

# Reconstructing and interacting with multi-scale dynamics of knowledge

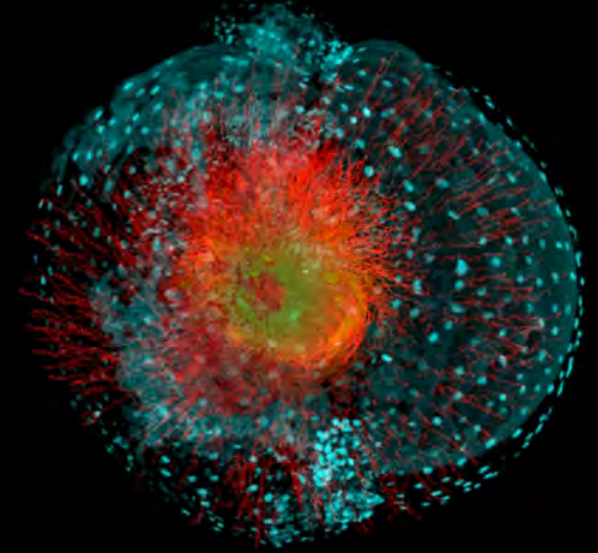
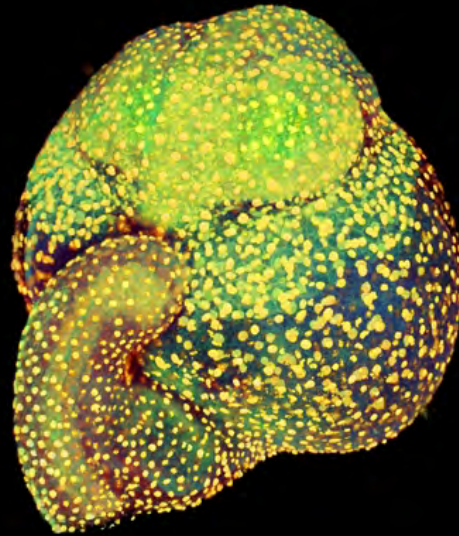
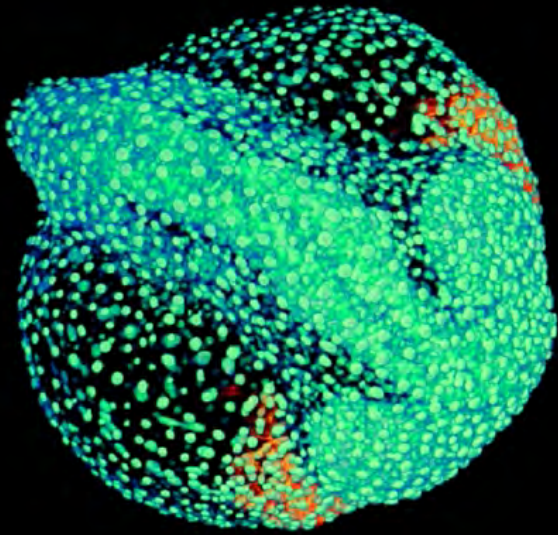
Quentin Lobbé

Researcher at the Complex Systems Institute of Paris IdF (ISCPIF)  
with David Chavalarias & Alexandre Delanoë



# Science as a living system

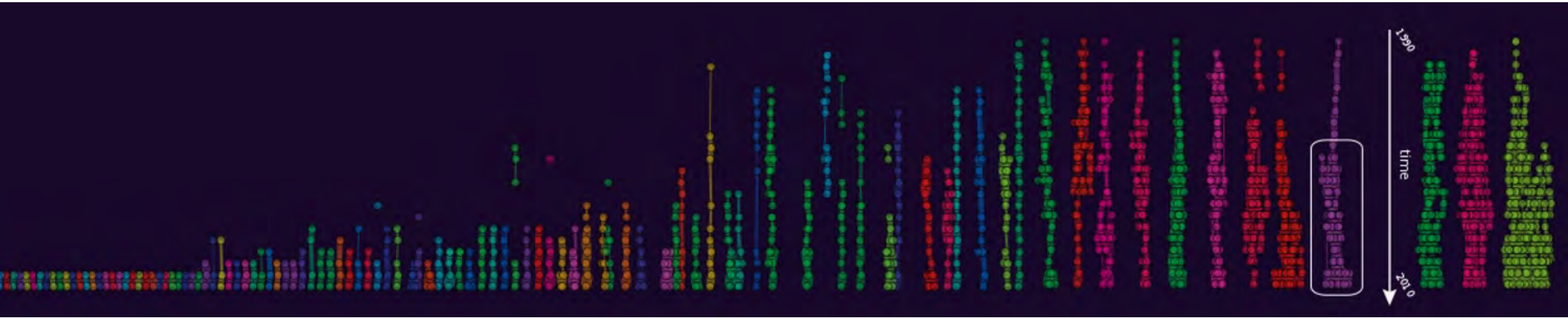
Our initial inspiration comes from embryogenesis reconstruction – *Bioemergences* project at ISCPIF (2005 - 2015)



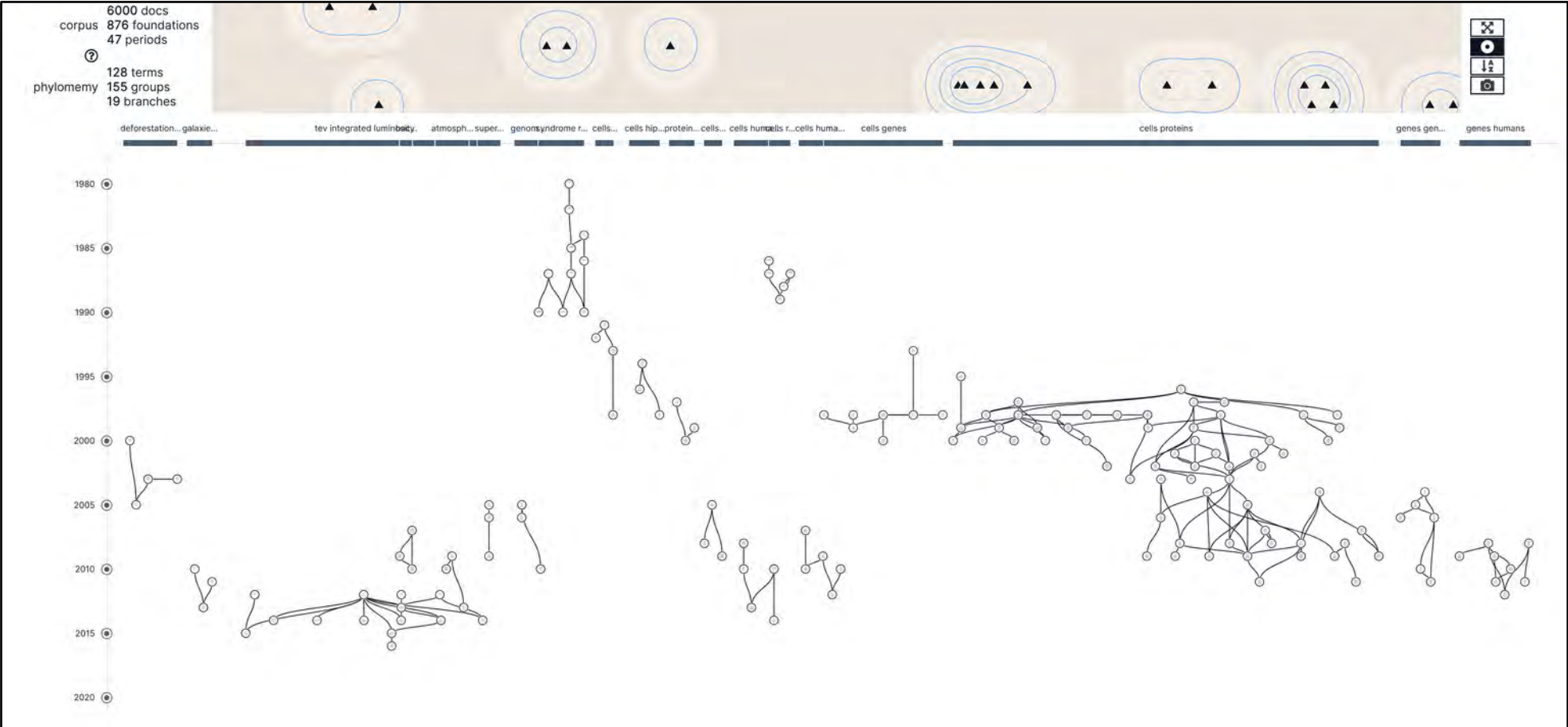
# Science as a living system

Science Evolution - Chavalarias & Cointet, PLoS One (2013)

Phylomemories are inheritance networks of elements of knowledge



# Science as a living system



## Phylomemy reconstruction (2020 ++)

How do we define a branch of knowledge ?

What is the meaning of an inter-temporal lineage ?

Which links should we consider ?

How do we translate local knowledge dynamics ?

Can we navigate through different levels of observation ?

Can we visualize multi-scale structures ?

# Phylometry reconstruction (2020 ++)

Phenomenological reconstruction methods can be summarized as three steps processes.

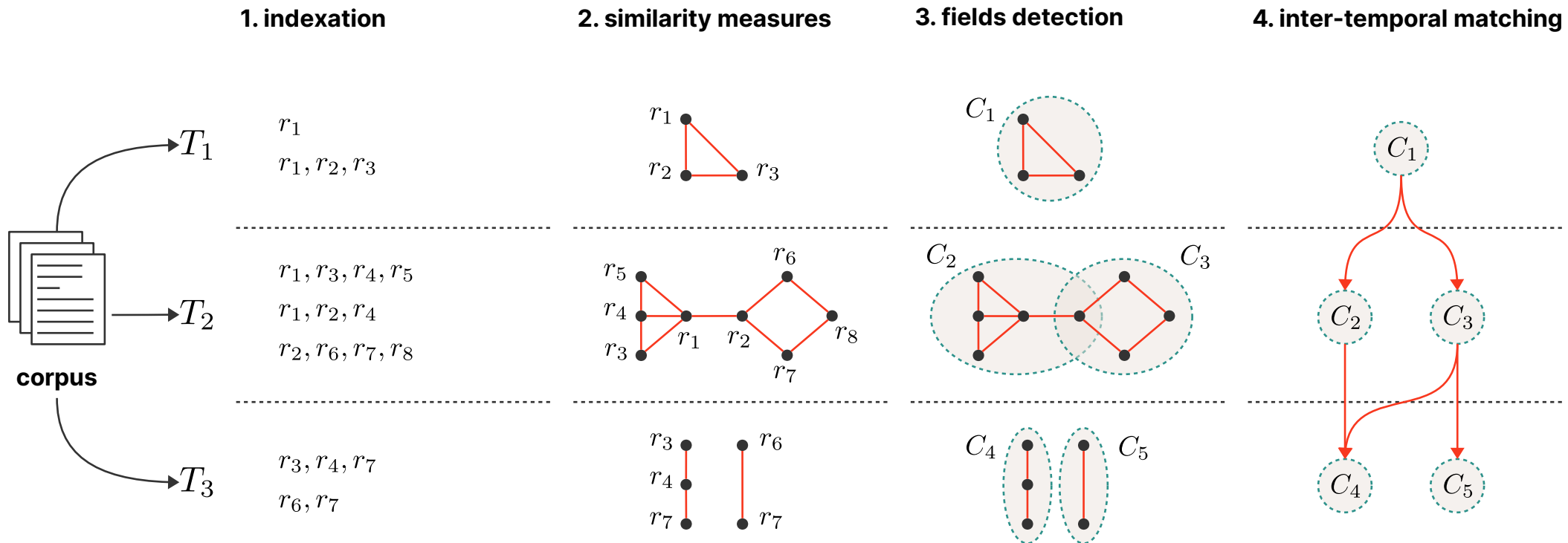


An object or a natural  
phenomenon

The reconstruction as a  
formal object

A projection in a  
visualization system

# Phylomemy reconstruction (2020 ++)



$\Phi = {}^4\Phi \circ {}^3\Phi \circ {}^2\Phi \circ {}^1\Phi$  is the reconstruction process that goes from **terms** to **groups**, **fields** and **branches** of knowledge clusters

# Upstream & downstream Inter-temporal matching ${}^4\Phi$

a weighted Jaccard index



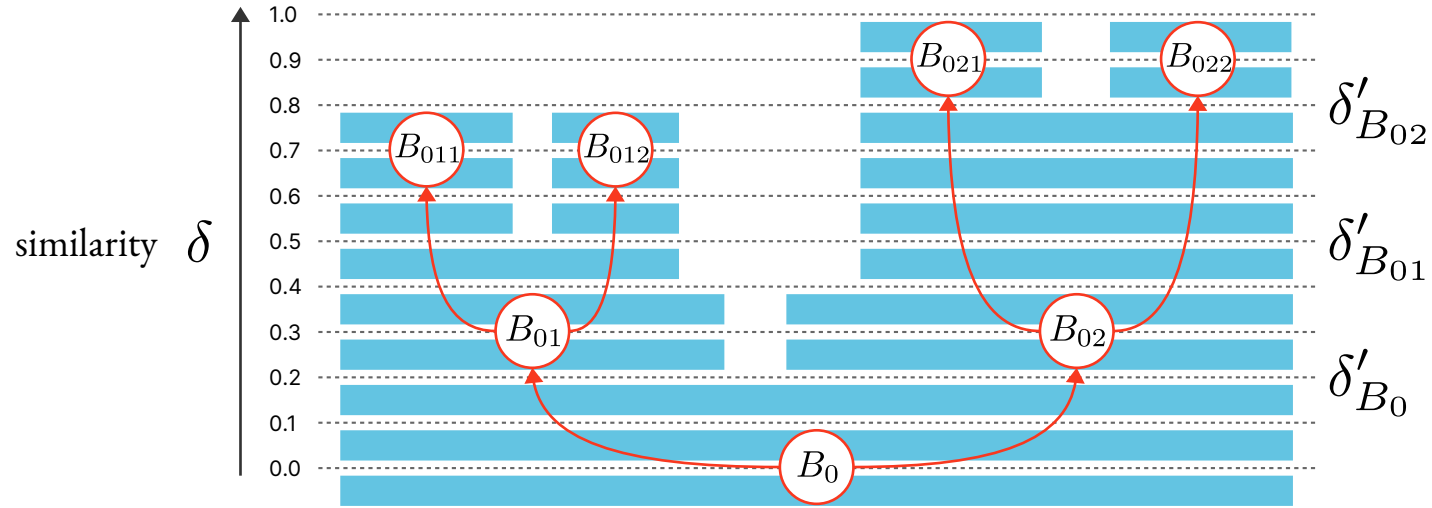
We consider the function  ${}^4\Phi^\gamma : \mathcal{C} \times [0, 1] \mapsto (\mathcal{P}(\mathcal{C}), w)$  so that  ${}^4\Phi^\gamma(C^T) = (\{C_j \in \kappa_{C^T}^\gamma\}, w)$

where

- $\kappa_{C^T}^\gamma = \operatorname{argmax}_{\Delta(C^T, \kappa)} [\operatorname{arg min}_{\{\kappa \subset \mathcal{C}^{T'} \succ T \mid \Delta(C^T, \kappa) \geq \delta\}} \{\tau(C^T, \kappa)\}]$
- $\mathcal{C}^{T'} \succ T = \{C^{T'} \in \mathcal{C} \mid T' \succ T\}$  is the set of all clusters of  $\mathcal{C}$  which period is strictly a posterior to  $T$
- $w \in [0, 1]$  is the association strength of  $\kappa_{C^T}^\gamma$



# Sea rise algorithm



by gradually elevating the similarity's threshold between the fields, we split the branches and we recursively reconstruct their local homogeneity

## Quality and level of observation $\lambda$

The quality of an answer  $\mathcal{Q}(x)$  can be assessed with respect to the desired level of observation  $\lambda \in [0, 1]$  between precision and recall with the following  $F$ -score function

$$F_\lambda(x, k) = \frac{(1 + f(\lambda)^2) \cdot (\xi_x^k \cdot \rho_x^k)}{\rho_x^k + f(\lambda)^2 \cdot \xi_x^k}$$

where  $f(\lambda) = \tan\left(\frac{\pi \cdot \lambda}{2}\right)$

$$\text{the precision } \xi_x^k = \frac{|\mathcal{C}_x \cap \mathcal{C}_{B_k}|_{\mathcal{T}_x}}{|\mathcal{C}_{B_k}|}$$

$$\text{the recall } \rho_x^k = \frac{|\mathcal{C}_x \cap \mathcal{C}_{B_k}|}{|\mathcal{C}_x|}$$

## Quality and level of observation $\lambda$

The quality of an answer  $\mathcal{Q}(x)$  can be assessed with respect to the desired level of observation  $\lambda \in [0, 1]$  between precision and recall with the following  $F$ -score function

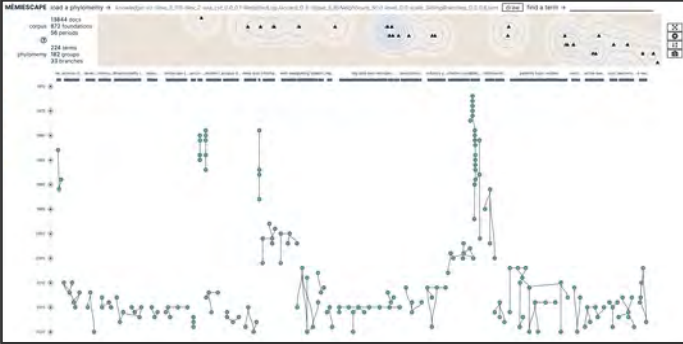
$$F_{\lambda}(x, k) = \frac{(1 + f(\lambda)^2) \cdot (\xi_x^k \cdot \rho_x^k)}{\rho_x^k + f(\lambda)^2 \cdot \xi_x^k}$$

The objective function for the evaluation of the relevance of  $\phi$  for answering  $\mathcal{Q}(x)$  thus becomes :

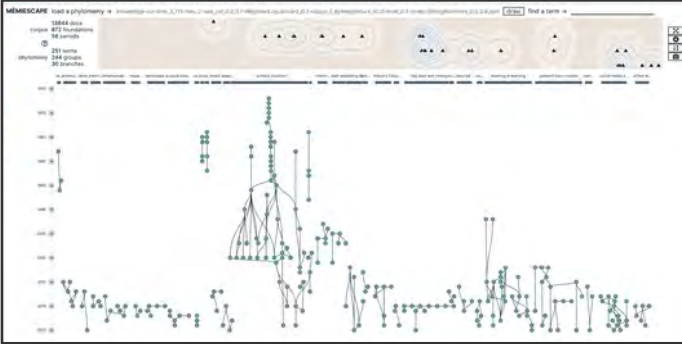
$$F_{\lambda}^x(\varphi) = \sum_{B_k \in \varphi | B_k \cap \mathcal{C}_x \neq \emptyset} \Psi_x(k) \cdot F_{\lambda}(x, k)$$

# Quality and level of observation $\lambda$

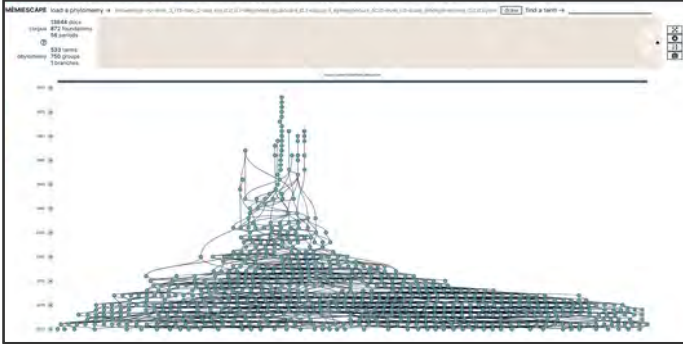
$\lambda = 0$



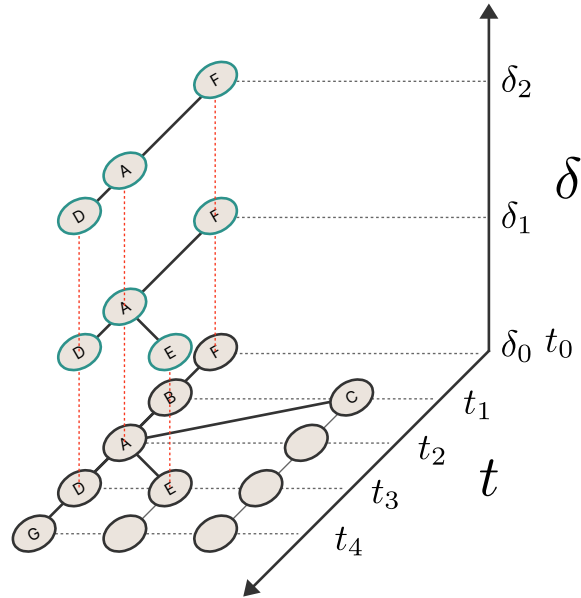
$\lambda = 0,5$



$\lambda = 1$



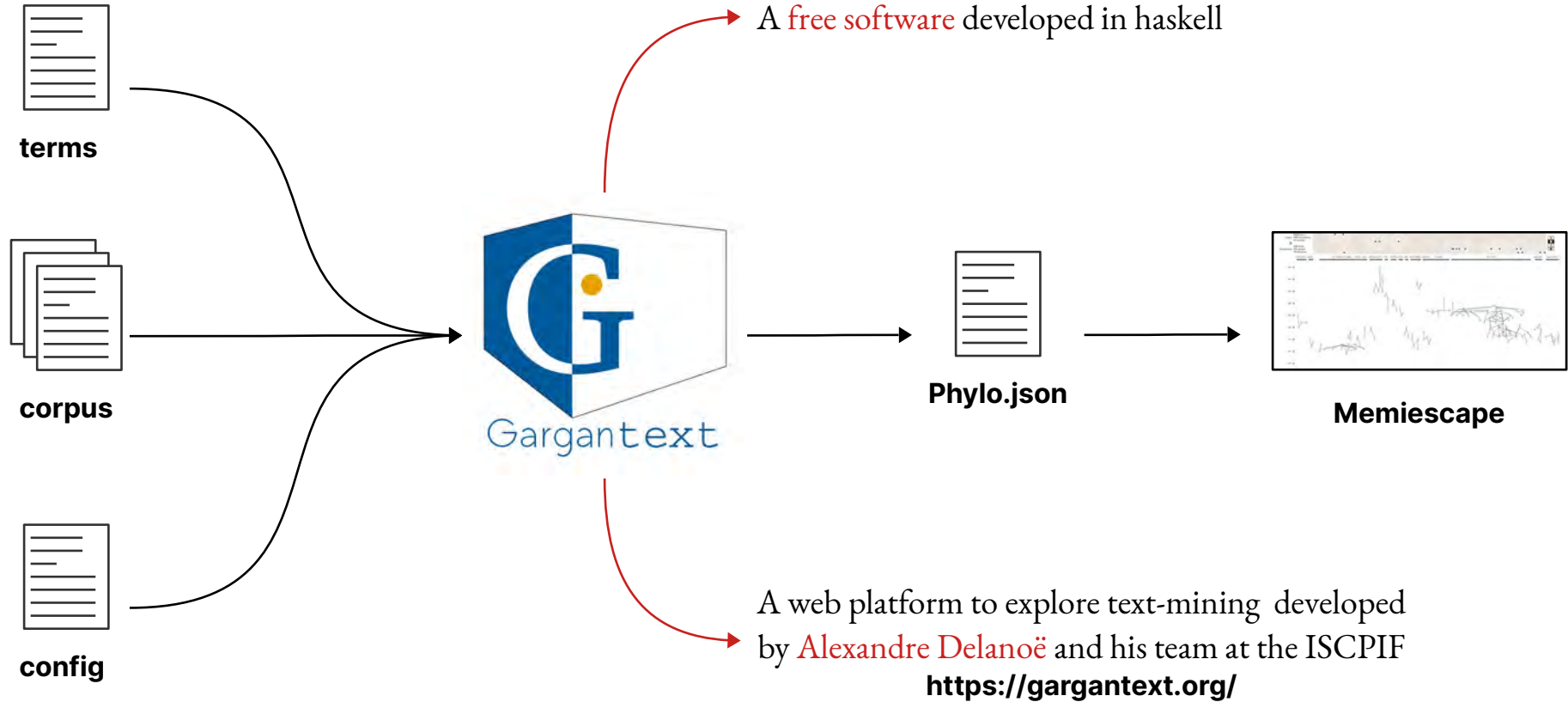
# Phylomemies as foliation on temporal series of clustering



A phylomemy  $\phi$  is a **foliation** on a **temporal series of clustering**  $\mathcal{C}^*$ . It describes, for any cluster  $C_j^T$  in temporal components of  $\mathcal{C}^*$  and any threshold  $\delta$ , the relevant inheritance linkages of  $C_j^T$ .

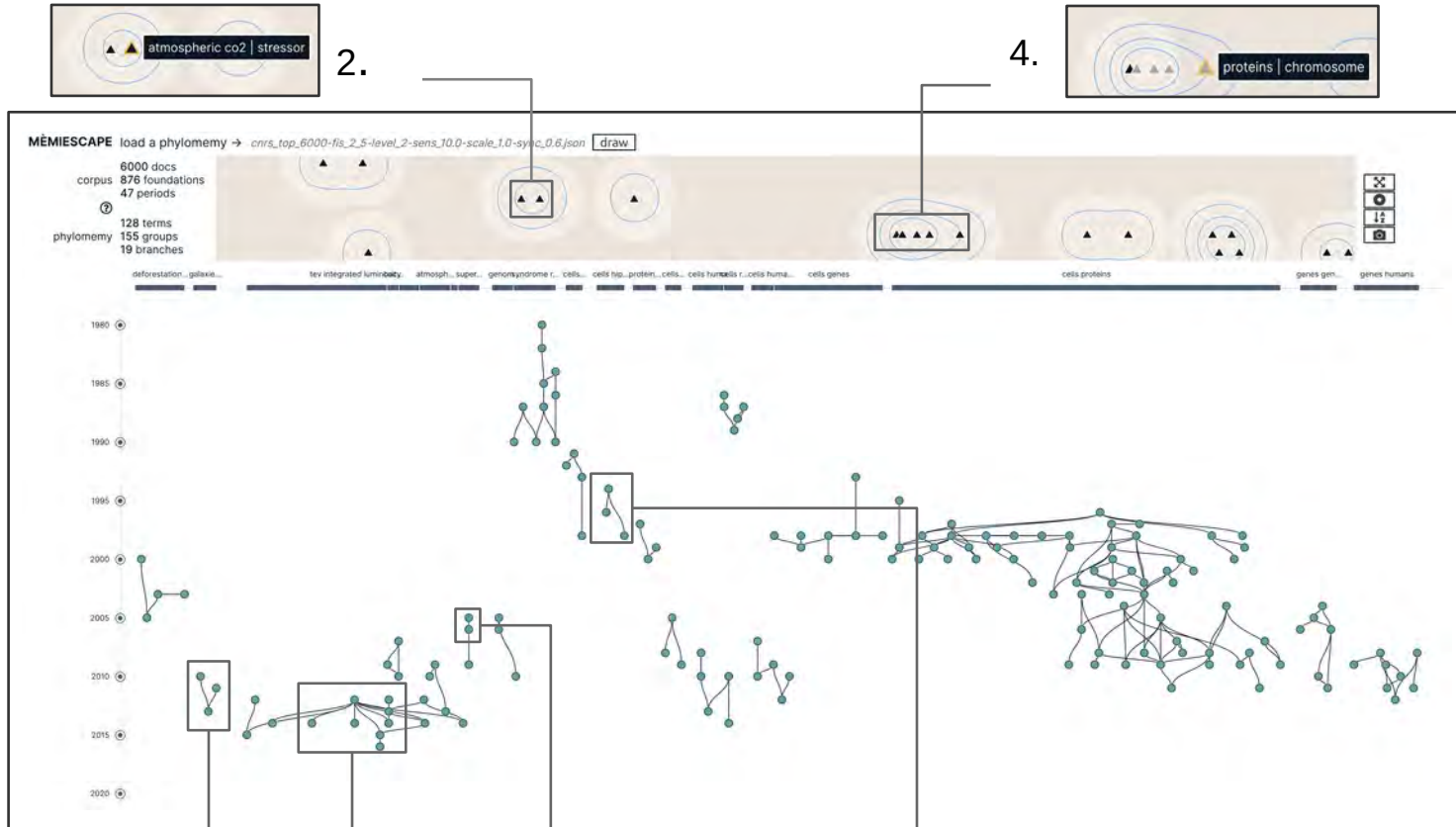
Consequently, for the study of knowledge dynamics,  $\mathcal{R}$  is the space of all foliations on temporal series of clustering.

# Implementation within Gargantext

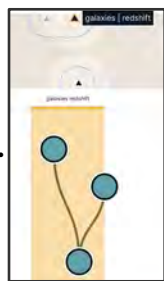


# The Memiescape

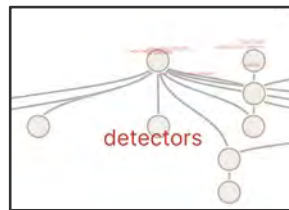
1.



3.



5.



6.



7.



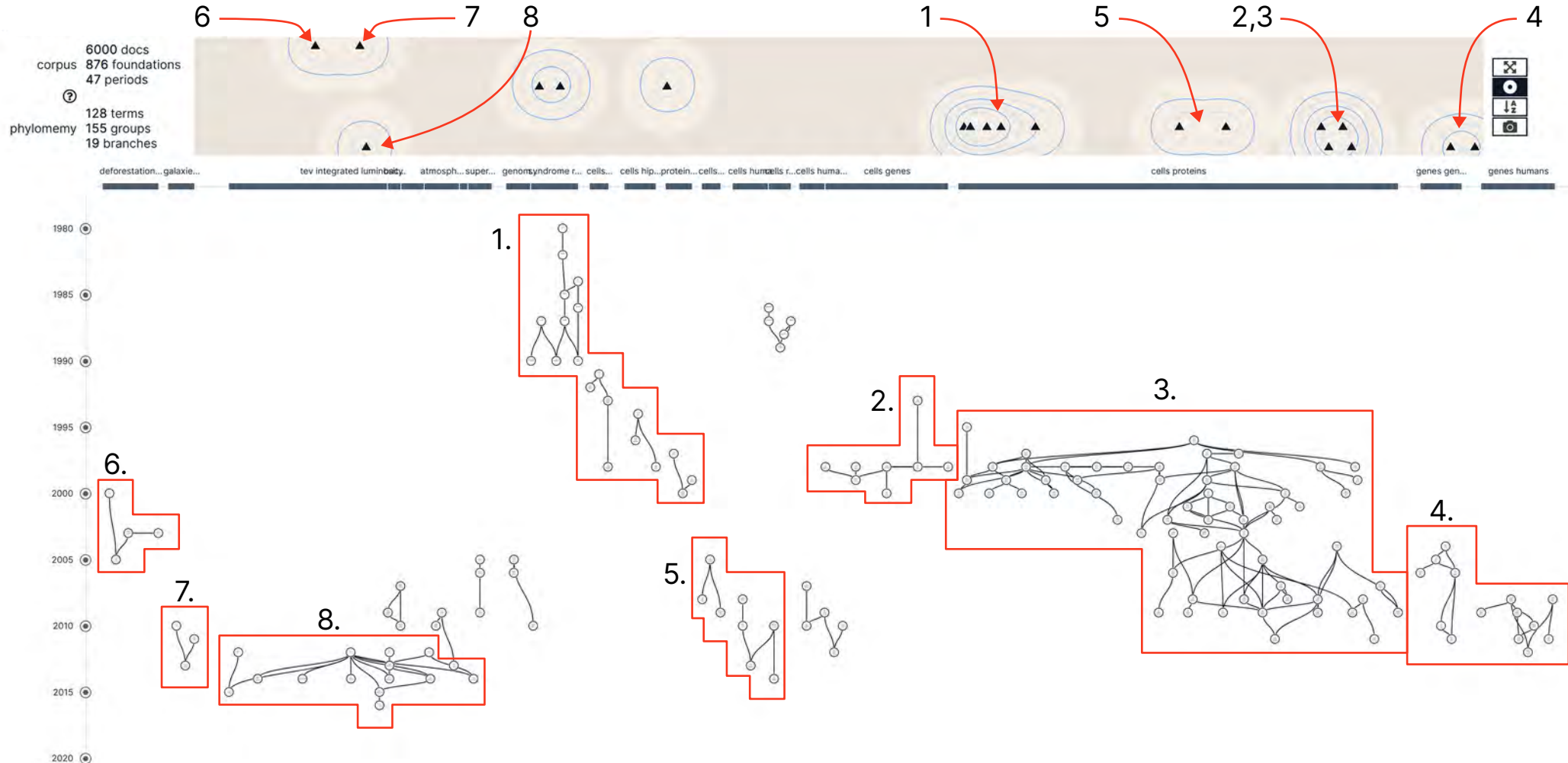
4.



2.



# Phylomemy of the CNRS 6000 most cited papers (WoS)





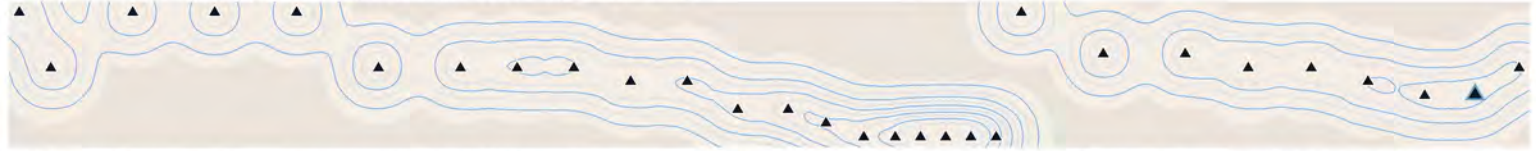
# Phylomemy of the state of the art of knowledge visualization (WoS)



# Phylomemy of the covid19 vaccines trials

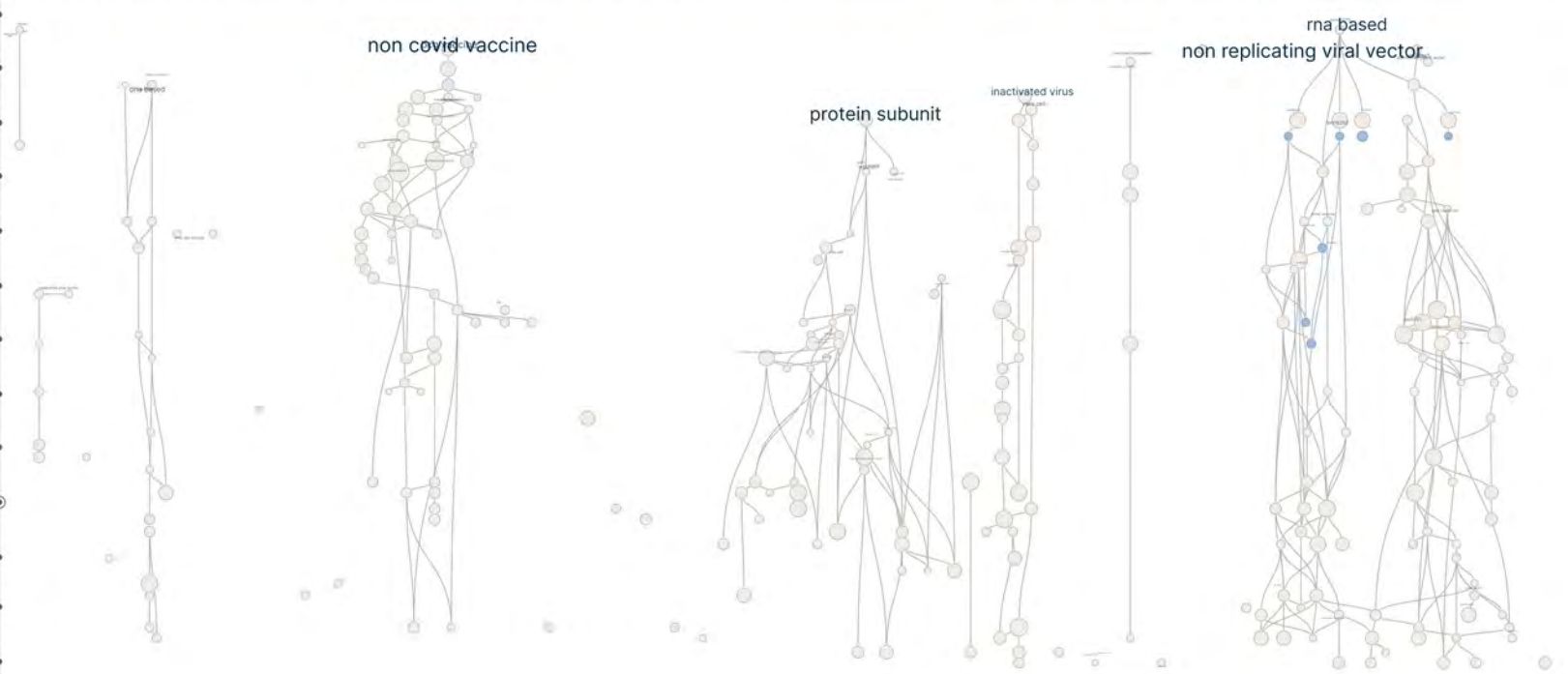
MÈMIESCAPE load a phylomemy →    find a term → \_\_\_\_\_

648 docs  
corpus 129 foundations  
52 periods  
129 terms  
phylomemy 259 groups  
29 branches



ce... replica... live... cel... electrop... vir... virus-ll... mycobac... razi co... tak... polio vaccine bcg vaccine. alumini... m1-ad.zbd a.virus- gbp5... nbp... recombinant protein vaccine adjuvant vbi... inactivated v... pdv... sam-ln.vaccin.sam-l... non rep... 13 c... rna based non replicating viral vector vaccl...

Apr  
May  
Jun  
Jul  
Aug  
Sep  
Oct  
Nov  
Dec  
2021  
Feb  
Mar  
Apr



rna based  
bnt162b1  
bnt162b3  
bnt162a1  
bnt162b2  
bnt162c2

# Publications



Chavalarias, D., Lobbé, Q., & Delanoë, A. (2021). Draw me Science - multi-level and multi-scale reconstruction of knowledge dynamics with phylomemies.  
<https://hal.archives-ouvertes.fr/hal-03180347>

Lobbé, Q., Delanoë, A. & Chavalarias, D. (2021). Exploring, browsing and interacting with multi-scale structures of knowledge.  
<https://hal.archives-ouvertes.fr/hal-03181233>

Thank you !

David Chavalarias

Alexandre Delanoë



Quentin Lobbé

Bruno Gaume