# Complex systems

a dynamical point of view

*Mauro Faccin*

March 7, 2023

# AutoInformation state aggregation

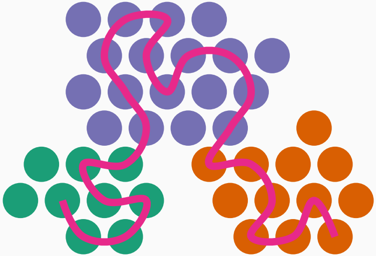
# Projected Markov Chain

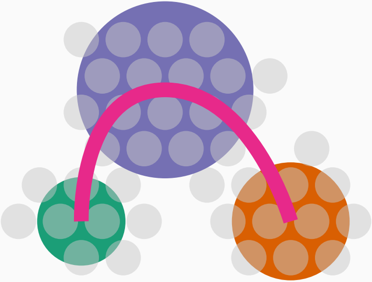
## Markov Chain

$$\ldots, \mathbf{x}_{\text{past}}, \mathbf{x}_{\text{now}}, \mathbf{x}_{\text{future}}, \ldots$$

## Projection

$$\ldots, \mathbf{y}_{\text{past}}, \mathbf{y}_{\text{now}}, \mathbf{y}_{\text{future}}, \ldots$$





# Non-linear correlations

**AutoInformation**

$$\mathbf{I}(\mathbf{y_t};\mathbf{y_{t-\tau}})$$

Non-linear *correlation* between successive time-steps

$$\mathbf{I}(\mathbf{y_t};\mathbf{y_{t-\tau}}) = \underbrace{\mathbf{I}(\mathbf{y_t};\mathbf{y_{t-\tau}},\ldots)}_{^{1}\text{Predictability}} - \underbrace{\mathbf{I}(\mathbf{y_t};\mathbf{y_{t-2\tau}},\ldots|\mathbf{y_{t-\tau}})}_{^{2}\text{non-Markovianity}}$$

where $\tau$ represents a time-scale parameter.

**We ask to:**

1. maximize predictability of the dynamics;
2. minimize non-Markovianity (effective memories from the projection).

M.F. et al, Journal of Complex Networks, 2018

M.F. et al, PRL, 2021

## Links to the literature

### Modularity

$\chi_c$ characteristic function of class $c$

$$Q = \sum_c \mathrm{Cov}\left(\chi_c(t), \chi_c(t+1)\right)$$

Auto-covariance of the dynamics on the partition space.

Linear correlation between consecutive time-steps.

Shen et al. (2010) PRE, 82, 016114

### DC-SBM

$$\mathcal{S} \propto \frac{1}{2} \sum_{cd} e_{cd} \log \frac{e_{cd}}{e_c e_d}$$

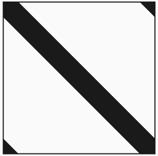
Fitting a generative model (e.g. DC-SBM) to the data through log-likelihood maximization can be seen as maximizing the AutoInformation for paths of lenght $\tau = 1$ (e.g. links).

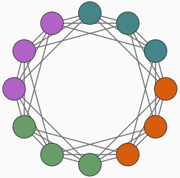
Karrer and Newman (2011), PRE 83, 016107.

# Example: A cycle

A regular ring lattice with **N** nodes, each connected with **k** neighbours.

How many classes?

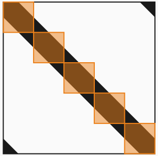
Adj: 

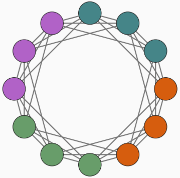


# Example: A cycle

A regular ring lattice with **N** nodes, each connected with **k** neighbours.

How many classes?

Adj: 

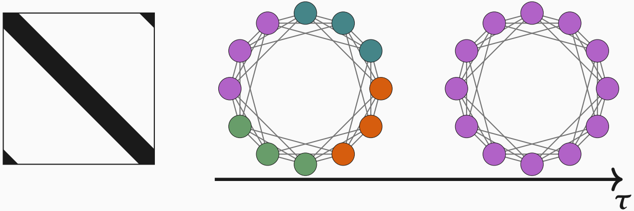


# Example: A cycle

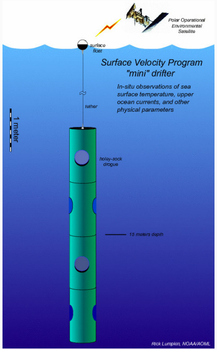
A regular ring lattice with **N** nodes, each connected with **k** neighbours.

How many classes?

Adj:

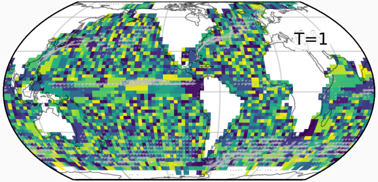


# Example: Ocean buoys

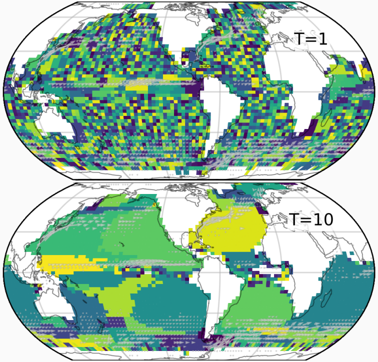

Global Drifter Program

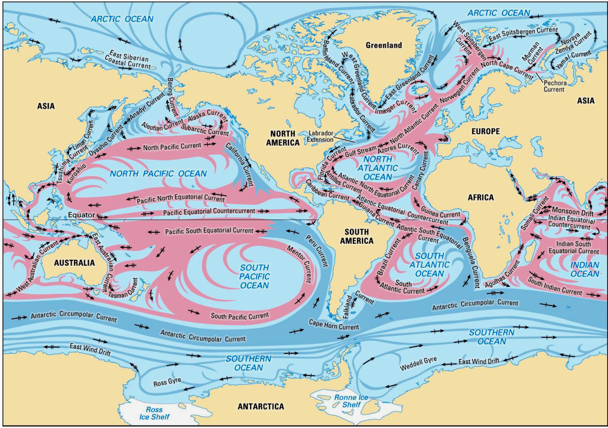
# Example: Ocean buoys




Each time step lasts 16 days.

# Example: Ocean buoys

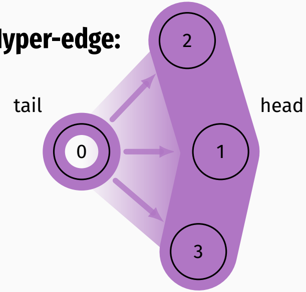



Each time step lasts 16 days.

# Dynamics on hypergraphs

**Hyper-edge:**




| node | role |
|---|---|
| 0 | tail |
| 1, 2, 3 | head |

**Hypergraph:** $\{\mathbf{N}, \mathbf{E}\}$: nodes and hyperedges

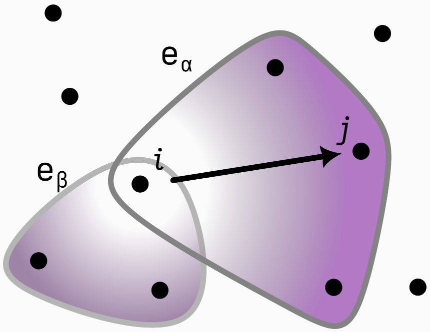
**Nodes:** same as before

**Hyperedges:** $\mathbf{e}_\alpha = \{\text{tail}, \text{head}\} \in \mathbf{E}$

# Random walker on a hypergraph

The walker

- sits on a node $\mathbf{i}$
- chooses a hyperedge $\mathbf{e}_\alpha$ incident on $\mathbf{i}$ in its tail (user $\mathbf{i}$ tweeted $\boldsymbol{\alpha}$) with probability dependant on the hyperedge size (with parameter $\boldsymbol{\tau}$);
- chooses an exit node $\mathbf{j}$ from the head of $\mathbf{e}_\alpha$ ($\boldsymbol{\alpha}$ get retweeted by user $\mathbf{j}$) with flat probability.

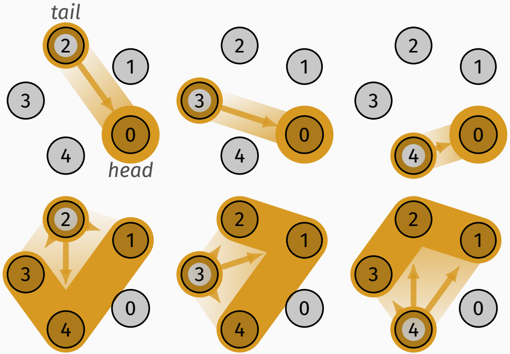


## Yes but why?

**Measure the dynamics through its transition matrix instead of extending to the hypergraph framework with the corresponding complexity overhead.**

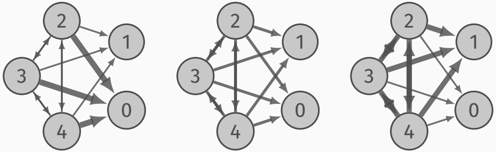
## ↓ Ranking nodes

Hyper-Edges

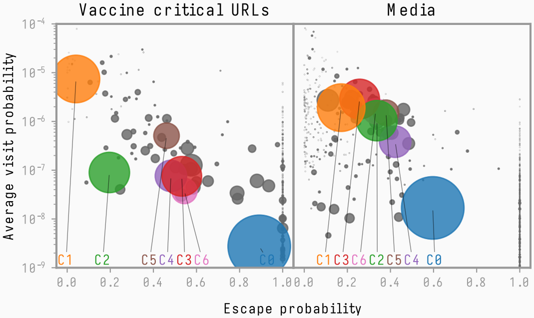



| $\tau = -1$ | | $\tau = 0$ | | $\tau = 1$ | |
|---|---|---|---|---|---|
| ranking | prob | ranking | prob | ranking | prob |
| 0 | 0.325 | 1 | 0.224 | 1 | 0.251 |
| 1 | 0.186 | 0 | 0.224 | 2 | 0.2 |
| 2 | 0.163 | 2 | 0.184 | 3 | 0.2 |
| 3 | 0.163 | 3 | 0.184 | 4 | 0.2 |
| 4 | 0.163 | 4 | 0.184 | 0 | 0.149 |

Effective graphs



# Communities and roles



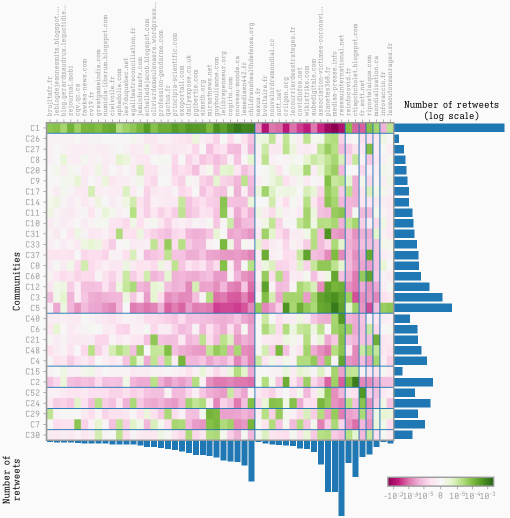

**Visiting probability** probability of being visited by a random walker (a retweet)

**Escape probability** probability of reaching other communities (being retweeted outside one's bubble)

| Comm. | Interpretation |
|---|---|
| $C_0$ | media aggregators or web influencers. |
| $C_1$ | Far right groups. |
| $C_2$ | health institutions and MDs |
| $C_3$ | French news media. |
| $C_4$ | international news media. |
| $C_5$ | Far left and trade unions. |
| $C_6$ | government representatives. |
| $C_7$ | Canada |

$C_1$ and $C_5$ are the main actors in spreading vaccine-critical content (high visiting and escape probability).

# Community URL usage pattern.




Clustering of communities by URL usage pattern.

**right wing $C_1$** use an original body of URLs.

**left wing and news media** use a similar set or sources.

**health institutions** use an original set of sources.

## Compartmental models

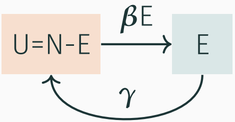
## SIS Compartmental model

**Engagement**

Engaged users share a URL from the set within a time window $\tau = 3$ days.

$$\mathbf{dE_t} = \alpha_t \frac{\mathbf{E_t}(\mathbf{N_t} - \mathbf{E_t})}{\mathbf{N_t}} - \beta_t \mathbf{E_t}$$

$$\mathbf{R_t} = \frac{\alpha_t}{\beta_t}$$

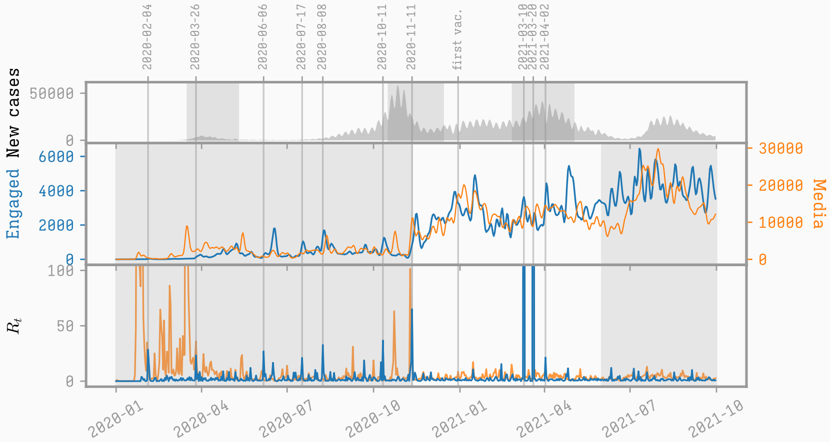



$\alpha_t$ engagement rate

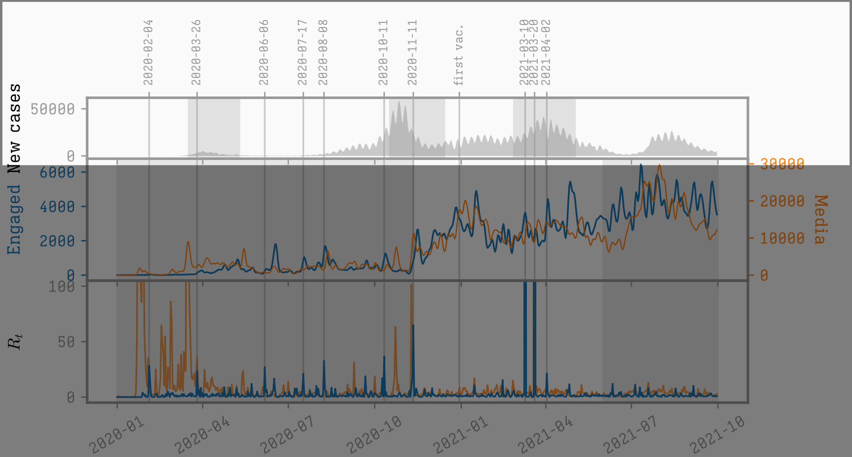
$\beta_t$ unengagement rate

$\mathbf{R_t}$ reproduction number

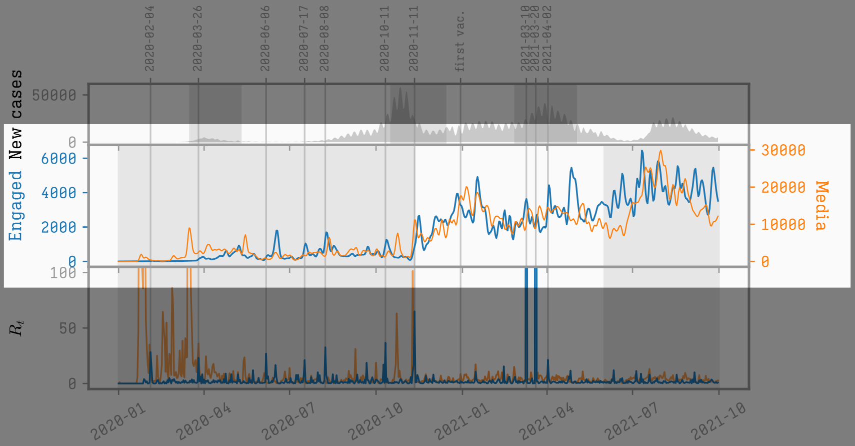
# Engagement evolution



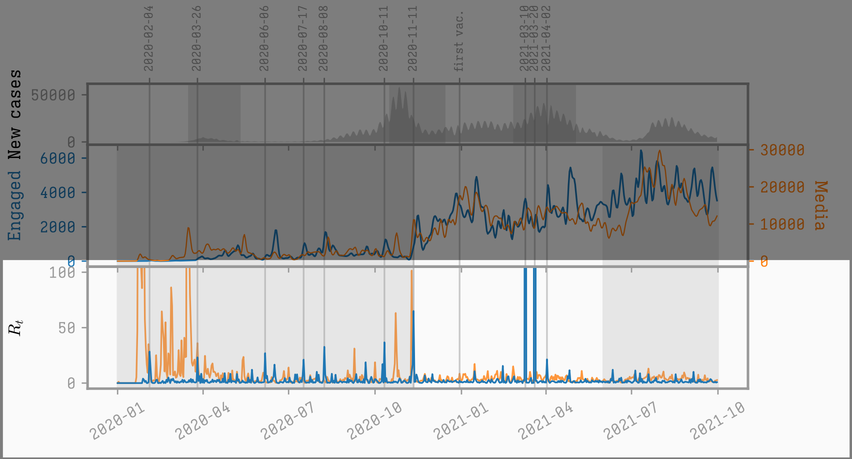
# Engagement evolution



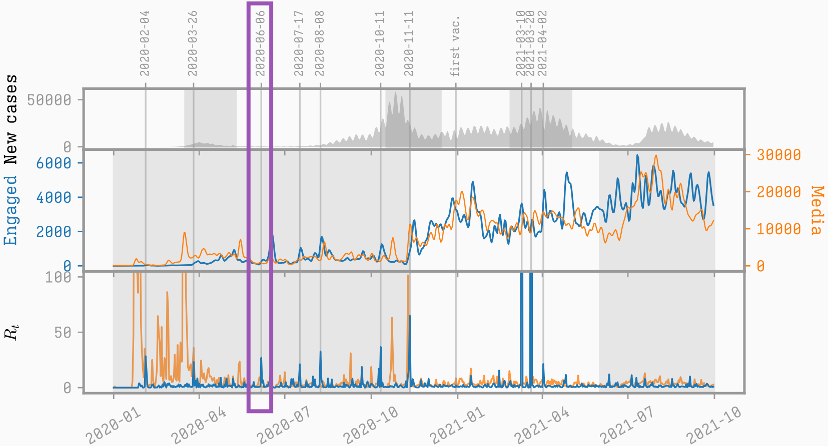
# Engagement evolution



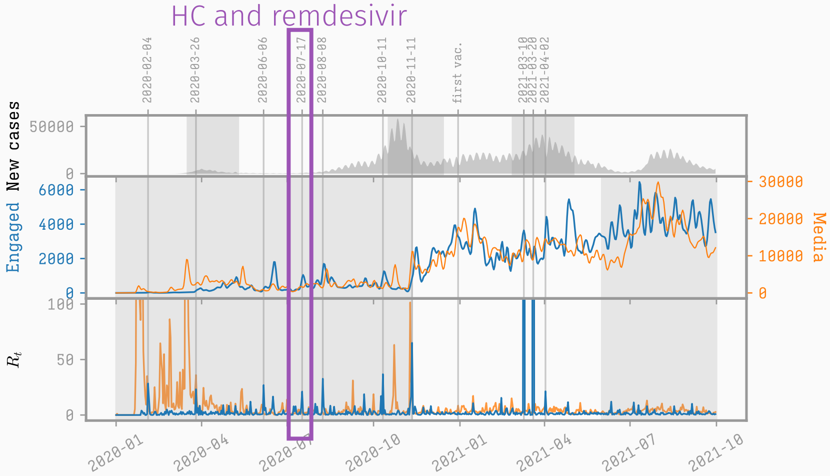
# Engagement evolution



deaths in clinical trials

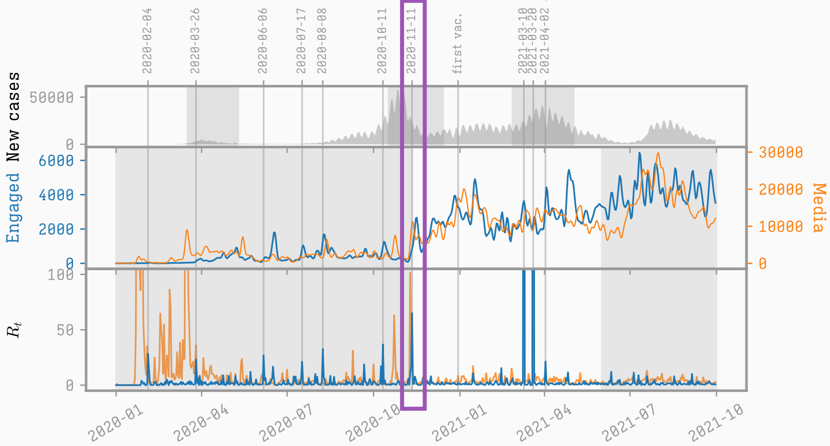


# Engagement evolution

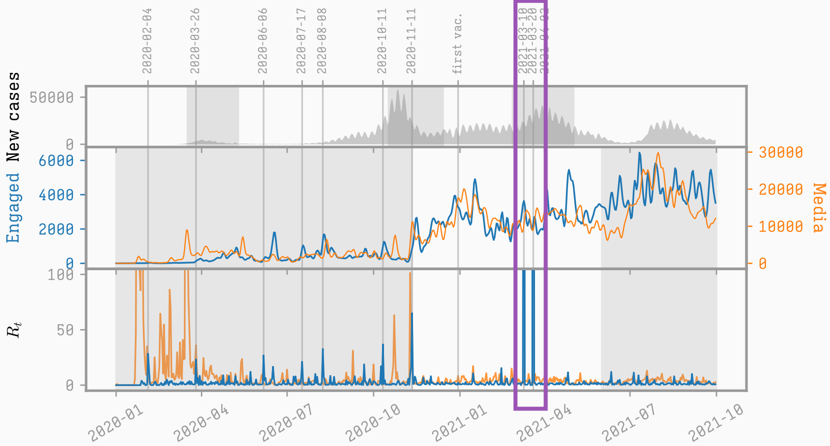


# Engagement evolution

Pfizer announcement / Holdup movie



suspension of Astrazeneca



## Tuberculosis under-detection

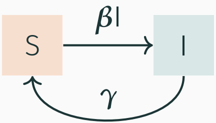
- 4 million of undetected/untreated TB cases, yearly [WHO]
- Hard to reach communities with high levels of TB incidence

**Efficient approach to ACF**

Data driven estimation of TB incidence rates to focus health-care interventions such as Active-Case Findings (ACF)

*Active-Case Finding:* systematic screening of the population for active TB cases.

Use of compartmental models (SIS) to disaggregate the reported cases.




**Assumption**

**Endemic** disease with slowly evolving **well mixed population**

$$\begin{cases} \frac{d\mathbf{S}}{dt} & = \gamma \mathbf{I} - \beta \frac{\mathbf{IS}}{\mathbf{S}+\mathbf{I}} \\ \frac{d\mathbf{I}}{dt} & = \beta \frac{\mathbf{IS}}{\mathbf{S}+\mathbf{I}} - \gamma \mathbf{I} \,. \end{cases}$$

$\longrightarrow$ fit the parameters to:

- the number of cases reported by the **local health system** subunits;
- the population density as estimated by [Worldpop]

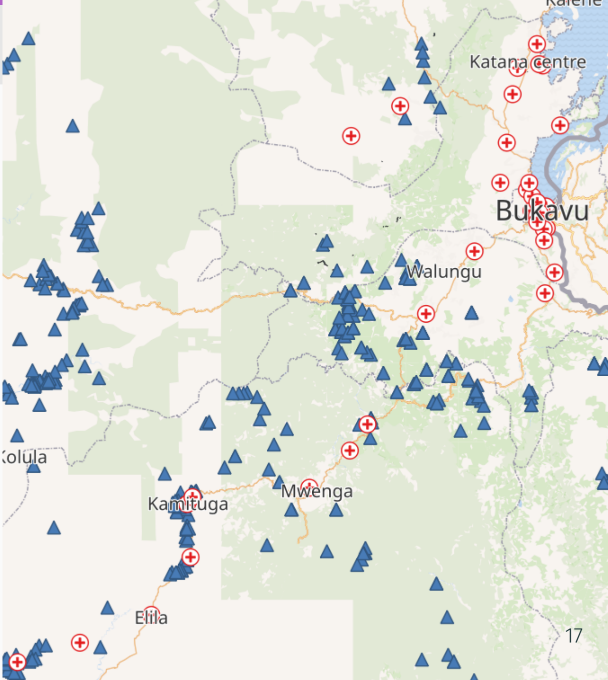
# Model refinement

We include additional data to refine the estimation:

- Mine locations;
- Health facilities,

from of openly available sources:

- OpenStreetMap
- IPIS Research Project (mines)
- Healthsites

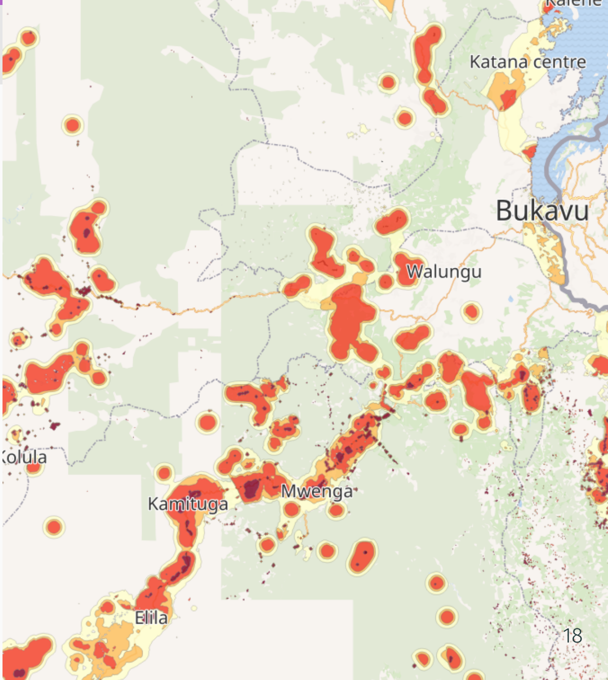


## Incidence rate estimation

The estimated incidence rates highlight the location of population pockets with high risk of TB.

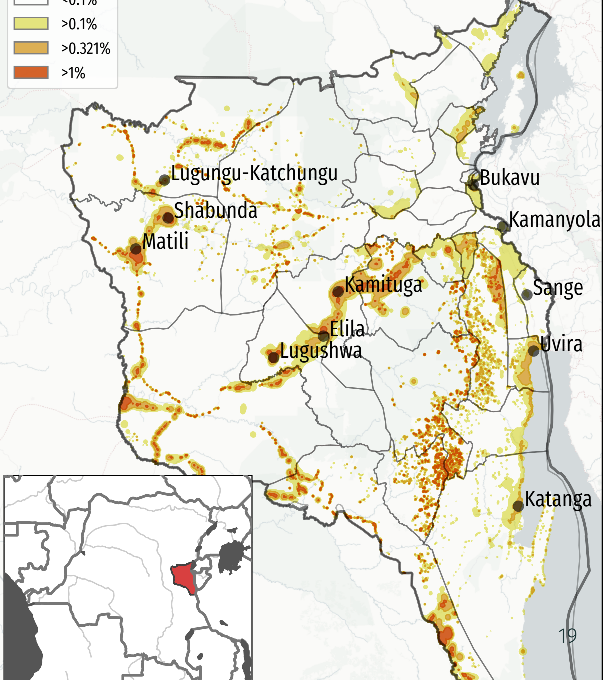
Leads to efficient ACF interventions.

| color | incidence rate |
| --- | --- |
| light yellow | >0.1% |
| orange | >0.32% |
| red | >0.5% |
| dark red | >1% |



## Multicentric clinical trial

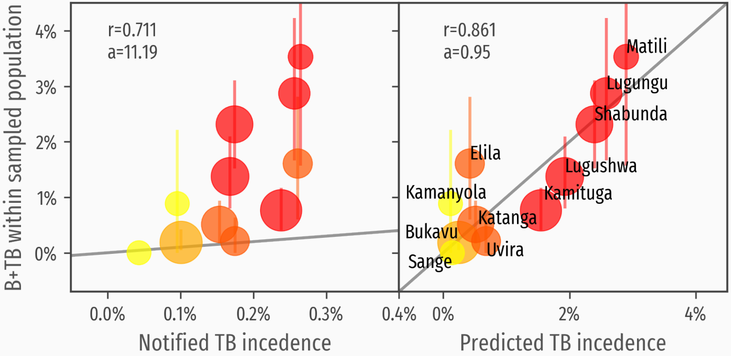
We performed a ***multicentric clinical trial*** addressing 11 locations with heterogeneous estimated incidence rates.



## Results

| | |
|---|---|
| screenings | 13.841 |
| lab tests | 1153 |
| positive cases | 112 |

> **80%** of positive cases originated from locations at high risk (estimated incidence rate higher than 1%).



# 🏃 Finally...

# Questions?

## Joint work with:

JC Delvenne

UCLouvain

M Schaub

RWTH AACHEN UNIVERSITY

F Gargiulo

SORBONNE UNIVERSITÉ

J Ward

Université Paris Cité

E. André

UZ LEUVEN

https://maurofaccin.github.io
mauro.fccn@gmail.com

Code at:

https://maurofaccin.github.io/aisa
https://maurofaccin.github.io/cartotb