Towards a Unified Theory of Time-Varying Data

Benjamin Merlin Bumpus, James Fairbanks, Martti Karvonen, Wilmer Leal & Frédéric Simard

UNIVERSITY of FLORIDA

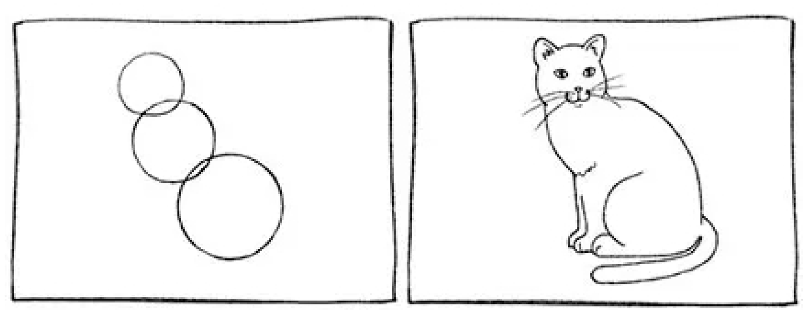
How does one build a robust and general theory of temporal data?

To address this question, we first draw inspiration from the theory of time-varying graphs. This theory has received considerable attention recently given the huge, growing number of data sets generated by underlying dynamics. Examples include human communication, collaboration, economic, biological, chemical networks, and epidemiological networks. We distill the lessons learned from temporal graph theory into the following set of *desiderata* for any mature theory of temporal data:

1. Categories of Temporal Data: Any theory of temporal data should define not only time-varying data, but also appropriate morphisms thereof.
2. Cumulative and Persistent Perspectives: In contrast to being a mere sequence, temporal data should explicitly record whether it is to be viewed cumulatively or persistently. Furthermore there should be methods of conversion between these two viewpoints.
3. Systematic 'Temporalization': Any theory of temporal data should come equipped with systematic ways of obtaining temporal analogues of notions relating to static data.
4. Object Agnosticism: Theories of temporal data should be object agnostic and applicable to any kinds of data originating from given underlying dynamics.
5. Sampling: Since temporal data naturally arises from some underlying dynamical system, any theory of temporal data should be seamlessly interoperable with theories of dynamical systems.

Building an object-agnostic data structure for temporal data

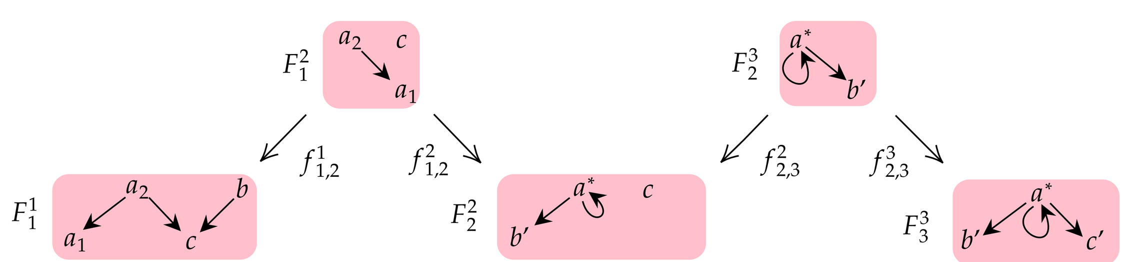
Narratives instead of mere sequences: Valuable information about the underlying dynamics is lost if temporal data is thought as a mere sequence of snapshots. As in the following sequence of instructions on how to draw a cat:



Data structures for temporal data should therefore be able to encode changes from one snapshot to another. *We can store many kinds of data transformation using morphisms between snapshots!* Consider the sets of cream companies that exist in a city each year:

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

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. This also works for other kinds of data, e.g., supplier (relational) data.



Computing long-term relationships: Which companies $\{a_1,a_2,b_1,c_1\}$ persisted over the course of three years? Just compute the pullback of the diagram involving sets F_1^2, F_2^2 and F_2^3.

$$\begin{tikzcd}[column sep=tiny] & & F_1^3 = \{(a_1,a),(a_2,a)\} \arrow[dl] \arrow[dr] \arrow[dd, phantom, "\lrcorner" rotate=-45] & & \\ & F_1^2 = \{a_1,a_2,c_1\} \arrow[dl] \arrow[dr] & & F_2^3 = \{a,b_2\} \arrow[dl] \arrow[dr] & \\ F_1^1 := \{a_1,a_2,b_1,c_1\} & & F_2^2 := \{a,b_2,c_1\} & & F_3^3 := \{a,b_2,c_2\} \end{tikzcd}$$

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

Dually, which data (companies) and relationships accumulated over the course of three years? Just compute the pushout of the diagram involving sets F_1^2, F_2^2 and F_2^3.

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

These data structures can be formalized as sheaves and cosheaves over time categories T (full subcategories of the subobject poset of $\mathbb{R}$ or $\mathbb{N}$ whose objects are intervals and whose morphism are inclusions). We call these data structures **persistent D-narratives** and **cumulative D-narratives** with T-time.

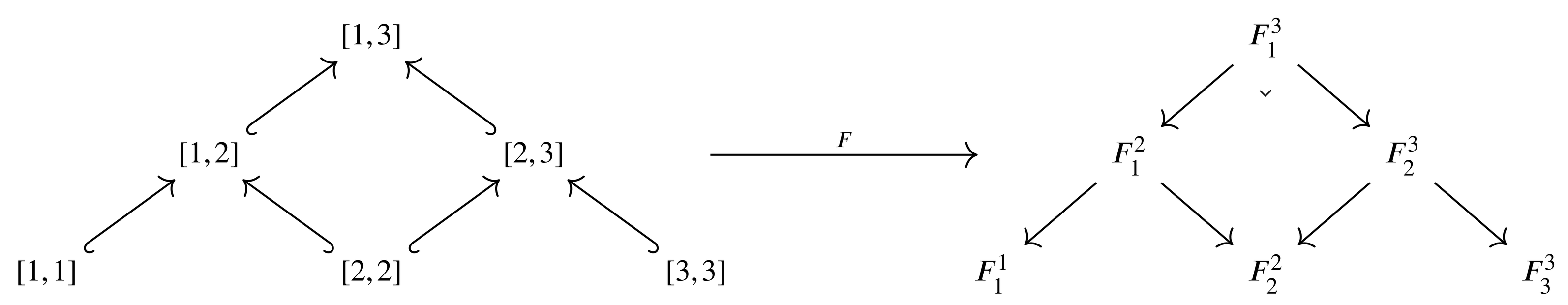
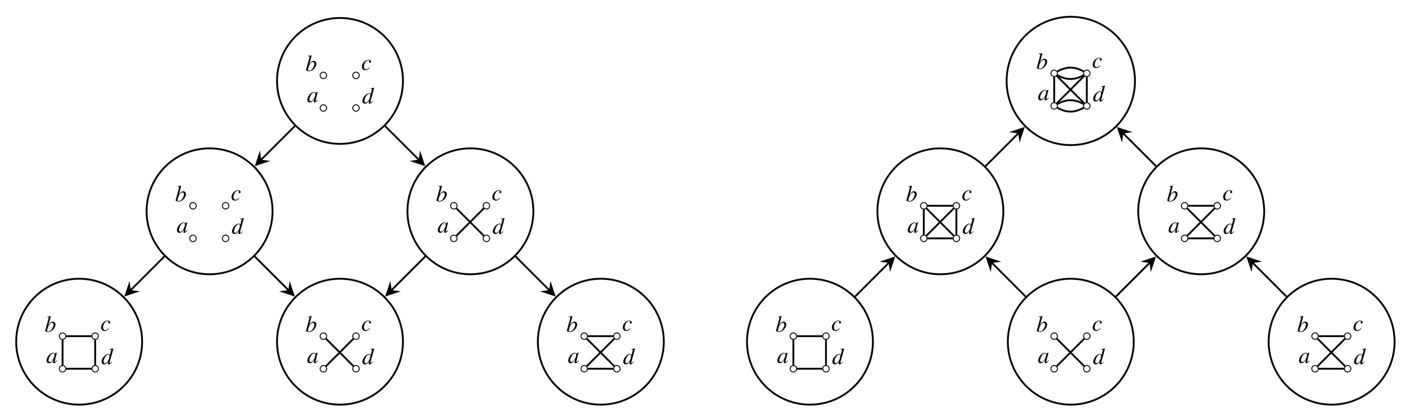


Figure 1. A schematic visualization of a sheaf on a discrete time category T (a persistent narrative) with three snapshots. The domain is a join-semi-lattice whose objects are intervals and whose morphisms are inclusions of intervals. The codomain of the sheaf is a category with pullbacks. We use the shorthand F_i^j (for $i \leq j$) to denote the data assigned the interval $[i,j]$ by F, i.e., $F_i^j := F([i,j])$.

Definition [Persistent and cumulative narratives]: *We denote by* $\mathsf{Pe}(\mathsf{T},\mathsf{D})$ *(resp.* $\mathsf{Cu}(\mathsf{T},\mathsf{D})$*) 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.*

Example [Temporal graphs]: *Persistent (left) and cumulative (right) narratives of temporal graphs.*



Adjunction between the cumulative and persistent perspectives

Theorem [Persistent vs cumulative]: *Let* D *be category with limits and colimits and* T *any time category. 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:*

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

Systematic 'Temporalization'

The next proposition tells us the kind of functors that can be used 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:*

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

This allow us to systematically define 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.

Example [Temporal paths]: *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* $(\mathfrak{P} \circ -)\colon \mathsf{Cu}(\mathsf{T},\mathsf{Paths}) \hookrightarrow \mathsf{Cu}(\mathsf{T},\mathsf{Grph})$ *whose objects are* ***temporal path-graphs****.*

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

From our persistent vs. cumulative adjunction theorem we have:

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

The change of base proposition applied to the full subcategory $\mathfrak{T}\colon \mathsf{Trees}_{\mathsf{mono}} \to \mathsf{Graph}_{\mathsf{mono}}$ yields:

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

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 $\mathsf{Graph}_{\mathsf{mono}}$-narratives, they give rise to the *same* narrative.

References

This poster is based on: *Towards a Unified Theory of Time-varying Data*. BM Bumpus, J Fairbanks, M Karvonen, W Leal, F Simard (**2024**). arXiv:2402.00206