_1, a_2, b_1, c_1\} \arrow[ur] & & F_2^2 := \{a, b_2, c_1\} \arrow[ul, hook'] \arrow[ur, hook] & & F_3^3 := \{a, b_2, c_2\} \arrow[ul, hook']
\end{tikzcd}
$$

Compute the pushout of the diagram involving sets F_1^2, F_2^2 and F_2^3.

Question 1: What are the data structures for temporal data?

This temporal data structure can be seen as a functor defined on certain time category:

$$
\begin{tikzcd}[column sep=small]
 & & {[1,3]} & & \\
 & {[1,2]} \arrow[ur, hook] & & {[2,3]} \arrow[ul, hook'] & \\
{[1,1]} \arrow[ur, hook] & & {[2,2]} \arrow[ul, hook'] \arrow[ur, hook] & & {[3,3]} \arrow[ul, hook']
\end{tikzcd}
\xrightarrow{F}
\begin{tikzcd}[column sep=small]
 & & F_1^3 \arrow[dl] \arrow[dr] \arrow[dd, phantom, "\lrcorner", very near start] & & \\
 & F_1^2 \arrow[dl] \arrow[dr] & & F_2^3 \arrow[dl] \arrow[dr] & \\
F_1^1 & & F_2^2 & & F_3^3
\end{tikzcd}
$$

With the additional condition that $F([a, b])$ is the pullback $F([a, p]) \times_{F([p,p])} F([p, b])$ for any $p \in [a, b]$.

Question 1: What are the data structures for temporal data?

We borrow interval categories from Schultz, Spivak and Vasilakopoulou:

Definition (Interval categories[2])

The **category of intervals**, denoted Int, is the category having closed intervals $[a, b]$ in $\mathbb{R}+$ (the non-negative reals) as objects and orientation-preserving isometries as morphisms. Analogously, one can define the category $\mathrm{Int}_{\mathbb{N}}$ of **discrete intervals** by restricting only to $\mathbb{N}$-valued intervals.
These categories can be turned into sites by equipping them with the Johnstone coverage which stipulates that a **cover** of any interval $[a, b]$ is a partition into two closed intervals $([a, p], [p, b])$.

[2]Schultz, P., Spivak, D. I., and Vasilakopoulou, C. Dynamical systems and sheaves. Applied Categorical Structures 28, 1 (2020), 1–57.

Question 1: What are the data structures for temporal data?

We borrow interval categories from Schultz, Spivak and Vasilakopoulou:

Definition (Strict Embedding Intervals)

We denote by I *(resp.* $\mathsf{I}_{\mathbb{N}}$*) the full subcategory (specifically a join-semilattice) of the subobject poset of* $\mathbb{R}$ *(resp.* $\mathbb{N}$*) whose objects are intervals. We will refer to* I, $\mathsf{I}_{\mathbb{N}}$ *and any sub-join-semilattices thereof as* **time categories**.

The following lemma states that time categories can be given Grothendieck topologies in much the same way as the interval categories:

Lemma

Any time category forms a site when equipped with the Johnstone coverage.

Question 1: What are the data structures for temporal data?

Now we are ready to give the definition of a sheaf with respect to any of the sites described in the previous Lemma.

Proposition (T-sheaves)

Let T *be any time category equipped with the Johnstone coverage. Suppose* D *is a category with pullbacks, then a* D-**valued sheaf on** T *is a presheaf* $F\colon \mathsf{T}^{op} \to \mathsf{D}$ *satisfying the following additional condition: for any interval* $[a,b]$ *and any cover* $([a,p],[p,b])$ *of this interval,* $F([a,b]) = F([a,p]) \times_{F([p,p])} F([p,b])$.

$$
\begin{tikzcd}
 & & {[1,3]} & & & & & & F_1^3 \arrow[dl] \arrow[dr] \arrow[dd, phantom, "\lrcorner", very near start] & & \\
 & {[1,2]} \arrow[ur, hook] & & {[2,3]} \arrow[ul, hook'] \arrow[rrrr, "F"] & & & & F_1^2 \arrow[dl] \arrow[dr] & & F_2^3 \arrow[dl] \arrow[dr] & \\
{[1,1]} \arrow[ur, hook] & & {[2,2]} \arrow[ul, hook'] \arrow[ur, hook] & & {[3,3]} \arrow[ul, hook'] & & F_1^1 & & F_2^2 & & F_3^3
\end{tikzcd}
$$

Question 1: What are the data structures for temporal data?

Similarly, we can define the temporal data structure that can encode accumulated data and relationships using cosheaves:

Proposition (T-cosheaves)

Let T *be any time category equipped with the Johnstone coverage. Suppose* D *is a category with pushouts, then a* D-**valued cosheaf on** T *is a copresheaf* $\hat{F}\colon \mathsf{T} \to \mathsf{D}$ *satisfying the following additional condition: for any interval* $[a,b]$ *and any cover* $([a,p],[p,b])$ *of this interval,* $\hat{F}([a,b])$ *is the pushout* $\hat{F}([a,p]) +_{\hat{F}([p,p])} \hat{F}([p,b])$.

$$\begin{tikzcd} & & {[1,3]} & & \\ & {[1,2]} \arrow[ur, hook] & & {[2,3]} \arrow[ul, hook'] & \\ {[1,1]} \arrow[ur, hook] & & {[2,2]} \arrow[ul, hook'] \arrow[ur, hook] & & {[3.3]} \arrow[ul, hook'] \end{tikzcd} \xrightarrow{\quad\hat{F}\quad} \begin{tikzcd} & & \hat{F}_1^3 & & \\ & \hat{F}_1^2 \arrow[ur] & & \hat{F}_2^3 \arrow[ul] & \\ \hat{F}_1^1 \arrow[ur] & & \hat{F}_2^2 \arrow[ul] \arrow[ur] & & \hat{F}_3^3 \arrow[ul] \end{tikzcd}$$

Question 1: What are the data structures for temporal data?

Definition (Persistent and cumulative)

We denote by $\mathrm{Pe}(T, \mathrm{D})$ (resp. $\mathrm{Cu}(T, \mathrm{D})$) the category of D-valued sheaves (resp. cosheaves) on T and we call it the category of **persistent** D-**narratives** (resp. **cumulative** D-**narratives**) with T-time.

$$
\begin{tikzcd}
& & {[1,3]} & & & & & & F_1^3 \arrow[dl] \arrow[dr] \arrow[dd, phantom, "\lrcorner"{anchor=center, pos=0.125, rotate=-45}] & & \\
& {[1,2]} \arrow[ur, hook] & & {[2,3]} \arrow[ul, hook'] & {} \arrow[rr, "F"] & & {} & F_1^2 \arrow[dl] \arrow[dr] & & F_2^3 \arrow[dl] \arrow[dr] & \\
{[1,1]} \arrow[ur, hook] & & {[2,2]} \arrow[ul, hook'] \arrow[ur, hook] & & {[3,3]} \arrow[ul, hook'] & & F_1^1 & & F_2^2 & & F_3^3
\end{tikzcd}
$$

$$
\begin{tikzcd}
& & {[1,3]} & & & & & & \hat{F}_1^3 \arrow[dd, phantom, "\lrcorner"{anchor=center, pos=0.125, rotate=-45}] & & \\
& {[1,2]} \arrow[ur, hook] & & {[2,3]} \arrow[ul, hook'] & {} \arrow[rr, "\hat{F}"] & & {} & \hat{F}_1^2 \arrow[ur] & & \hat{F}_2^3 \arrow[ul] & \\
{[1,1]} \arrow[ur, hook] & & {[2,2]} \arrow[ul, hook'] \arrow[ur, hook] & & {[3.3]} \arrow[ul, hook'] & & \hat{F}_1^1 \arrow[ur] & & \hat{F}_2^2 \arrow[ul] \arrow[ur] & & \hat{F}_3^3 \arrow[ul]
\end{tikzcd}
$$

Question 2: How are the persistent and cumulative perspectives related?

Theorem

Let D *be category with limits and colimits.*
There exist functors $\mathscr{P}\colon \mathsf{Cu}(\mathsf{T},\mathsf{D}) \to \mathsf{Pe}(\mathsf{T},\mathsf{D})$ *and* $\mathscr{K}\colon \mathsf{Pe}(\mathsf{T},\mathsf{D}) \to \mathsf{Cu}(\mathsf{T},\mathsf{D})$.
Moreover, these functors are adjoint to each other:

$$\begin{tikzcd} \mathsf{Cu}(\mathsf{T},\mathsf{D}) \arrow[r, bend left, "\mathscr{P}"] \arrow[r, phantom, "\perp"] & \mathsf{Pe}(\mathsf{T},\mathsf{D}) \arrow[l, bend left, "\mathscr{K}"] \end{tikzcd}$$

Proof.

Define $\mathscr{P}$ to be the map that assigns to any cosheaf $\hat{F}\colon T \to \mathsf{D}$ the sheaf $\mathscr{P}(\hat{F})\colon T^{op} \to \mathsf{D}$ defined on objects by:

$$\mathscr{P}(\hat{F})\colon [a,b] \mapsto \lim(\mathsf{T}(-,[a,b]) \hookrightarrow T \xrightarrow{\hat{F}} \mathsf{D}).$$

$\square$

Question 2: How are the persistent and cumulative perspectives related?

Theorem

Let D *be category with limits and colimits.*
There exist functors $\mathscr{P}\colon \mathsf{Cu}(\mathsf{T},\mathsf{D}) \to \mathsf{Pe}(\mathsf{T},\mathsf{D})$ *and* $\mathscr{K}\colon \mathsf{Pe}(\mathsf{T},\mathsf{D}) \to \mathsf{Cu}(\mathsf{T},\mathsf{D})$.
Moreover, these functors are adjoint to each other:

$$\begin{tikzcd} \mathsf{Cu}(\mathsf{T},\mathsf{D}) \arrow[r, bend left, "\mathscr{P}"] \arrow[r, phantom, "\perp"] & \mathsf{Pe}(\mathsf{T},\mathsf{D}) \arrow[l, bend left, "\mathscr{K}"] \end{tikzcd}$$

Proof.

Define $\mathscr{K}\colon \mathsf{Pe}(\mathsf{T},\mathsf{D}) \to \mathsf{Cu}(\mathsf{T},\mathsf{D})$ as the map that takes any sheaf $F\colon T^{op} \to \mathsf{D}$ to the cosheaf $\mathscr{K}(F)\colon T \to \mathsf{D}^{op}$ defined on objects by:

$$\mathscr{K}(F)\colon [a,b] \mapsto \mathrm{colim}(\mathsf{T}(-,[a,b]) \hookrightarrow T \xrightarrow{F} \mathsf{D}).$$

$\square$

Temporal networks can be written as narratives

Definition (Temporal networks[3])

Take $\mathbb{T}$ to be either $\mathbb{N}$ or $\mathbb{R}$. A $\mathbb{T}$-temporal (directed) network is a quintuple $(G, \rho_e, \eta_e, \rho_v, \eta_v)$ where G is a (directed) graph and ρ_e, η_e, ρ_v and η_v are functions of the following types:

$$\rho_e : E(G) \times \mathbb{T} \to \{\perp, \top\}, \qquad \eta_e : E(G) \times \mathbb{T} \to \mathbb{T},$$
$$\rho_v : V(G) \times \mathbb{T} \to \{\perp, \top\}, \qquad \eta_v : V(G) \times \mathbb{T} \to \mathbb{T}$$

where ρ_e and ρ_v are are functions indicating whether an edge or vertex is active at a given time and where η_e and η_v are latency functions indicating the amount of time required to traverse an edge or vertex.

[3] Casteigts, Flocchini, Quattrociocchi and Santoro. Time-varying graphs and dynamic networks. International Journal of Parallel, Emergent and Distributed Systems 27, 5 (2012), 387–408.

Temporal networks can be written as narratives

- Consider the monoid of natural numbers viewed as a single-vertex graph with a loop edge for each natural number $G_{B\mathbb{N}} : \mathsf{SGr} \to \mathsf{Set}$ having $G_{B\mathbb{N}}(V) = 1$ and $G_{B\mathbb{N}}(E) = \mathbb{N}$
- The slice category $\mathsf{Set}^{\mathsf{SGr}}/G_{B\mathbb{N}}$ will have pairs $(G, \lambda\colon G \to G_{B\mathbb{N}})$ as objects where G is a graph and λ is a graph homomorphism that assigns a natural number label to each edge of G.
- Thus narratives valued in $\mathsf{Set}^{\mathsf{SGr}}/G_{B\mathbb{N}}$ can be interpreted as time-varying graphs whose edges come equipped with **latencies** (which can change with time).

Temporal networks can be written as narratives

- Consider the monoid of natural numbers viewed as a single-vertex graph with a loop edge for each natural number $G_{B\mathbb{N}} : \mathsf{SGr} \to \mathsf{Set}$ having $G_{B\mathbb{N}}(V) = 1$ and $G_{B\mathbb{N}}(E) = \mathbb{N}$
- The slice category $\mathsf{Set}^{\mathsf{SGr}}/G_{B\mathbb{N}}$ will have pairs $(G, \lambda\colon G \to G_{B\mathbb{N}})$ as objects where G is a graph and λ is a graph homomorphism that assigns a natural number label to each edge of G.
- Thus narratives valued in $\mathsf{Set}^{\mathsf{SGr}}/G_{B\mathbb{N}}$ can be interpreted as time-varying graphs whose edges come equipped with **latencies** (which can change with time).

Many other kinds of temporal graphs and networks, including reflexive graphs, symmetric-and-reflexive graphs, half-edge graphs and Petri nets, can be constructed using this approach.

Question 3: How to change the base category?

Proposition (Covariant Change of base)

Let C *and* D *be categories with limits (resp. colimits) and let* T *be any time category. If* $K \colon \mathsf{C} \to \mathsf{D}$ *is a continuous functor, then composition with* K *determines a functor* $(K \circ -)$ *from persistent (resp. cumulative)* C*-narratives to persistent (resp. cumulative)* D*-narratives. Spelling this out explicitly for the case of persistent narratives, we have:*

$$
\begin{aligned}
&(K \circ -) \colon \mathsf{Pe}(\mathsf{T}, \mathsf{C}) \to \mathsf{Pe}(\mathsf{T}, \mathsf{D}) \\
&(K \circ -) \colon (F \colon T^{op} \to \mathsf{C}) \mapsto (K \circ F \colon T^{op} \to \mathsf{D}).
\end{aligned}
$$

Application: Defining temporal counterparts of static properties

- Any class P of objects in C can be identified with a subcategory

  $$P \colon \mathsf{P} \to \mathsf{C}$$

  this functor picks out those objects of C that satisfy a given property P.
- If this functor P is cocontinuous, then we can apply our previous proposition to identify a class

  $$(P \circ -) \colon \mathsf{Cu}(\mathsf{T}, \mathsf{P}) \to \mathsf{Cu}(\mathsf{T}, \mathsf{C})$$

  of C-narratives which satisfy the property P at all times.
- For example, consider the full subcategory $\mathfrak{P} \colon \mathsf{Paths} \hookrightarrow \mathsf{Grph}$ which defines the category of all paths and the morphisms between them. $\mathfrak{P}$ determines a subcategory $(\mathfrak{P} \circ -) \colon \mathsf{Cu}(T, \mathsf{Paths}) \hookrightarrow \mathsf{Cu}(T, \mathsf{Grph})$ whose objects are temporal path-graphs.

Application: the cumulative temporal tree is equivalent to the persistent temporal path problem

$$\begin{tikzcd} \mathsf{Cu}(\mathsf{T}, \mathsf{Grph}_{mono}) \arrow[r, bend left, "\mathscr{P}"] \arrow[r, phantom, "\perp"] & \mathsf{Pe}(\mathsf{T}, \mathsf{Grph}_{mono}) \arrow[l, bend left, "\mathscr{K}"] \end{tikzcd}$$

The change of base proposition applied to the full subcategory

$$\mathfrak{T} \colon \mathsf{Trees}_{mono} \to \mathsf{Grph}_{mono}$$

yields:

$$\begin{tikzcd} \mathsf{Cu}(\mathsf{T}, \mathsf{Trees}_{mono}) \arrow[r, "(\mathfrak{T} \circ -)"] & \mathsf{Cu}(\mathsf{T}, \mathsf{Grph}_{mono}) \\ \mathsf{Pe}(\mathsf{T}, \mathsf{Paths}_{mono}) \arrow[r, "(\mathfrak{P} \circ -)"] & \mathsf{Pe}(\mathsf{T}, \mathsf{Grph}_{mono}) \arrow[u, "\mathscr{K}"] \end{tikzcd}$$

Taking the pullback yields a category with pairs (T, P) as objects: a cumulative tree narrative T and a persistent path narrative P such that, when both are viewed as cumulative Grph_{mono}-narratives, they give rise to the *same* narrative.

Conclusions[4]

- **Categories of temporal data**: We have introduced a data structure for temporal data, *categories of narratives*.
- **Object-agnosticism**: All kind of mathematical structures representing data can be encoded as narratives, provided they have limits or colimits.
- **Cumulative and Persistent Perspectives**: Formally related via an adjuntion.
- **Systematic “Temporalization”**: Can lift static properties to temporal ones.
- **Sampling**: our framework of temporal opens the possibility of formally studying connections between temporal data and dynamical systems.

[4] Based on: *Towards a Unified Theory of Time-varying Data*. BM Bumpus, J Fairbanks, M Karvonen, W Leal, F Simard. arXiv:2402.00206