# Processing Probabilistic Signals on Causal Abstraction Networks

**GD**, Paolo Di Lorenzo, Sergio Barbarossa

*Sapienza University, CNIT*

GSP Workshop, June 08, 2026

SPAICOM
SIGNAL PROCESSING FOR ARTIFICIAL INTELLIGENCE AND COMMUNICATIONS

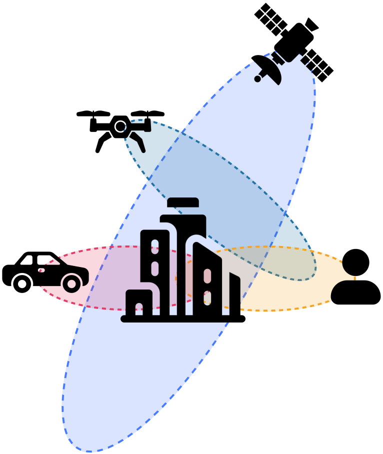
SAPIENZA
UNIVERSITÀ DI ROMA

cnit
consorzio nazionale interuniversitario per le telecomunicazioni

# The Multi-agent Problem

Modern networked systems involve **heterogeneous agents** that:

- Observe the same environment **partially** and at **different resolutions**
- Build **local models** shaped by their interactions
- Must **collaborate** despite heterogeneous perspectives



Richens & Everitt (2024). *Robust agents learn causal world models*. ICLR.

Richens, Everitt, Abel (2025). *General agents need world models*. ICML.

Bareimboim (2025). *Causal Artificial Intelligence: A Roadmap for Building Causally Intelligent Systems*. Online (draft version).

# Research Objectives

**Representation**

How to **formally organize** heterogeneous local models over a network?

**Diffusion & Alignment**

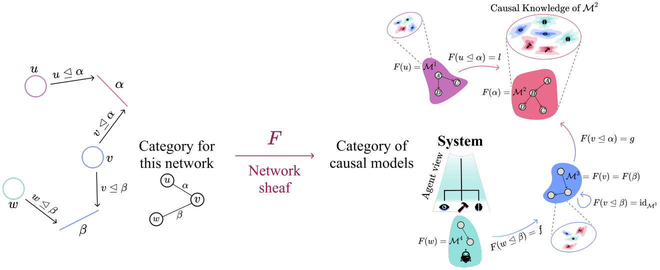
How does local knowledge **evolve, propagate, and align** across agents?

**Learning**

How to **learn** the representation model from data alone?

# Network Sheaf of Causal Knowledge





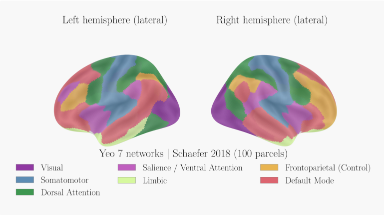

*A network sheaf organizes local causal models into a principled, globally consistent structure. It attaches causal models to nodes and edges of the network, and causal abstractions to node-edge incidence relations*

D'Acunto & Battiloro (2025). *The Relativity of Causal Knowledge*. UAI.

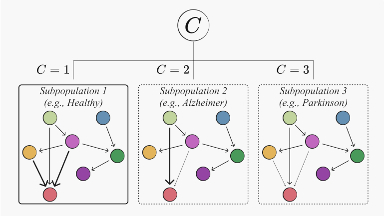
Fritz (2009). *Convex spaces I: Definition and examples*. ArXiv.

Pearl (2009). *Causality*. Cambridge university press.

# *Structural Causal Models and Mixtures*




(a) System




(b) Mixture Causal Model

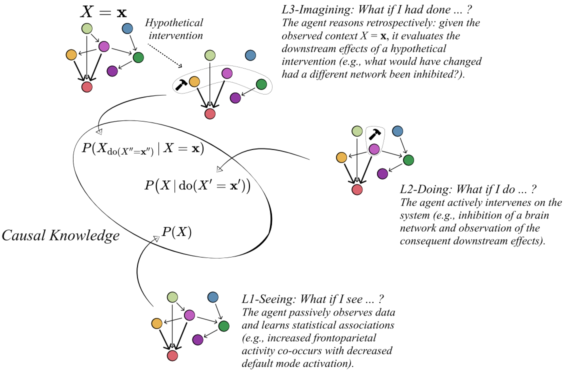
*Each agent holds a local mixture causal model: a probabilistic, context-dependent subjective belief of the system*


Pearl (2009). *Causality*. Cambridge University Press.

Geiger & Heckerman (1996). *Knowledge representation and inference in similarity networks and Bayesian multinets*. Artificial Intelligence.

Thiesson, Meek, Chickering, Heckerman (2013). *Learning mixtures of DAG models*. UAI.

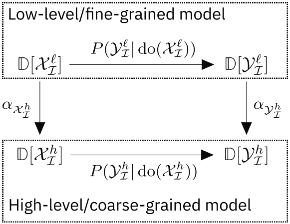
# The Ladder of Causation



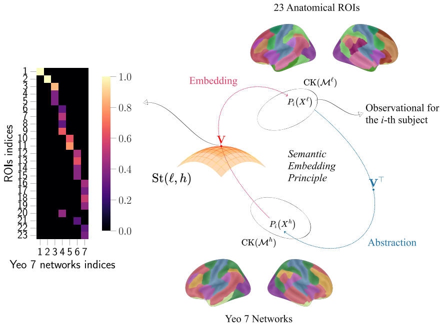

*An MCM supports reasoning at all three levels of Pearl's causal hierarchy: seeing, doing, and imagining*


Pearl & Mackenzie (2018). *The book of why: The new science of cause and effect*. Basic Books.

Recall $\mathrm{St}(d_i, d_j) \coloneqq \{\mathbf{V}_{ij} \in \mathbb{R}^{d_i \times d_j} \mid \mathbf{V}_{ij}^\top \mathbf{V}_{ij} = \mathbf{I}_{d_j}\}$




(c) $\boldsymbol{\alpha} \coloneqq \langle \mathcal{R}, m, \alpha \rangle$




(d) Constructive Linear CA

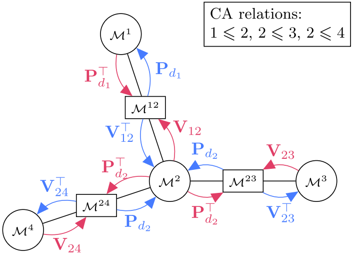
*Causal abstractions formalize interventionally consistent mappings between causal models of the same system working at different levels of granularity*

Beckers & Halpern (2019). *Abstracting causal models*. AAAI.

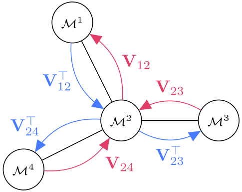
Rischel & Weichwald (2021). *Compositional abstraction error*. UAI.

D'Acunto, Zennaro, Felekis, Di Lorenzo (2025). *Causal Abstraction Learning based on the semantic embedding principle*. ICML.

# Causal Abstraction Network (CAN)




Network (co)sheaf representation



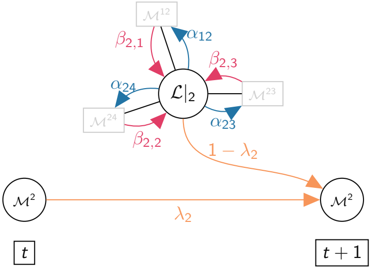

Reparameterized representation

*CAN synthetizes the network sheaf (from node to edges) and cosheaf (from edges to nodes) using constructive linear abstractions as transport maps. Reparameterization can be exploited.*

D'Acunto, Di Lorenzo, Barbarossa (2026). *Networks of causal abstractions: A sheaf-theoretic framework. ArXiv.*

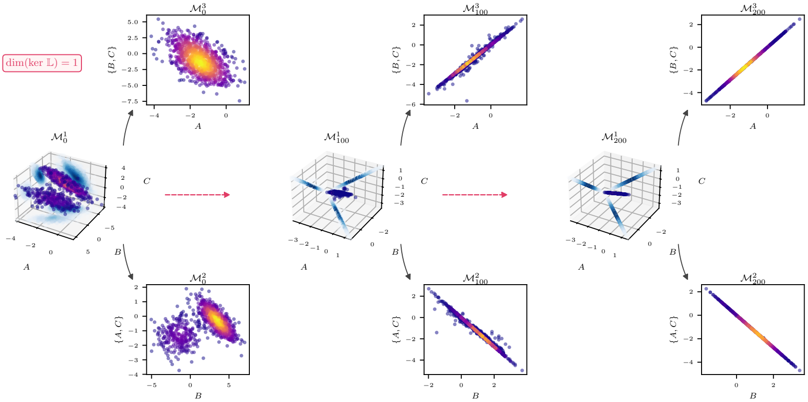
# The Laplacian operator

- Sheaf signal:
  $\mathcal{P}_0 = \{P(X^1)^0, \ldots, P(X^N)^0\}$
- Laplacian
  $\mathcal{L} : \mathcal{C}^0(\mathbb{G}; E) \to \mathcal{C}^0(\mathcal{G}; E)$,
  $\mathcal{L} = \delta^\star \circ \delta$
  Algebraic representation $\mathbb{L}$
- Coboundary
  $\delta : \mathcal{C}^0(\mathcal{G}; E) \to \mathcal{C}^1(\mathbb{G}; E)$;
  $\delta(\mathcal{P})\big|_{e_{ij}} = \mathrm{cc}_{\alpha_e}\left(P(X^i),\ P_{\mathbf{V}_{ij}}(X^j)\right)$
- Boundary
  $\delta^\star : \mathcal{C}^1(\mathcal{G}; E) \to \mathcal{C}^0(\mathbb{G}; E)$




$$\mathcal{P}_{t+1} = \mathrm{cc}_{\boldsymbol{\lambda}}(\mathcal{P}_t,\ \mathcal{L}(\mathcal{P}_t))$$

# Diffusion and Convergence




*The nonlinear sheaf Laplacian drives local distributions toward a common one-dimensional latent subspace* ($\dim(\ker \mathbb{L} = 1)$)*: the space of global sections*

- We work in non-Euclidean setting
- $P\left(X^i\right)$ ($d_i$-dimensional), $P\left(X^j\right)$ ($d_j$-dimensional), $d_i > d_j$
- Is $P\left(X^j\right)$ a constructive linear abstraction of $P\left(X^i\right)$?
- $D_{\mathbf{V}_{ij}^\top}(P\left(X^j\right), P\left(X^i\right))$
  - Gaussian setting: $\mathrm{KL}(P\left(X^j\right) || P_{\mathbf{V}_{ij}^\top}(X^i))$
  - Mixtures of Gaussians: $\mathrm{MPW}_2(P\left(X^i\right), P\left(X^j\right))$

$$\mathrm{MPW}_2(P\left(X^i\right), P\left(X^j\right)) := \inf_{\substack{\mathbf{\Omega}\in\Pi(\mathbf{w}_i,\mathbf{w}_j)\\ \mathbf{V}_{ij}\in\mathrm{St}(d_i,d_j)}} \sum_{\substack{p\in[I]\\ q\in[J]}} \omega_{pq}\,\mathrm{W}_2^2\Big(P_{\mathbf{V}_{ij}^\top}(X^{(i,p)}),\ P(X^{(j,q)})\Big)$$

D'Acunto, Zennaro, Felekis, Di Lorenzo (2025). *Causal Abstraction Learning based on the semantic embedding principle*. ICML.

Cai & Lim (2022). *Distances between probability distributions of different dimensions*. IEEE Transactions on Information Theory.

Salmona, Desolneux, Delon (2024). *Gromov-Wasserstein-like Distances in the Gaussian Mixture Models Space*. TMLR.

# Smoothness vs. Global Alignment

- Smoothness:

$$f(\mathcal{E}, \{\mathbf{V}_{ij}^\top\}) = \sum_{e_{ij} \in \mathcal{E}} D_{\mathbf{V}_{ij}^\top}(P\left(X^j\right), P\left(X^i\right))$$

- Global section condition:

$$P\left(X^i\right) = P_{\mathbf{V}_{ij}}(X^j), \quad \mathbf{V}_{ij} \in \mathrm{St}(d_i, d_j), \quad \forall e_{ij} \in \mathcal{E}$$

- *Global alignment* $\implies$ *Smoothness* (the converse does not hold)

# Learning CANs from Data

- Global nonconvex learning problem:

$$\max_{\mathcal{E}\subseteq\mathcal{F}} |\mathcal{E}|$$

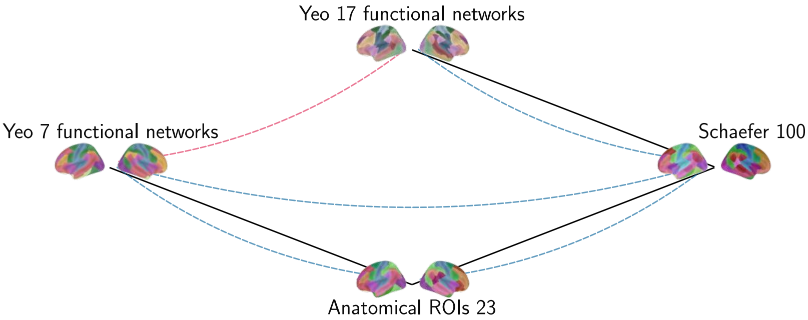
subject to $$\inf_{\substack{\mathbf{\Omega}\in\Pi(\mathbf{w}_i,\mathbf{w}_j) \\ \mathbf{V}_{ij}\in\mathrm{St}(d_i,d_j)}} \mathrm{MPW}_2(P\left(X^i\right), P\left(X^j\right)) \leq \tau_{\mathrm{MPW}}\,, \quad \forall\, e_{ij}\in\mathcal{E}\,.$$

- Search procedure over the candidate edges that exploits CLCA compositionality
- Local problems solved in alternating fashion over $\mathbf{\Omega}$ (entropic OT) and $\mathbf{V}_{ij}$ (first-order Riemannian optimization)


Peyré & Cuturi (2019). *Computational optimal transport: With applications to data science*. Now Foundations and Trends.

Lai & Osher (2014). *A splitting method for orthogonality constrained problems*. Journal of Scientific Computing.

# Networks of Brain Models based on Different Atlases



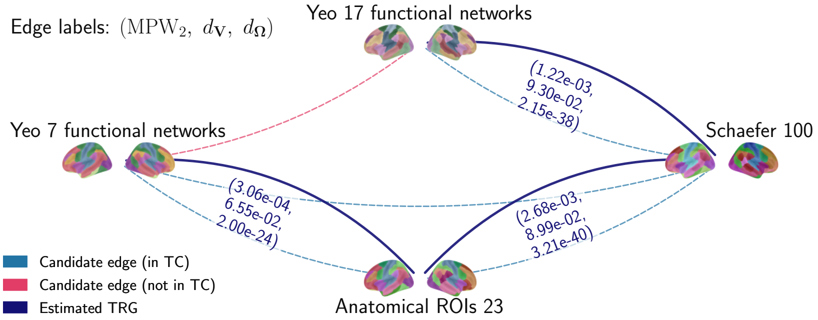

- Preprocessed multi-task HCP data, 100 healthy subjects
- Gaussian mixture models fitted at the granularity Schaefer 100 via EM and cross-validation
- Candidate edges (dashed) according to spectral interlacing necessary condition for the existence of causal abstractions

---

D'Acunto, Di Lorenzo, Barbarossa. (2026) *Multi-agent SCMs: From local to global knowledge via a sheaf-theoretic framework*. Submitted.

Santoro, Battiston, Petri, Amico (2023). *Higher-order organization of multivariate time series*. Nature Physics.

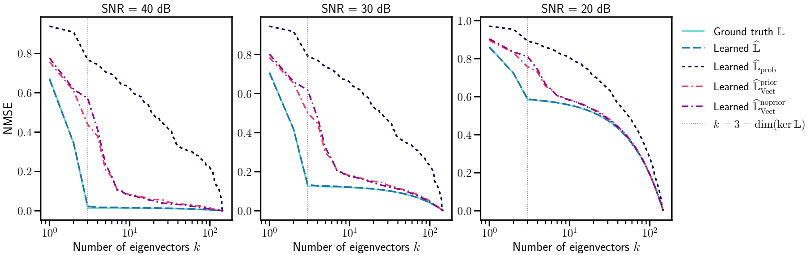
# Networks of Brain Models based on Different Atlases




- Our learning algorithm recovers the ground-truth V-shape topology and the CLCA maps from observational fMRI data alone

# *Signal Reconstruction*

- Perturbed global section $\mathcal{P} \in \mathbb{R}^{147 \times 162100}$ at varying degree of noise
- Baselines: learned sheaf Laplacian in $\mathsf{Vect}_{\mathbb{R}}$ with/out topological prior, learned sheaf Laplacian in CSprob (dense Stiefel maps)



# Conclusions

## Representation

Causal Abstraction Network

- Nodes and edges: Convex spaces of probability distributions
- Transport maps: constructive linear causal abstractions and their transpose

## Diffusion & Alignment

Nonlinear sheaf Laplacian $\mathcal{L}$

- Featured by hyper-parameters
- Spectral characterization via $\mathbb{L}$
- Convergent discrete-time dynamics under mild conditions

## Learning

Framework based on smoothness functional

- Separable over the edges
- Exploiting compositionality of causal abstraction
- Amenable to different metrics, according to non-Euclidean setting