

Séminaire MESCAL / MOAIS / NANOSIM

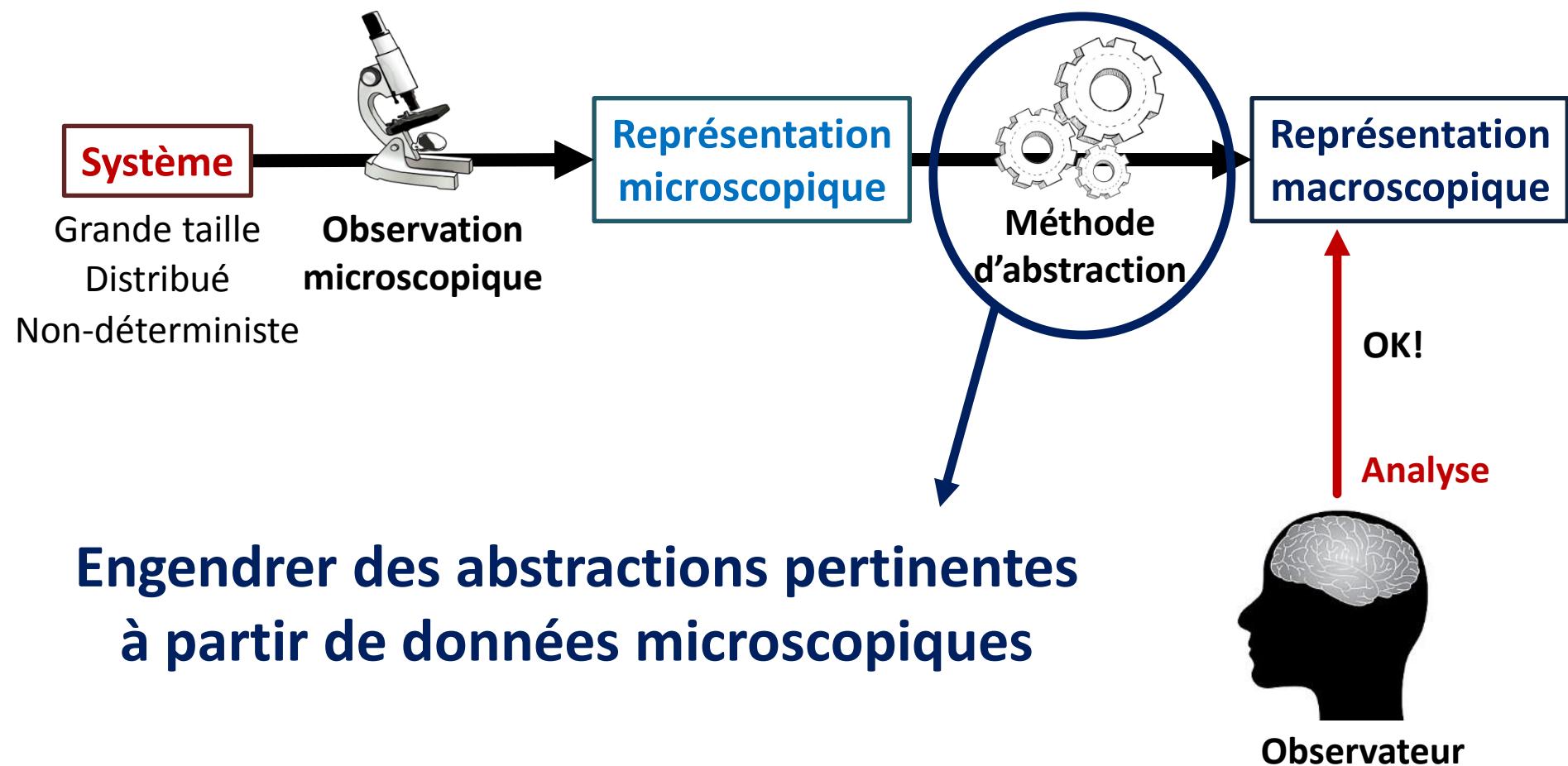
Montbonnot, 27 novembre 2014

Agrégation de données et analyse multiéchelle des systèmes complexes

Robin Lamarche-Perrin
Jean-Marc Vincent
Yves Demazeau



Analyse des grands systèmes



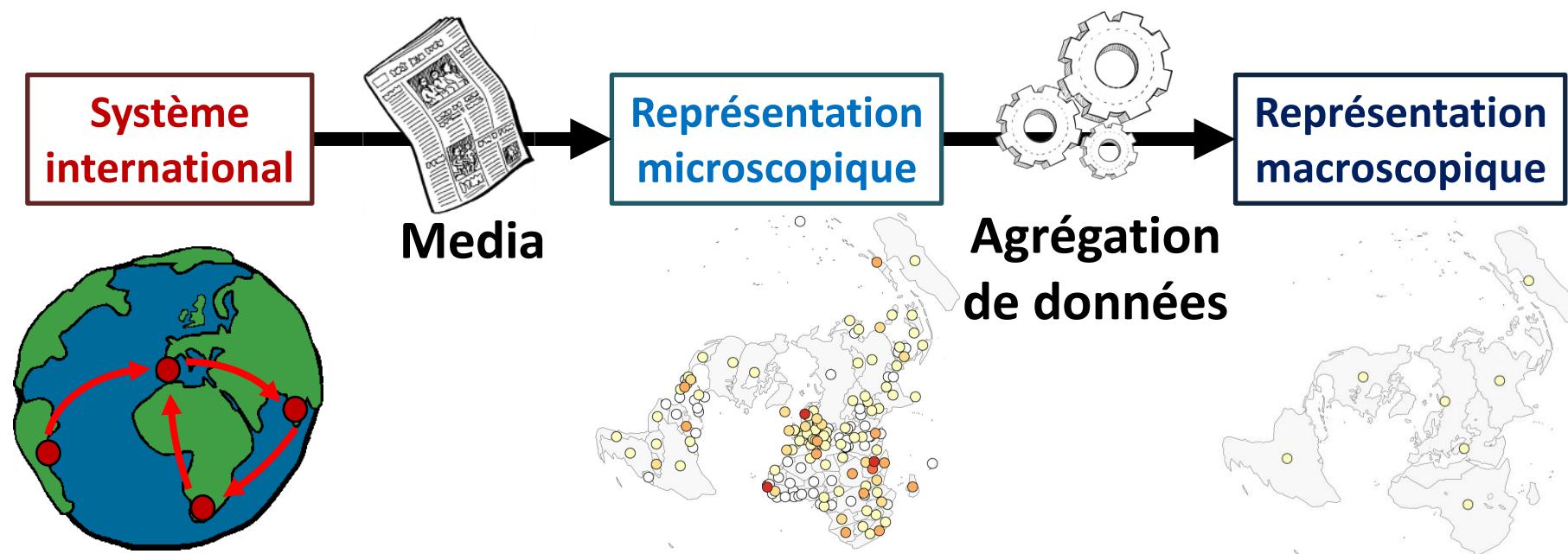
Analyse des relations internationales

Hypothèse : les médias constituent un instrument d'observation adéquat du niveau national

[Grasland *et al.*, 2011]



Géographe



Données médiatiques

THE GUARDIAN



“Japan”

THE TIMES OF INDIA



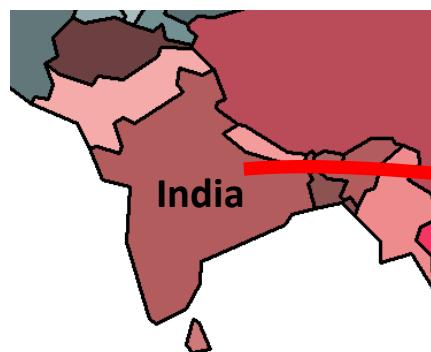
“Madrid”



“French”



“Spain”



Base de données GEOMEDIA
(ANR CORPUS GUI-AAP-04)

150 journaux

1,530,000 articles

INFORMATION GÉOGRAPHIQUE
193 pays (membres de l'ONU)

Données médiatiques

THE GUARDIAN

article 1



“Japan”

30 mai 2011

THE TIMES OF INDIA

article 2



“Madrid”

30 mai 2011

article 3



“French”

19 juillet 2012

Base de données GEOMEDIA
(ANR CORPUS GUI-AAP-04)

150 journaux

1,530,000 articles

INFORMATION GÉOGRAPHIQUE
193 pays (membres de l'ONU)

INFORMATION TEMPORELLE

630 jours / 90 semaines
(du 3 mai 2011 au 20 janvier 2013)

Représentation microscopique du système international

Journal LE MONDE

π

Espace

	USA	Libye	Syrie	France	Israël	...	Total
2 mai	25	12	11	10	4	...	142
9 mai	14	6	12	12	5	...	108
16 mai	20	11	12	6	9	...	142
23 mai	15	9	6	13	5	...	120
30 mai	10	16	17	9	4	...	137
6 juin	14	16	16	9	4	...	114
13 juin	15	14	17	9	6	...	119
20 juin	17	13	12	12	7	...	123
27 juin	7	6	7	20	2	...	103
4 juill.	12	13	8	10	6	...	129
11 juill.	21	10	10	14	3	...	107
18 juill.	7	3	8	4	5	...	61
25 juill.	16	7	6	13	4	...	128
1 août	21	1	9	7	4	...	88
...
Total	423	308	260	248	153	...	3520

t

Temps

Nombre de citations observées

$$v(\pi, t)$$

$$v(., t)$$

$$v(\pi, .)$$

$$v(., .)$$

Nombre de citations attendues

$$v^*(\pi, t) = \frac{v(\pi, .) v(., t)}{v(., .)}$$

Représentation géographique

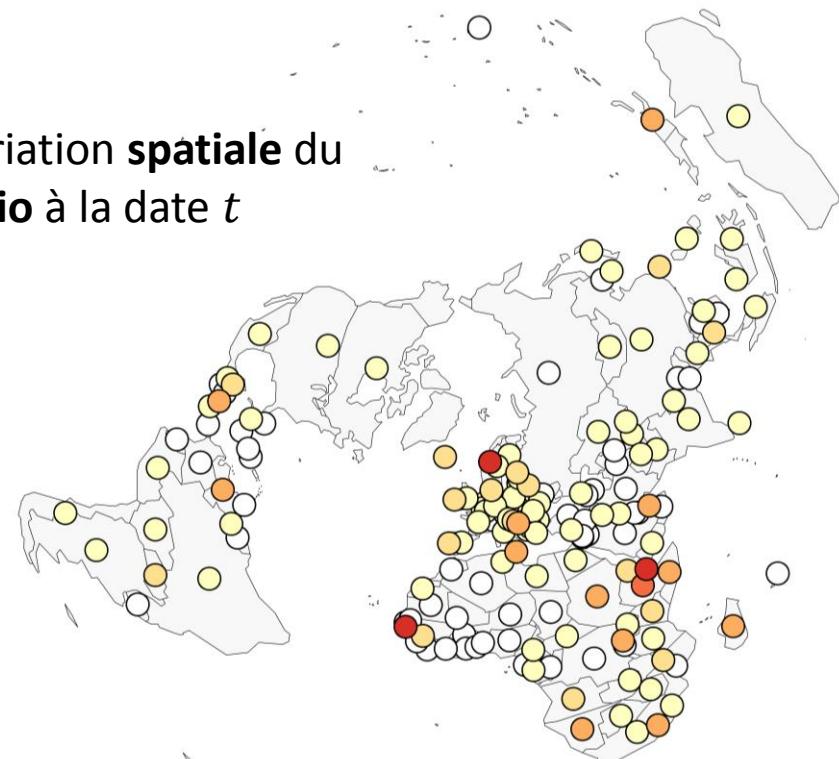
Espace

T
e
m
p
s

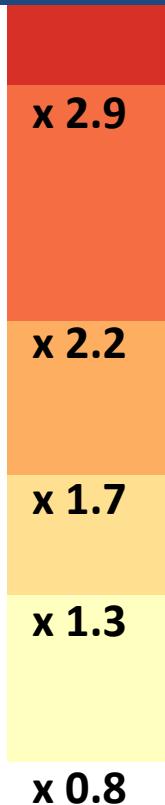
	USA	Libye	Syrie	France	Israël	...	Total
2 mai	25	12	11	10	4	...	142
9 mai	14	6	12	12	5	...	108
16 mai	20	11	12	6	9	...	142
23 mai	15	9	6	13	5	...	120
30 mai	10	16	17	9	4	...	137
6 juin	14	16	16	9	4	...	114
13 juin	15	14	17	9	6	...	119
20 juin	17	13	12	12	7	...	123
27 juin	7	6	7	20	2	...	103
4 juill.	12	13	8	10	6	...	129
11 juill.	21	10	10	14	3	...	107
18 juill.	7	3	8	4	5	...	61
25 juill.	16	7	6	13	4	...	128
1 août	21	1	9	7	4	...	88
...
Total	423	308	260	248	153	...	3520

Ratio des citations observées et des citations attendues

$$\rho(\pi, t) = \frac{v(\pi, t)}{v^*(\pi, t)} = \frac{v(\pi, t) v(., .)}{v(\pi, .) v(., t)}$$

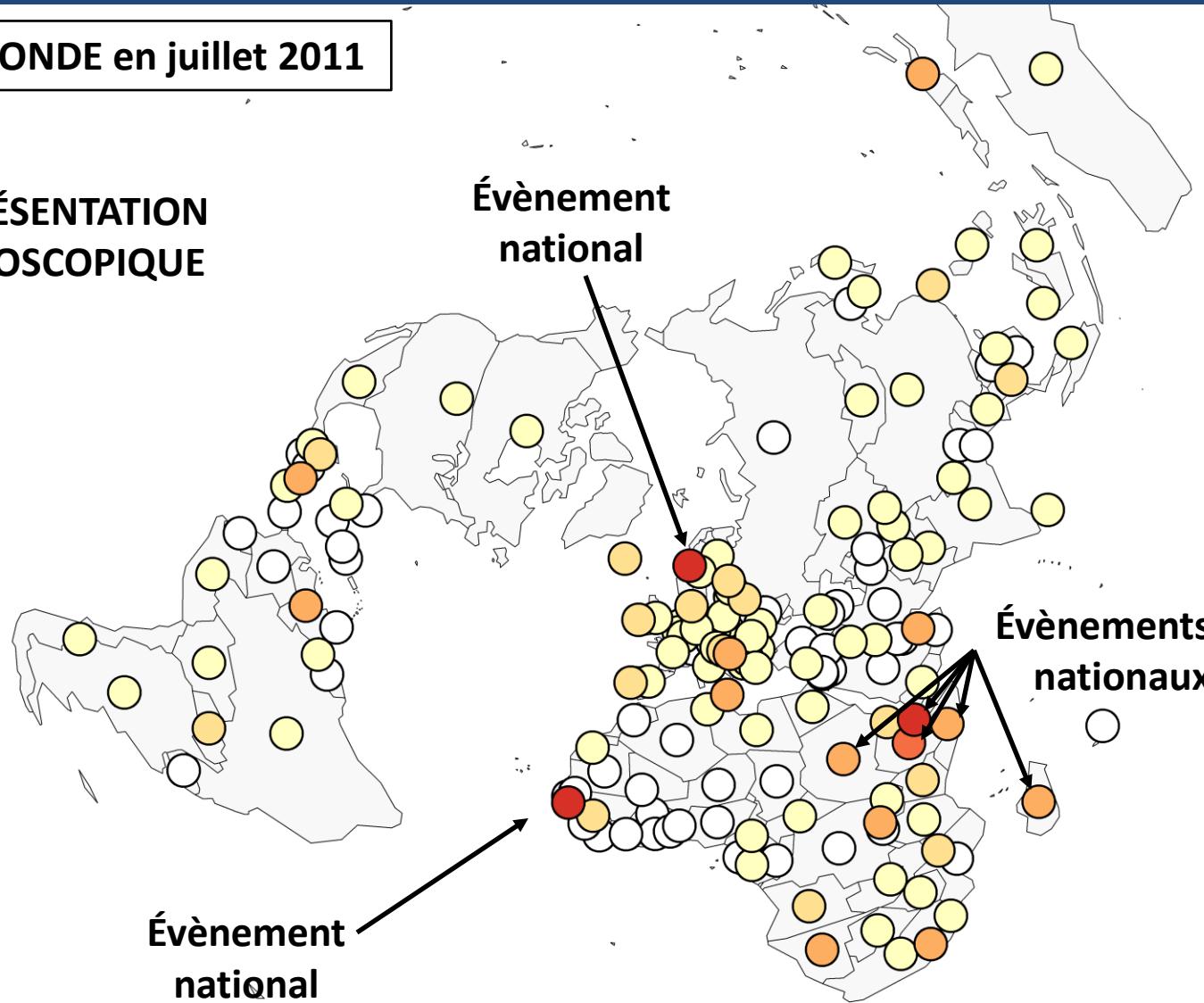


Détection d'évènements médiatiques

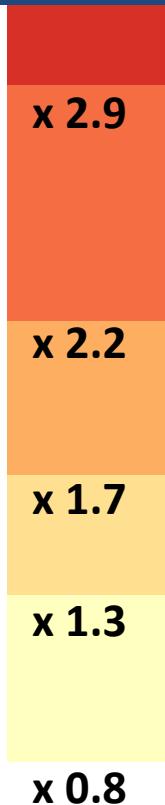


Journal LE MONDE en juillet 2011

REPRÉSENTATION
MICROSCOPIQUE

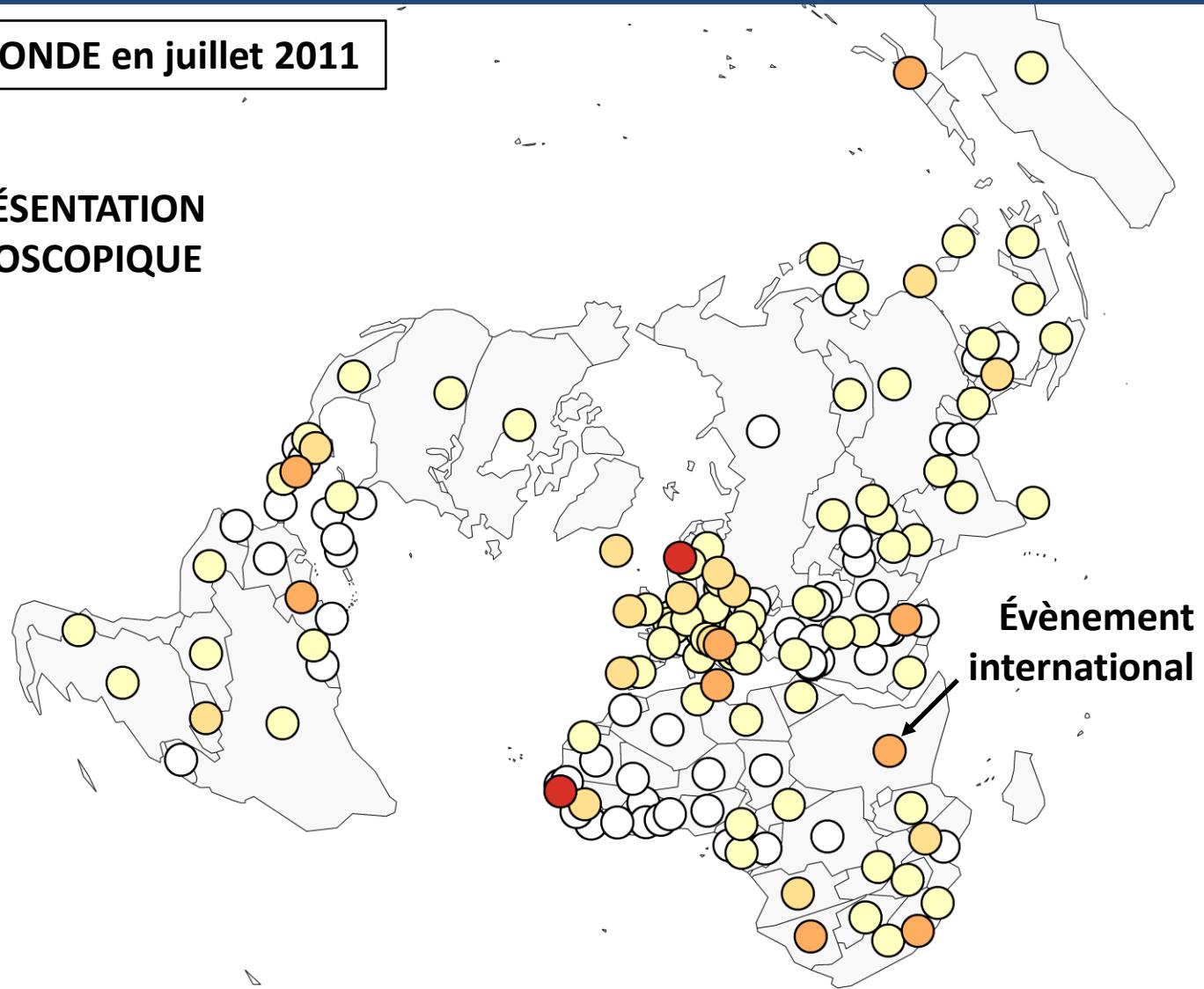


Détection d'évènements médiatiques



Journal LE MONDE en juillet 2011

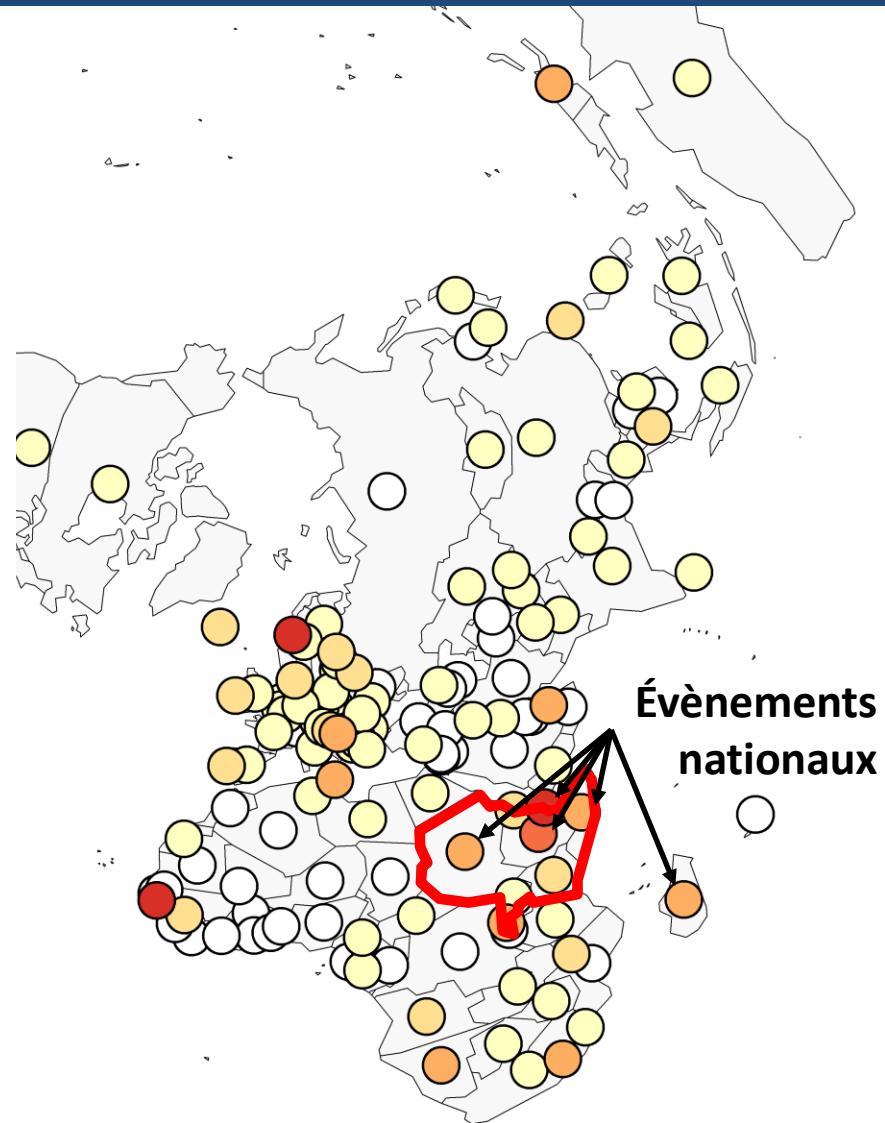
REPRÉSENTATION
MICROSCOPIQUE



Agrégation de données

Espace

	USA	Libye	Syrie	France	Israël	...	Total
2 mai	25	12	11	10	4	...	142
9 mai	14	6	12	12	5	...	108
16 mai	20	11	12	6	9	...	142
23 mai	15	9	6	13	5	...	120
30 mai	10	16	17	9	4	...	137
6 juin	14	16	16	9	4	...	114
13 juin	15	14	17	9	6	...	119
20 juin	17	13	12	12	7	...	123
27 juin	7	6	7	20	2	...	103
4 juill.	12	13	8	10	6	...	129
11 juill.	21	10	10	14	3	...	107
18 juill.	7	3	8	4	5	...	61
25 juill.	16	7	6	13	4	...	128
1 août	21	1	9	7	4	...	88
...
Total	423	308	260	248	153	...	3520



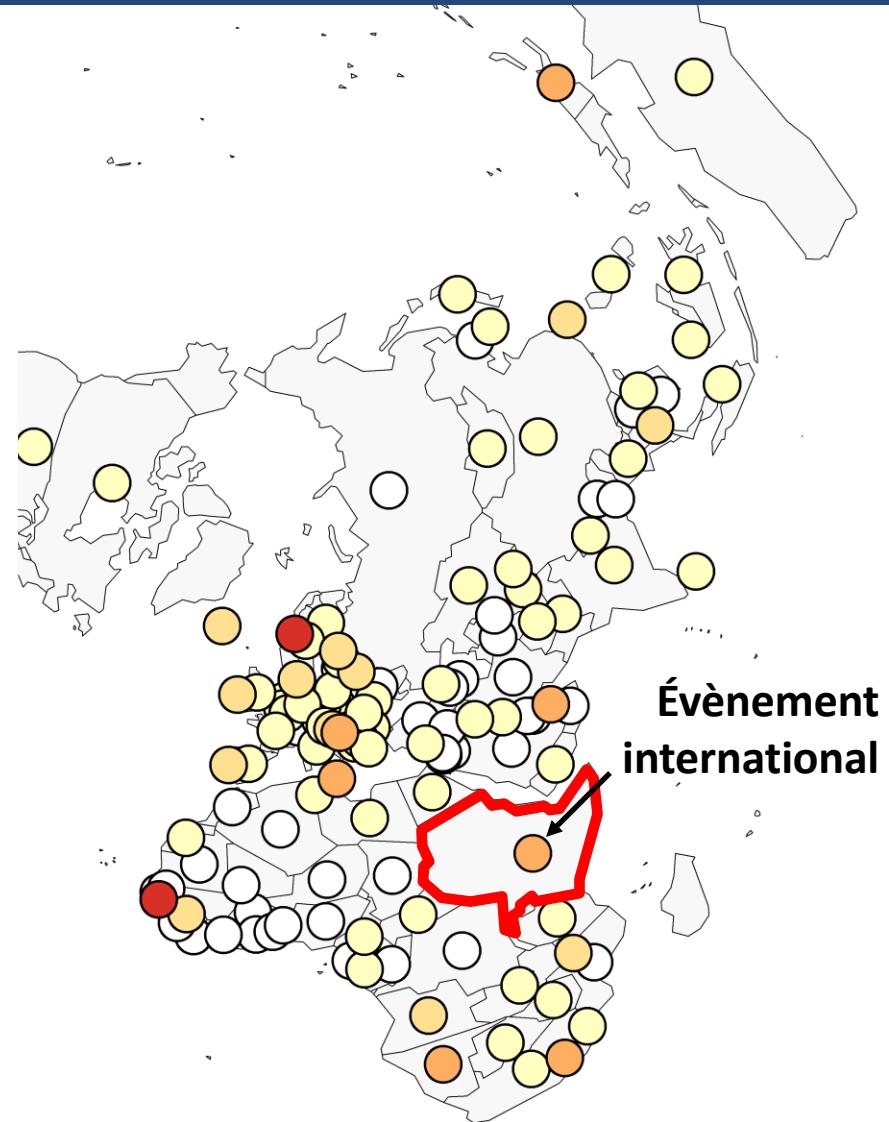
Agrégation de données

Espace

π₁ π₂ π₃

↑
Temp
↓

	USA	Agrégat	Israël	...	Total
2 mai	25	13+11+10	4	...	142
9 mai	14	6+12+12	5	...	108
16 mai	20	11+12+6	9	...	142
23 mai	15	9+6+13	5	...	120
30 mai	10	16+17+9	4	...	137
6 juin	14	16+16+9	4	...	114
13 juin	15	14+17+9	6	...	119
20 juin	17	13+12+12	7	...	123
27 juin	7	6+7+20	2	...	103
4 juill.	12	13+8+10	6	...	129
11 juill.	21	10+10+14	3	...	107
18 juill.	7	3+8+4	5	...	61
25 juill.	16	7+6+13	4	...	128
1 août	21	1+9+7	4	...	88
...
Total	423	308+260+248	153	...	3520



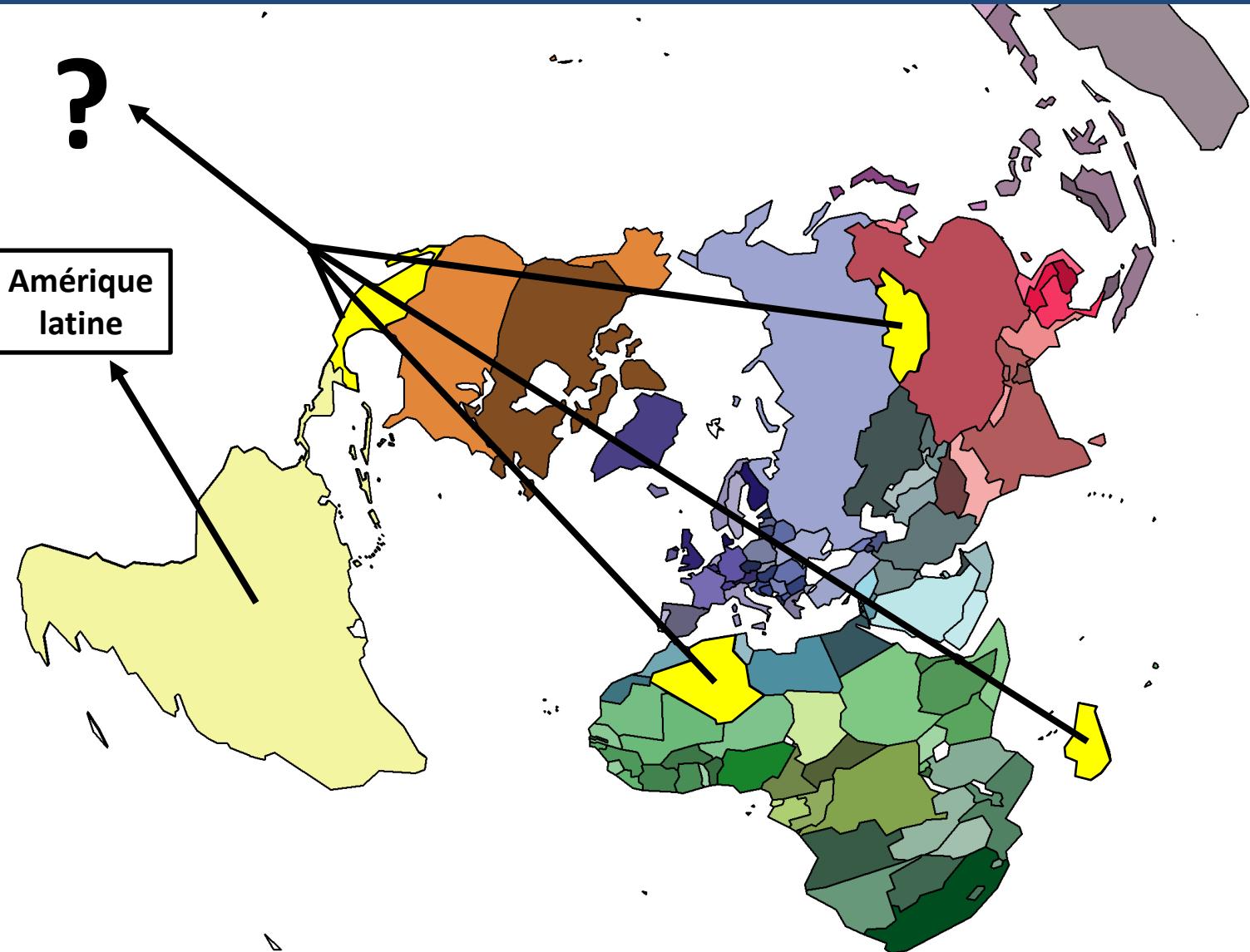
P1 : Sémantique géographique des agrégats



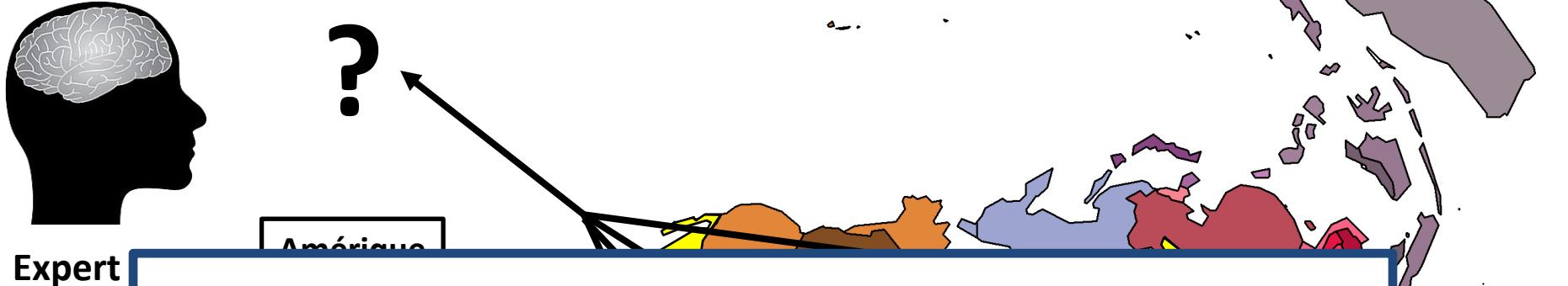
Expert

?

Amérique
latine

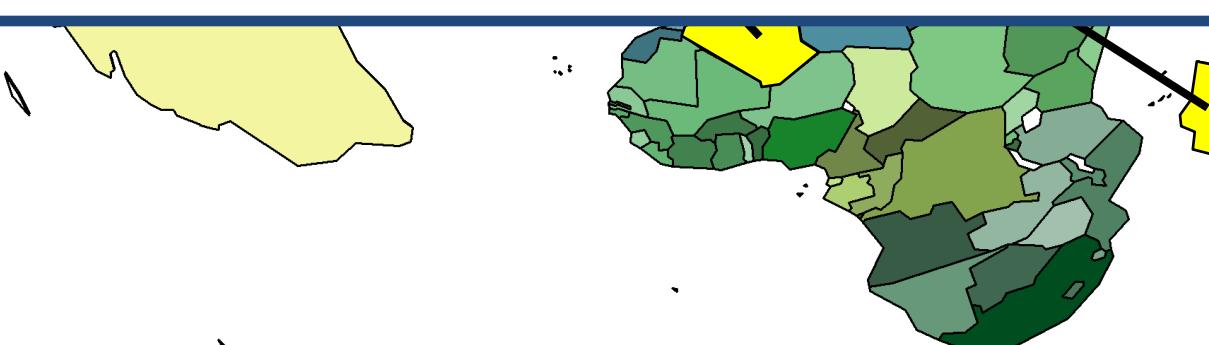


P1 : Sémantique géographique des agrégats

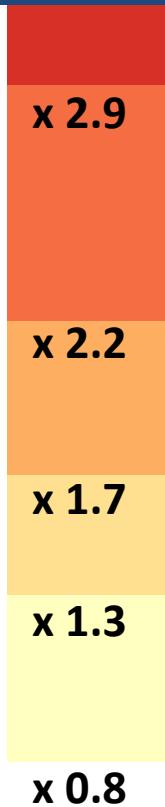


Expert

Problème 1 : Comment engendrer des abstractions cohérentes avec l'espace géographique ?

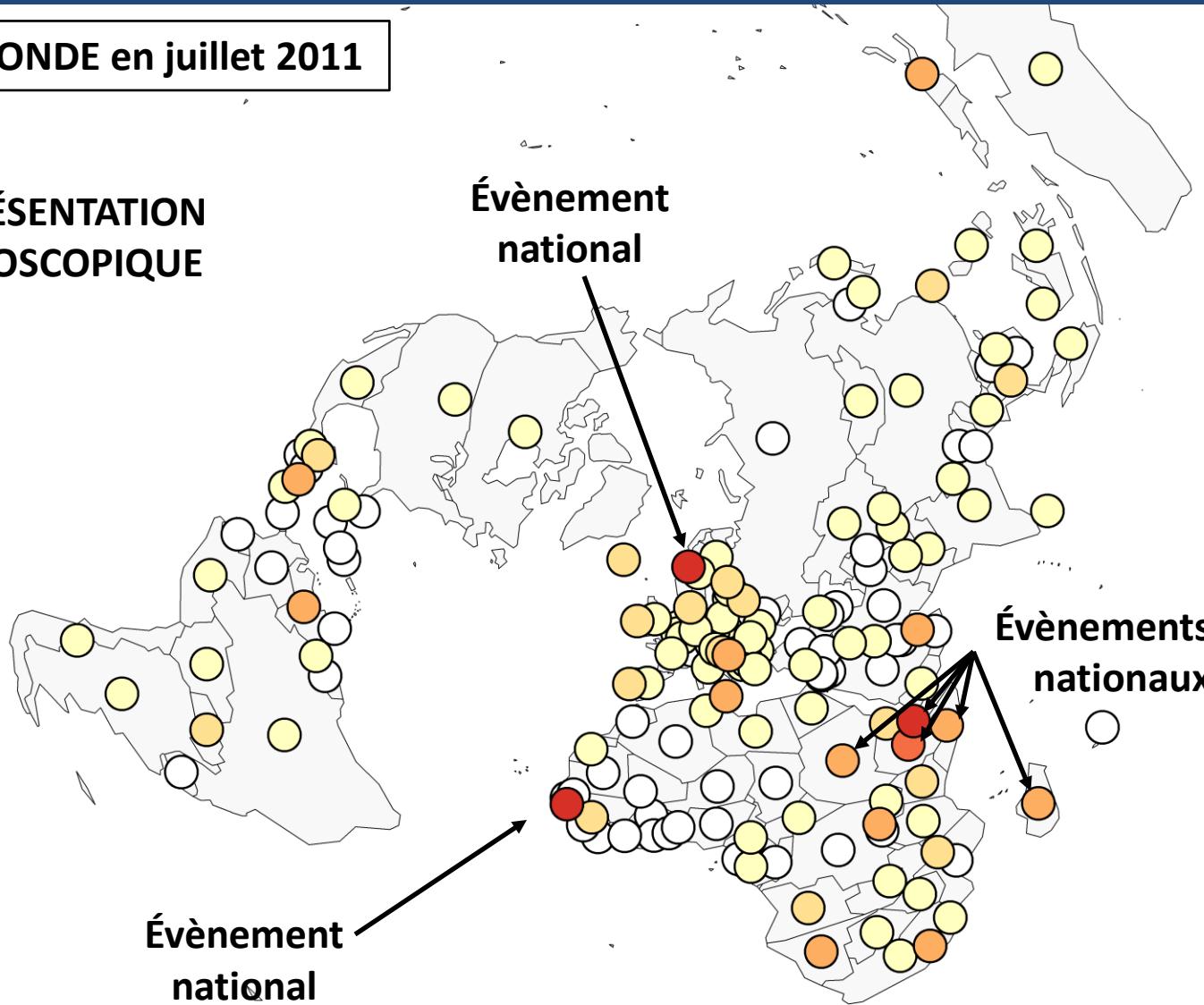


P2 : Niveaux de représentation

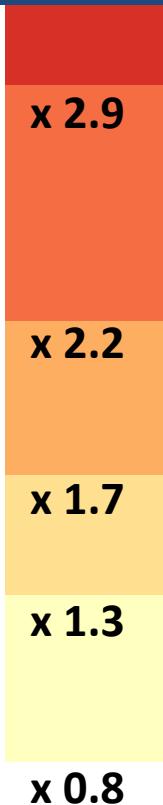


Journal LE MONDE en juillet 2011

REPRÉSENTATION
MICROSCOPIQUE

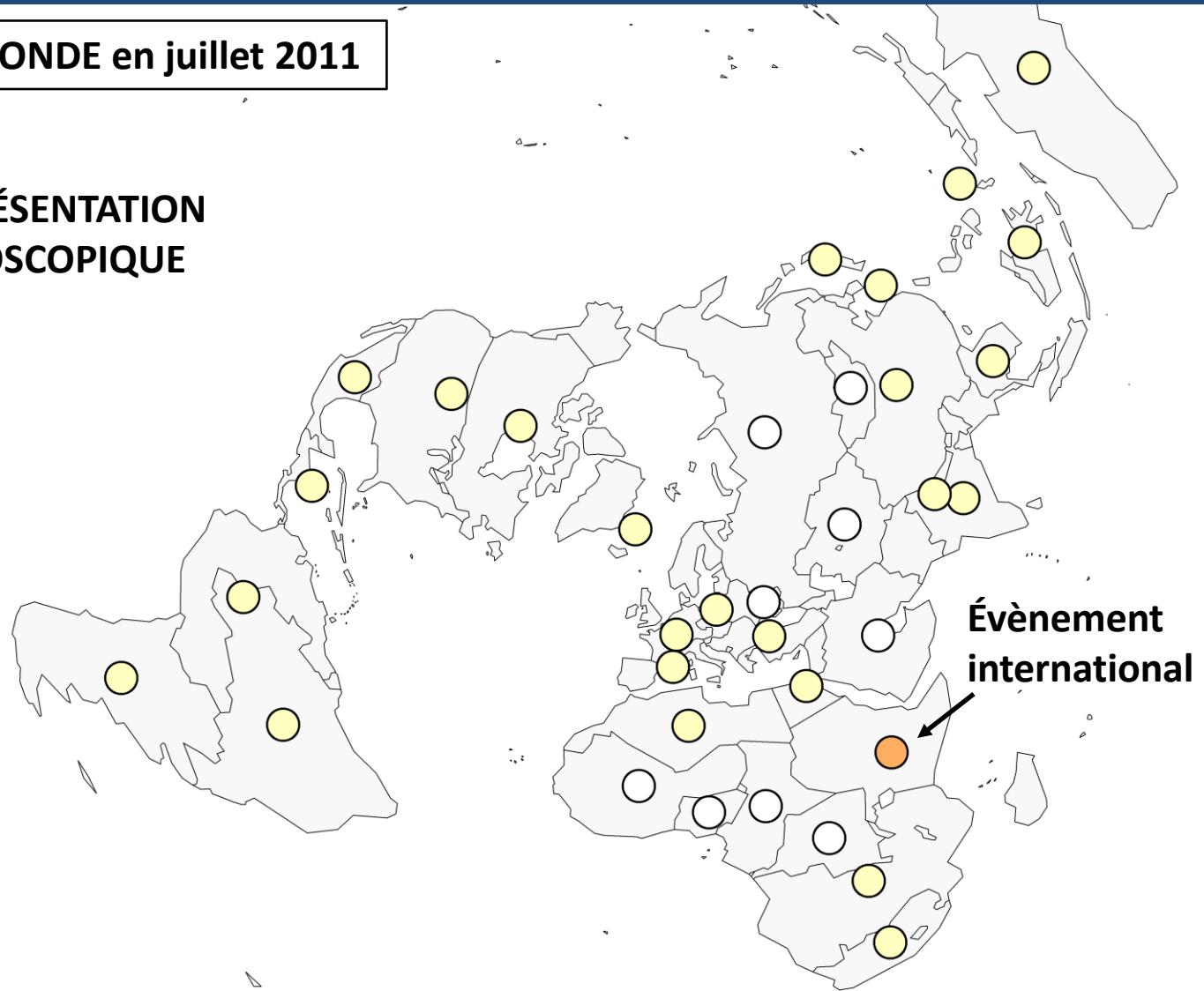


P2 : Niveaux de représentation

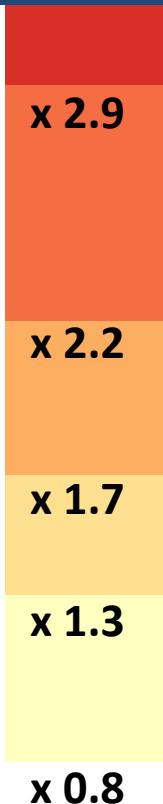


Journal LE MONDE en juillet 2011

REPRÉSENTATION MÉSOSCOPIQUE

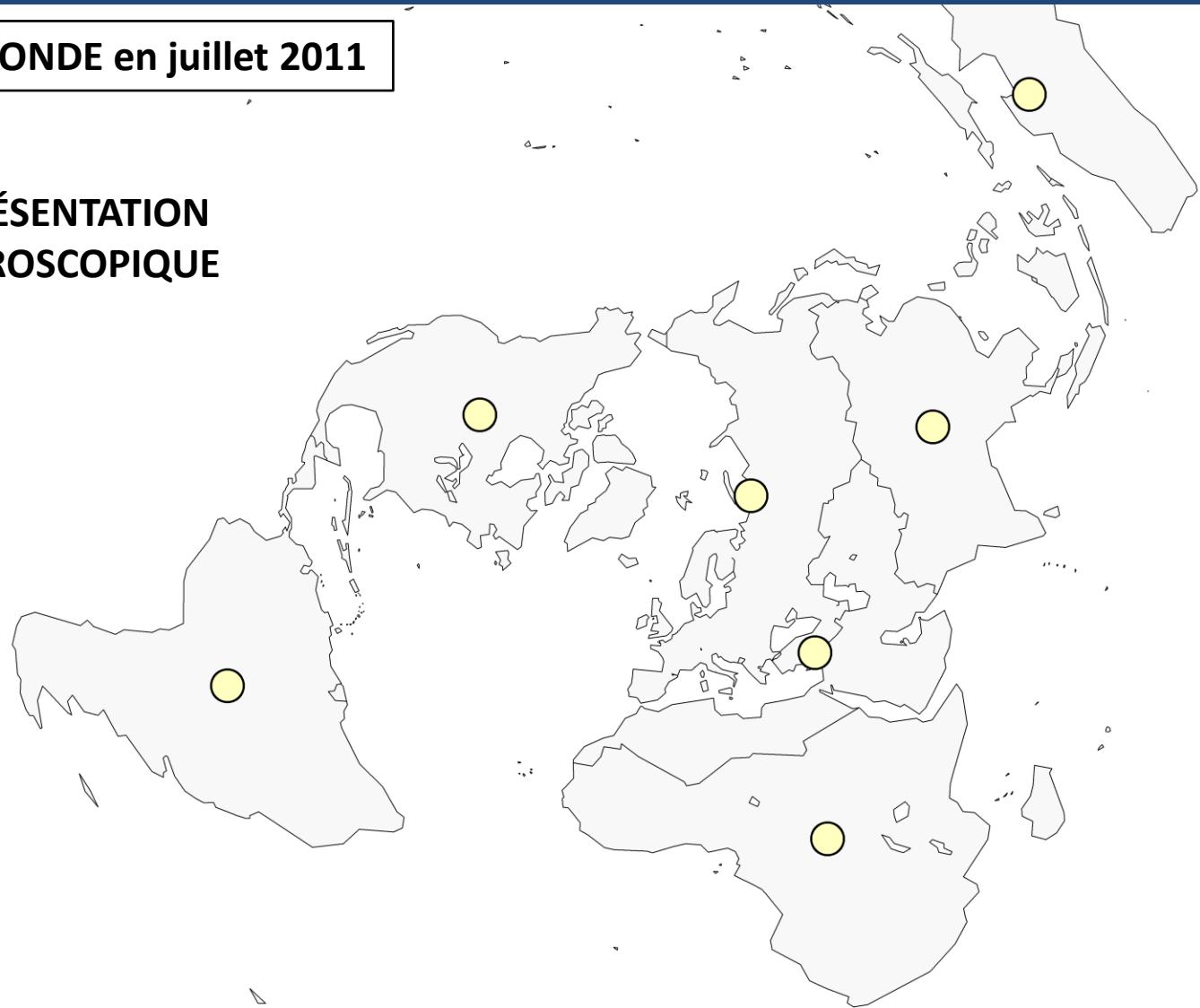


P2 : Niveaux de représentation

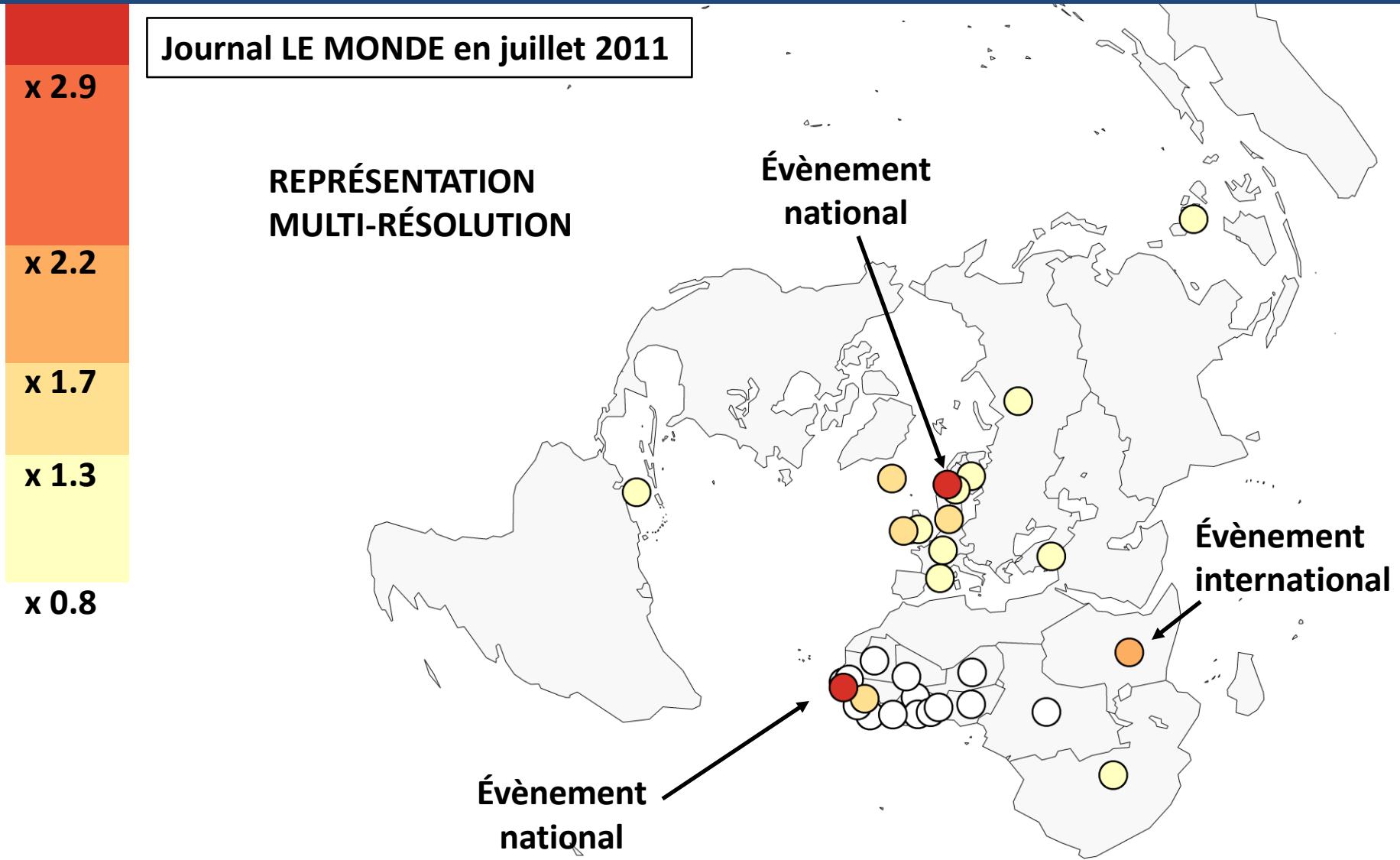


Journal LE MONDE en juillet 2011

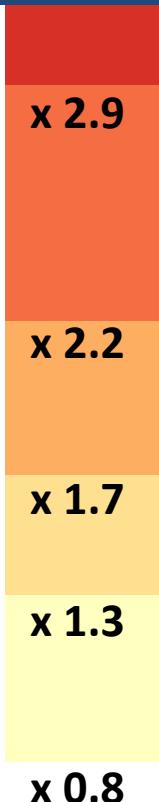
REPRÉSENTATION
MACROSCOPIQUE



P2 : Niveaux de représentation



P2 : Niveaux de représentation



Journal LE MONDE en juillet 2011

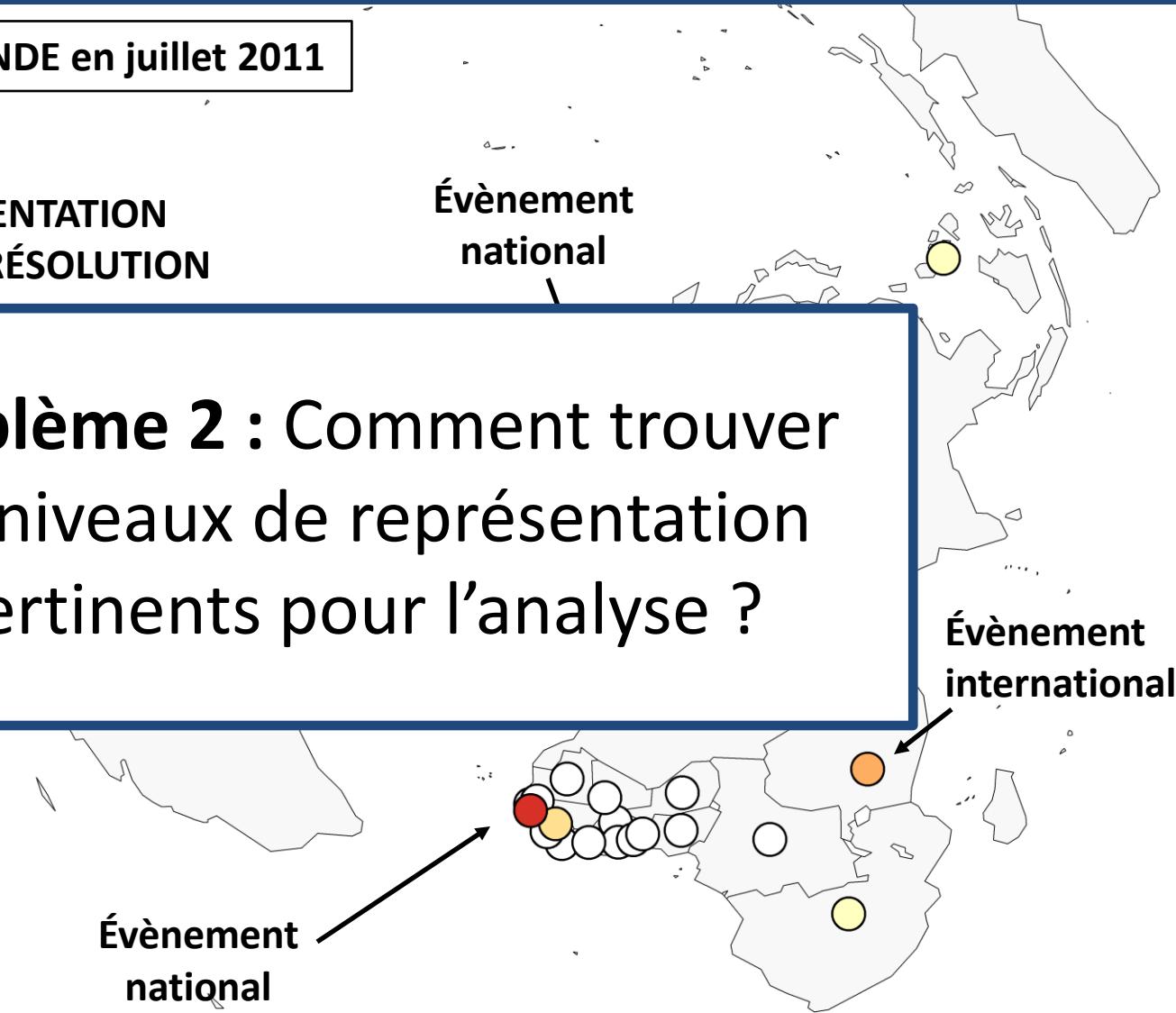
REPRÉSENTATION
MULTI-RÉSOLUTION

**Problème 2 : Comment trouver
les niveaux de représentation
pertinents pour l'analyse ?**

Évènement
national

Évènement
international

Évènement
national



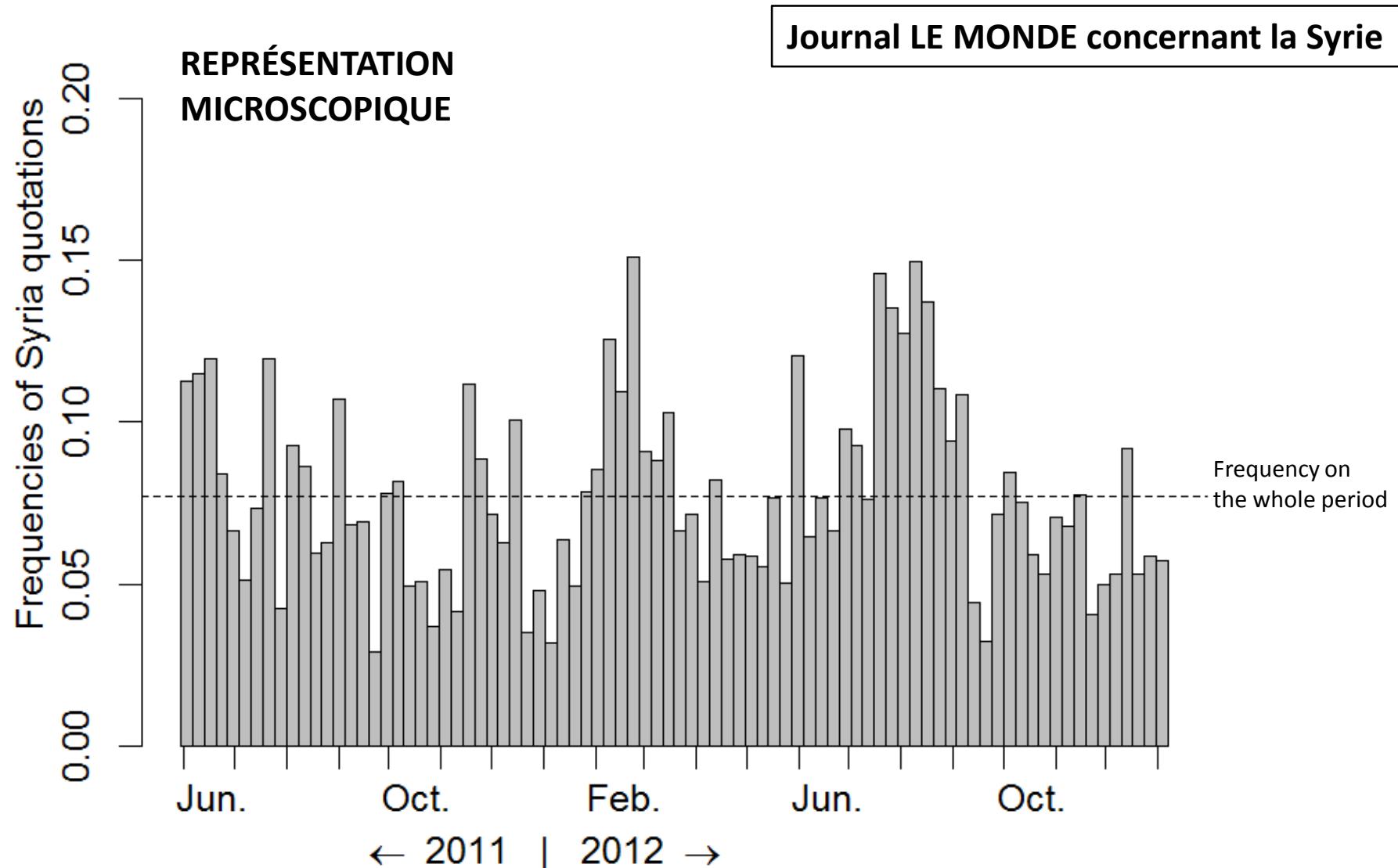
P3 : Calcul des meilleures représentations

- Le nombre de représentations possibles dépend **exponentiellement** de la taille de la représentation microscopique

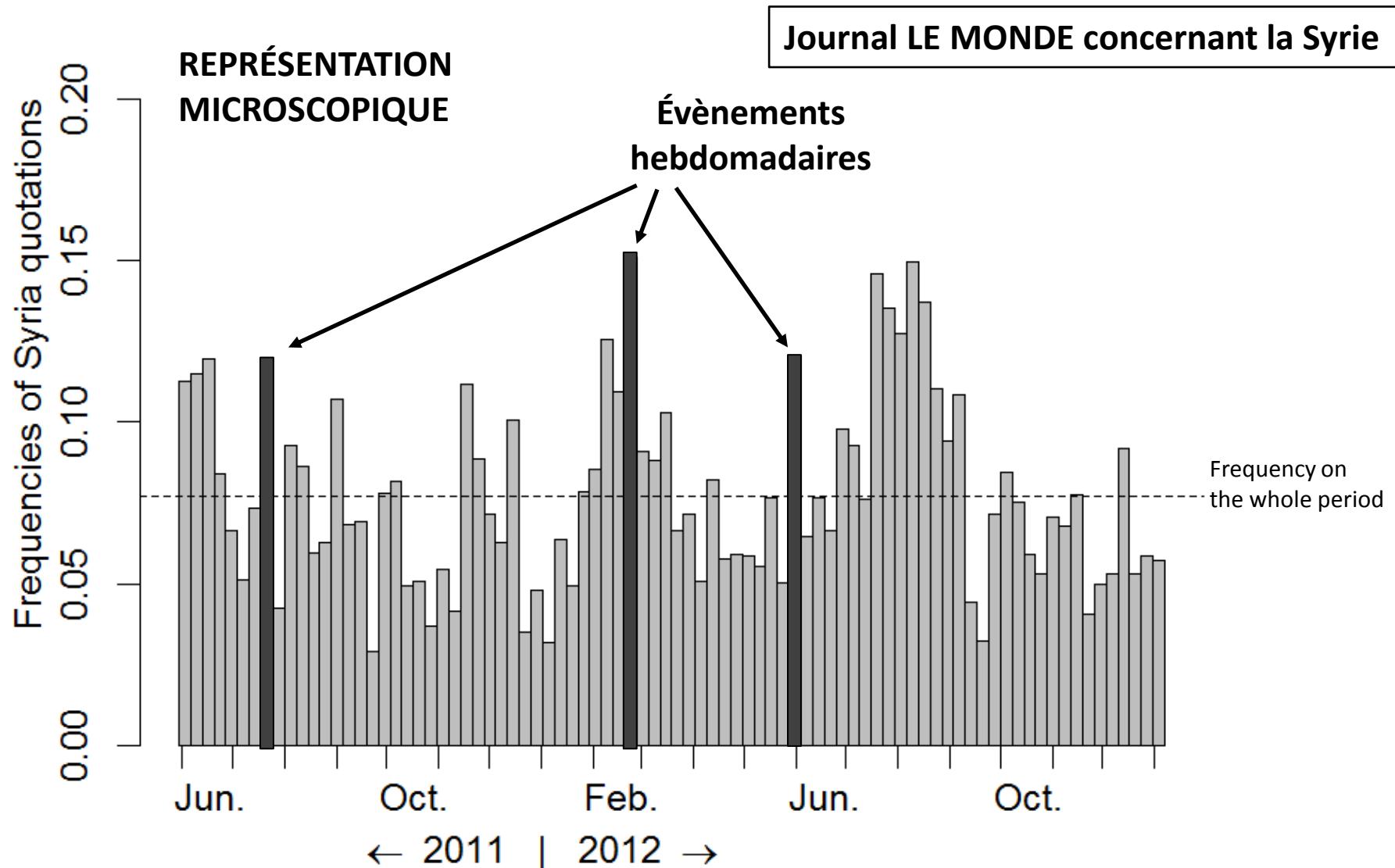
P3 : Calcul des meilleures représentations

Problème 3 : Comment calculer les « meilleures » représentations de manière efficace ?

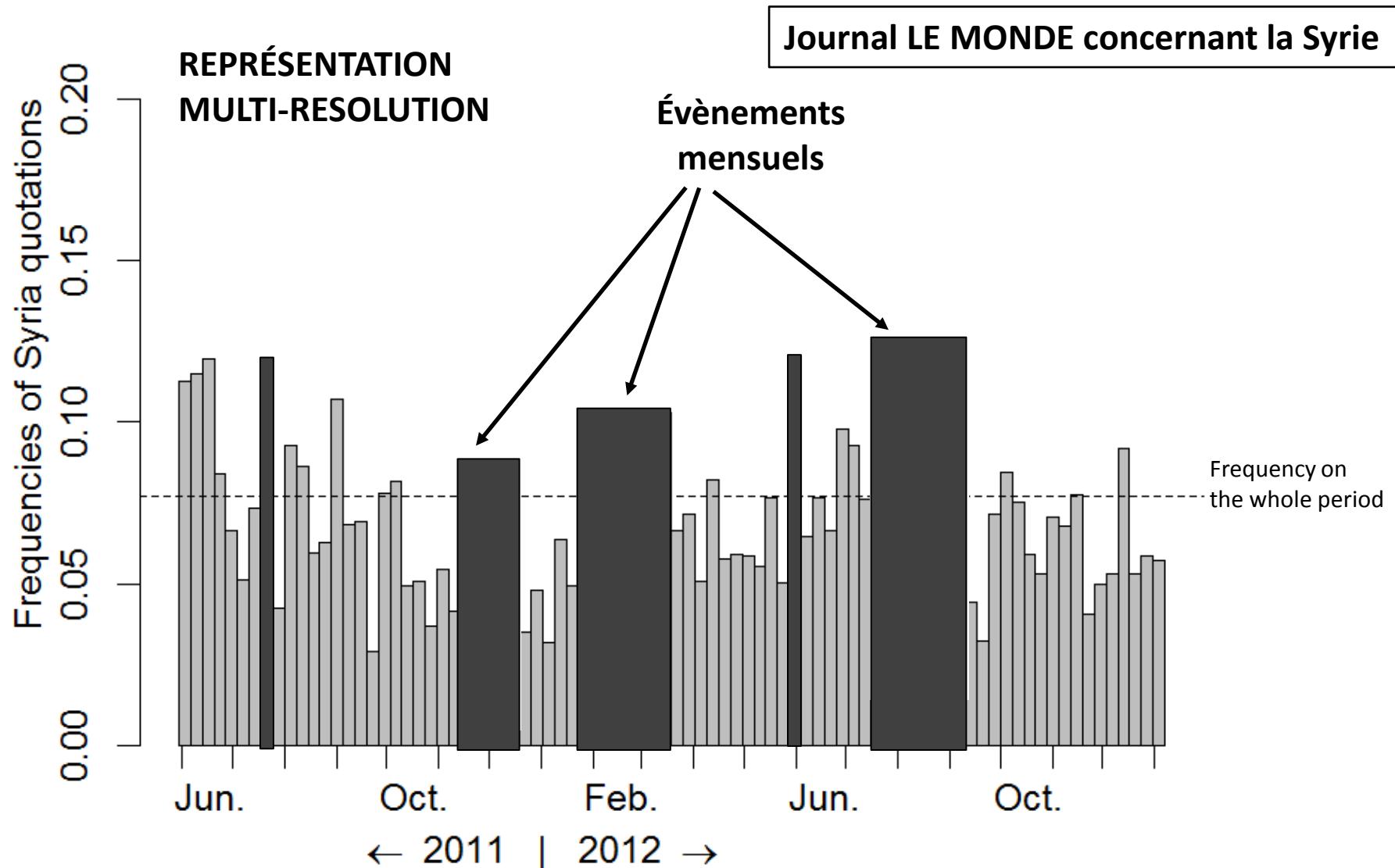
Et pour d'autres dimensions du système ?



Et pour d'autres dimensions du système ?



Et pour d'autres dimensions du système ?



Mon approche

P0

Caractériser le processus d'agrégation

→ Algèbre des partitions possibles

P1

Conserver la sémantique du système

→ Contraindre le partitionnement

Agréger selon plusieurs dimensions

→ Des contraintes adaptées aux topologies du système

P2

Évaluer et comparer les représentations

→ Mesurer la complexité et l'information

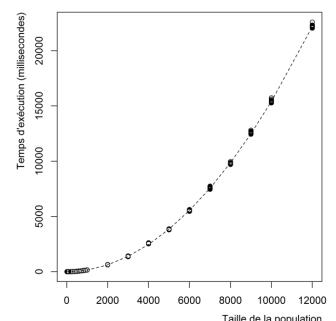
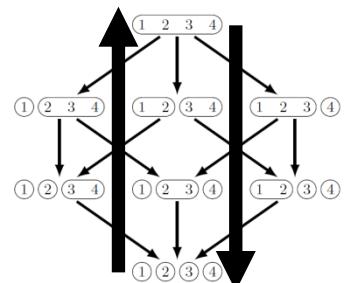
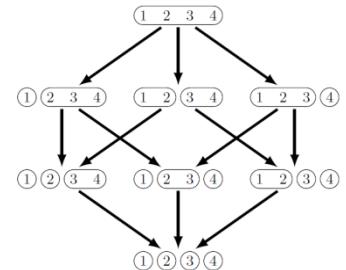
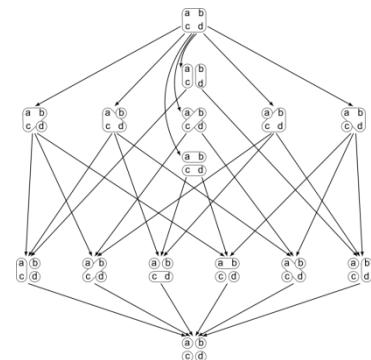
Proposer plusieurs niveau de granularité

→ Optimiser un compromis de qualité

P3

Calculer les meilleures représentations

→ Algorithme générique d'optimisation sous contraintes



Mon approche

P0

Caractériser le processus d'agrégation

→ Algèbre des partitions possibles

P1

Conserver la sémantique du système

→ Contraindre le partitionnement

Agréger selon plusieurs dimensions

→ Des contraintes adaptées aux topologies du système

P2

Évaluer et comparer les représentations

→ Mesurer la complexité et l'information

Proposer plusieurs niveau de granularité

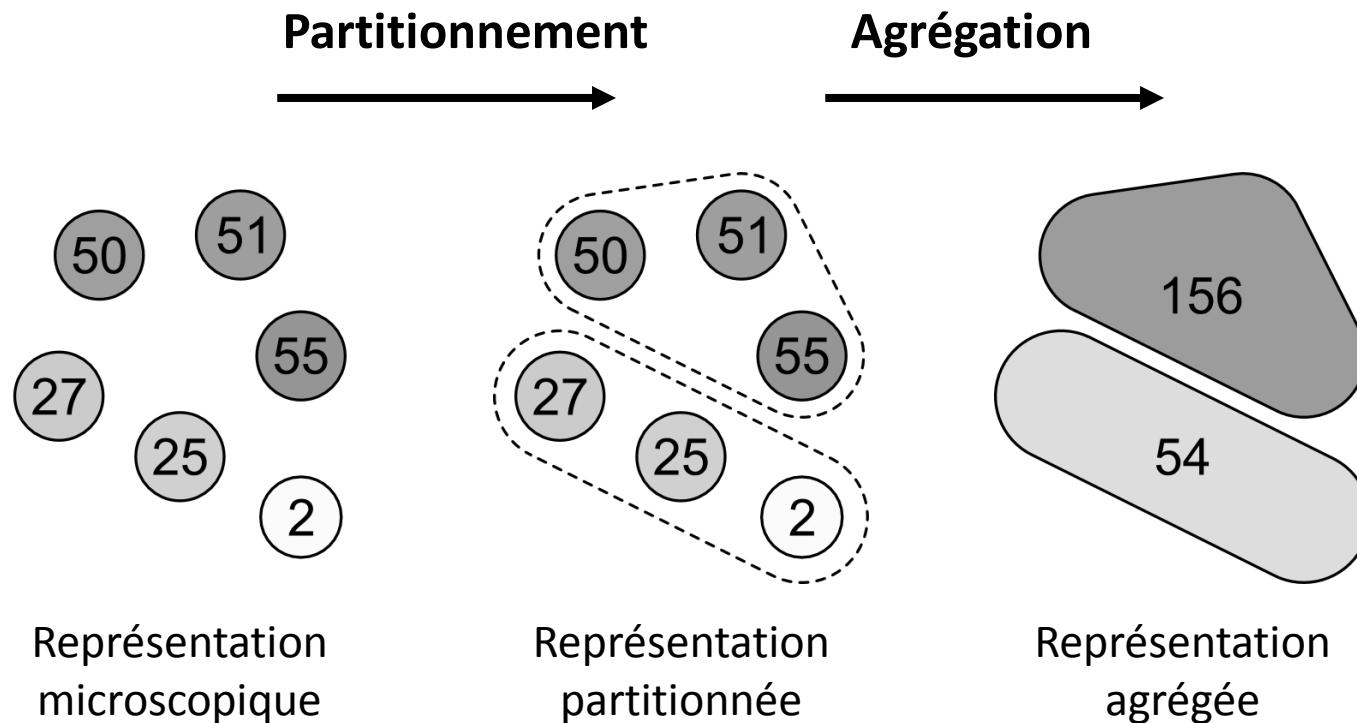
→ Optimiser un compromis de qualité

P3

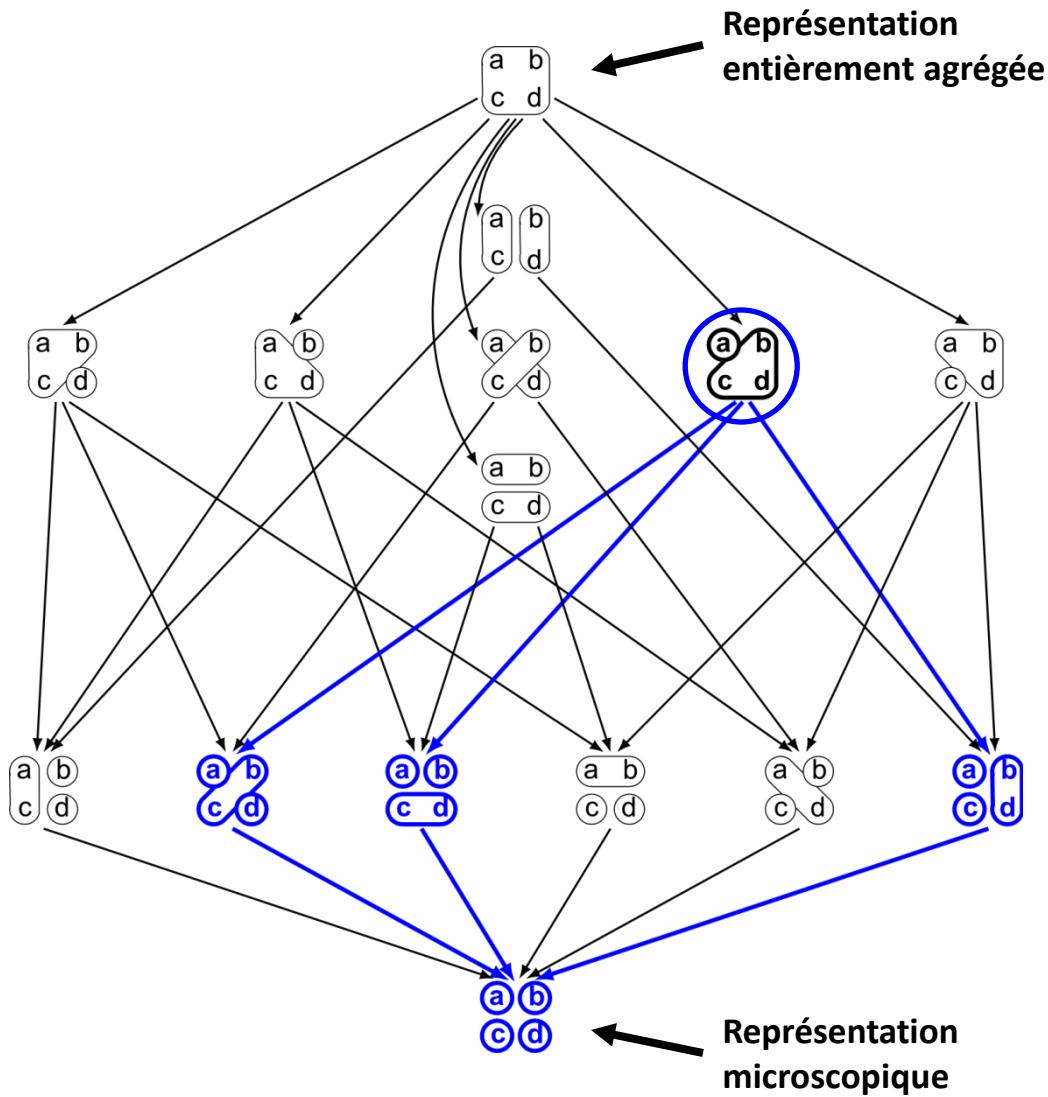
Calculer les meilleures représentations

→ Algorithme générique d'optimisation sous contraintes

Le processus d'agrégation



Ensemble des partitions possibles

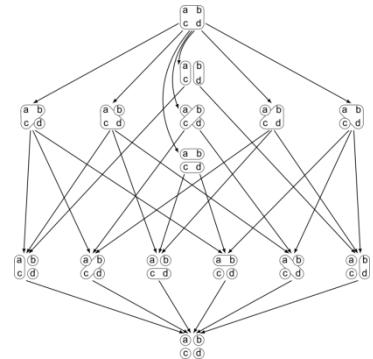


Structure algébrique

Ordre partiel sur l'ensemble des partitions possibles

→ relations de raffinement

Mon approche



P0 Caractériser le processus d'agrégation

→ Algèbre des partitions possibles

P1 Conserver la sémantique du système

→ Contraindre le partitionnement

Agréger selon plusieurs dimensions

→ Des contraintes adaptées aux topologies du système

P2 Évaluer et comparer les représentations

→ Mesurer la complexité et l'information

Proposer plusieurs niveau de granularité

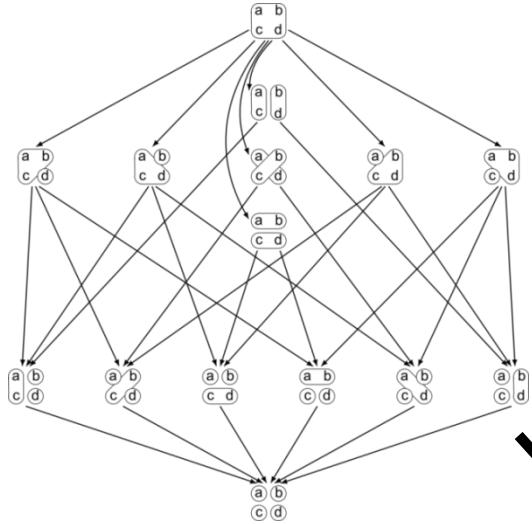
→ Optimiser un compromis de qualité

P3 Calculer les meilleures représentations

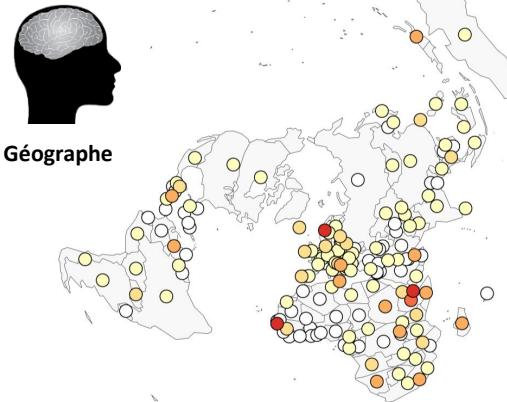
→ Algorithme générique d'optimisation sous contraintes

Problème et objectif

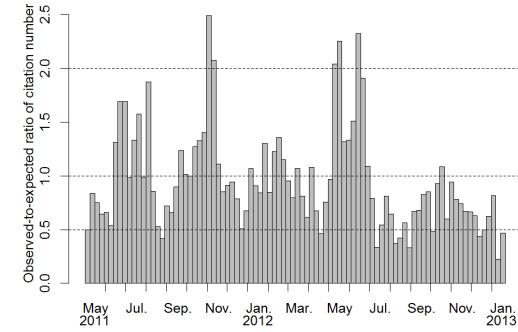
Ensemble des partitions possibles



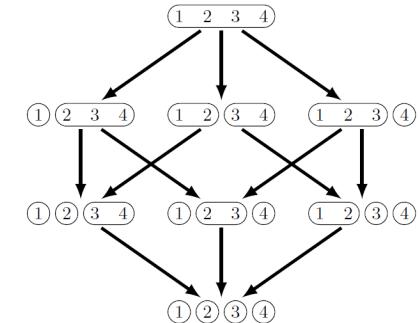
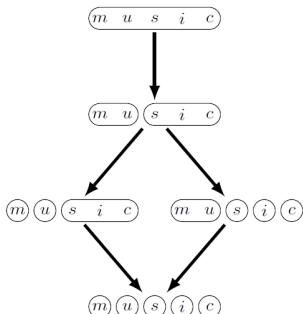
Sémantique géographique



Sémantique temporelle

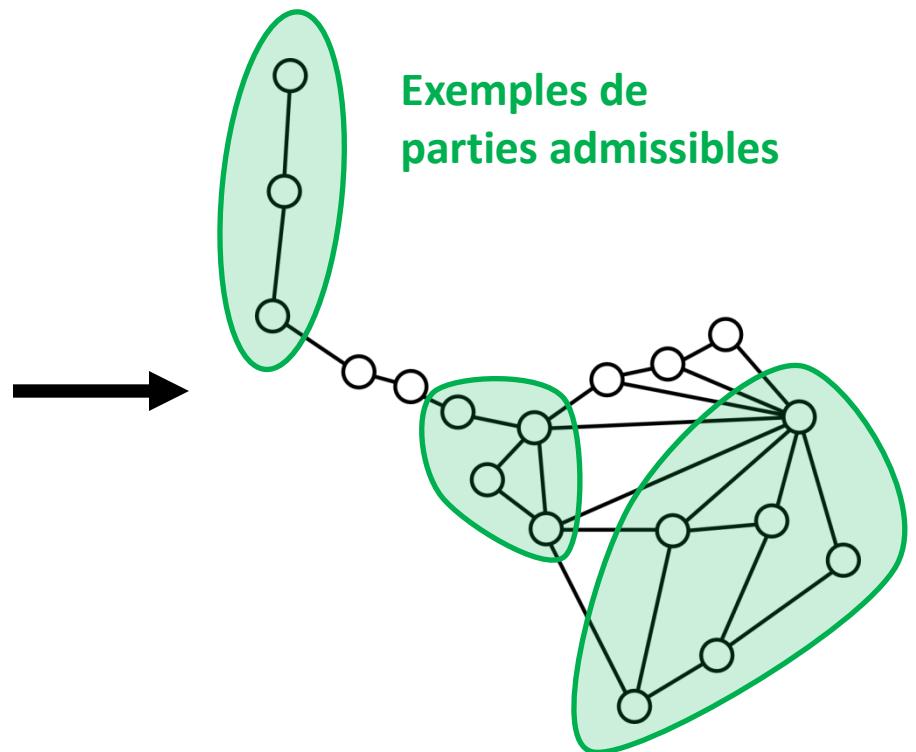


**Construire l'ensemble
des partitions admissibles**



Conserver la relation de voisinage

Parties admissibles : ensembles de pays connexes vis-à-vis du graphe de voisinage

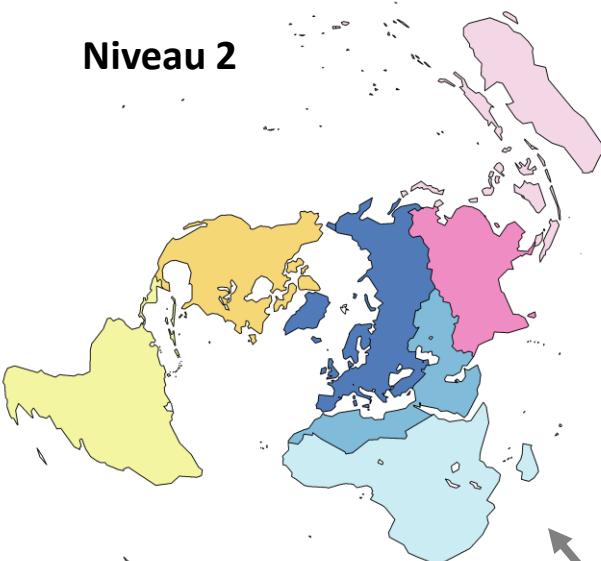


Exemples de parties admissibles

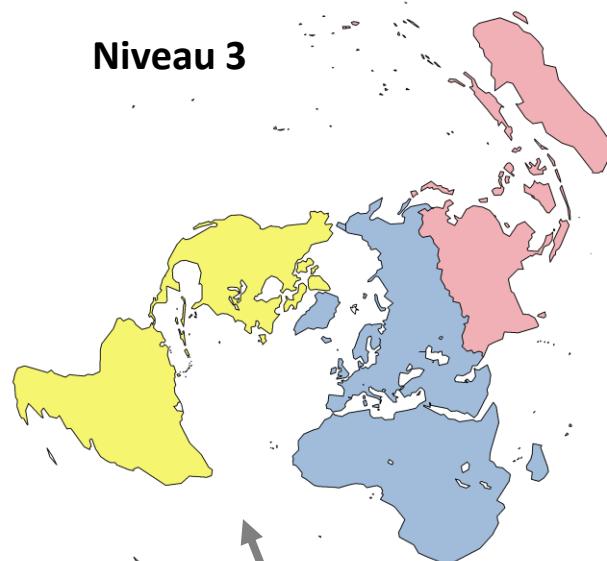
La hiérarchie WUTS

[Grasland et Didelon, 2007]

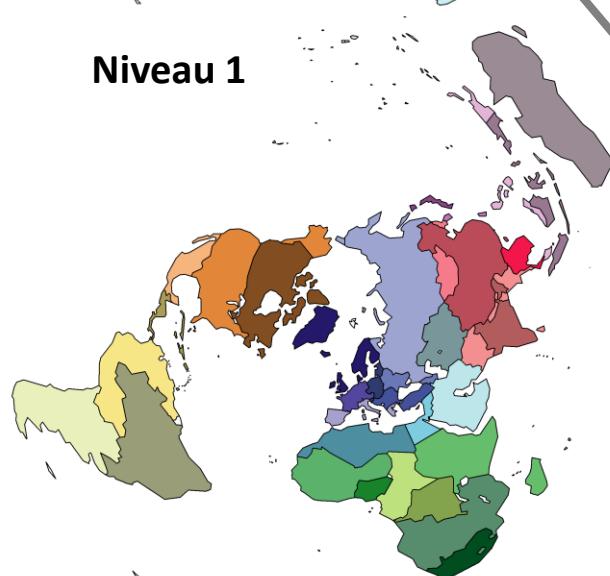
Niveau 2



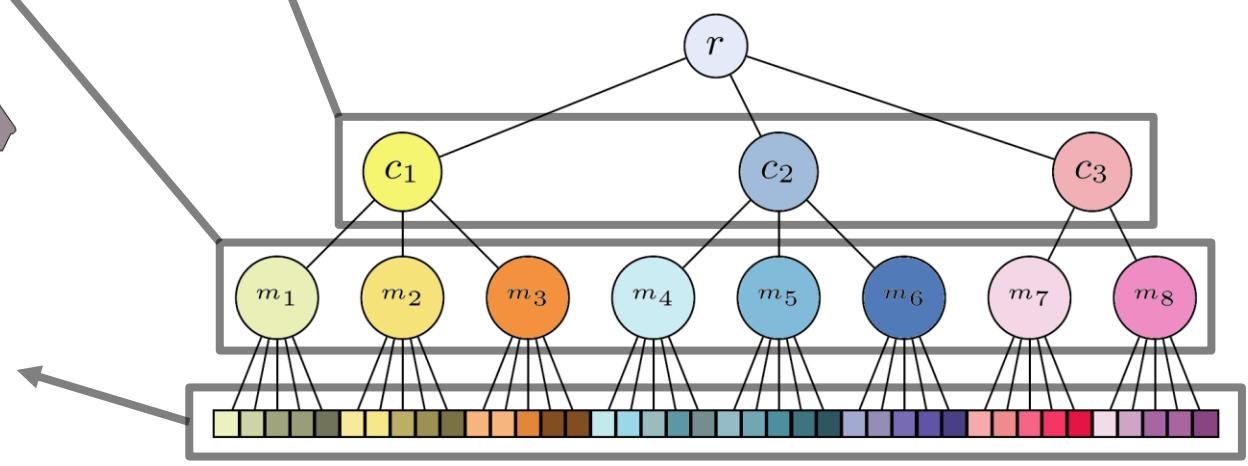
Niveau 3



Niveau 1



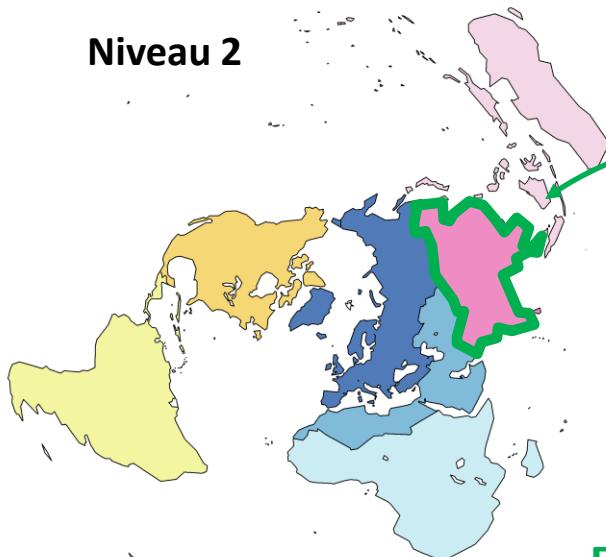
Parties admissibles :
ensembles de pays proches
sur le plan politique,
culturel, économique, etc.



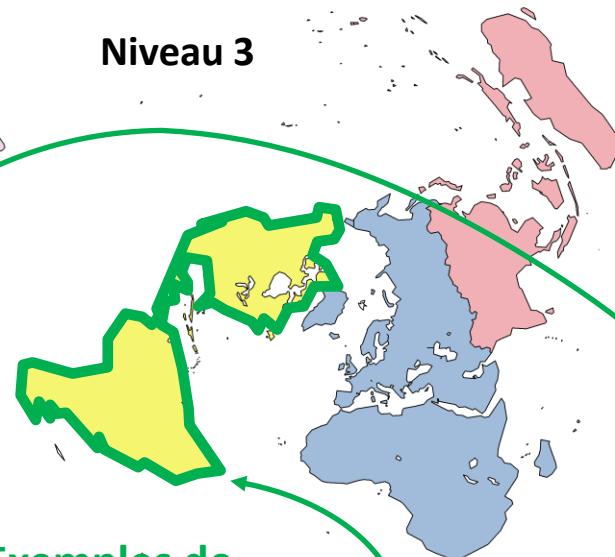
La hiérarchie WUTS

[Grasland et Didelon, 2007]

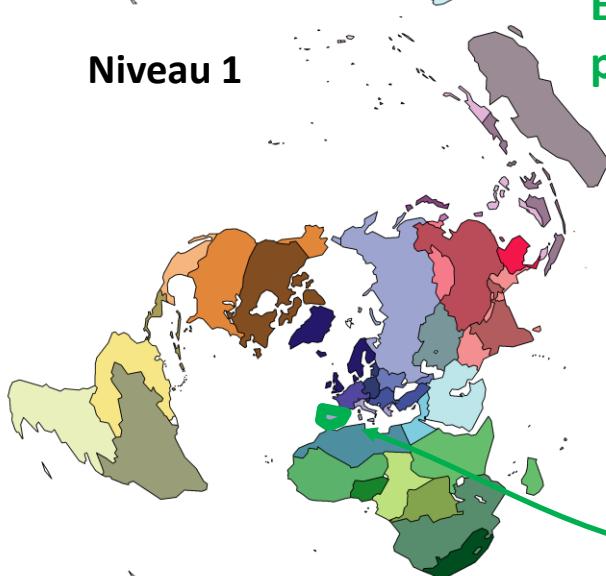
Niveau 2



Niveau 3

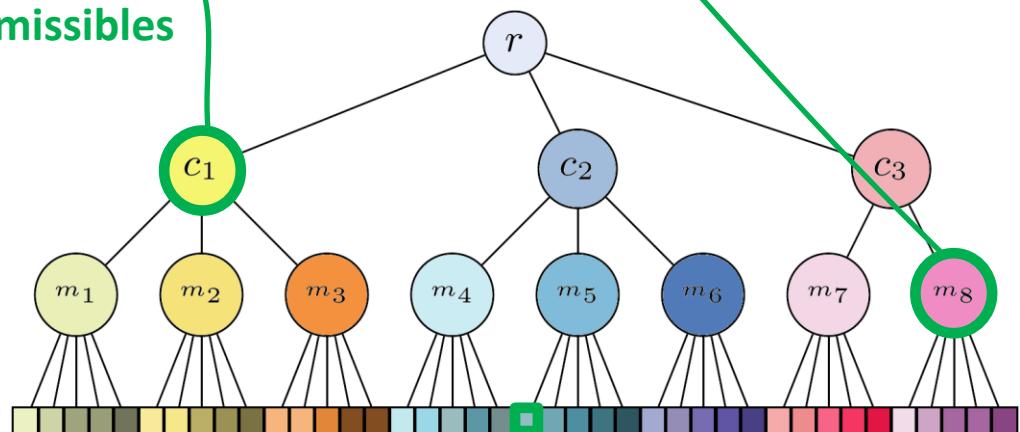


Niveau 1



Exemples de
parties admissibles

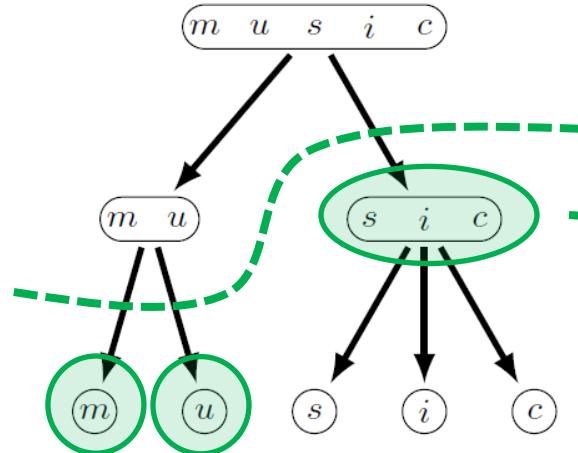
Parties admissibles :
ensembles de pays proches
sur le plan politique,
culturel, économique, etc.



Agrégation selon une hiérarchie

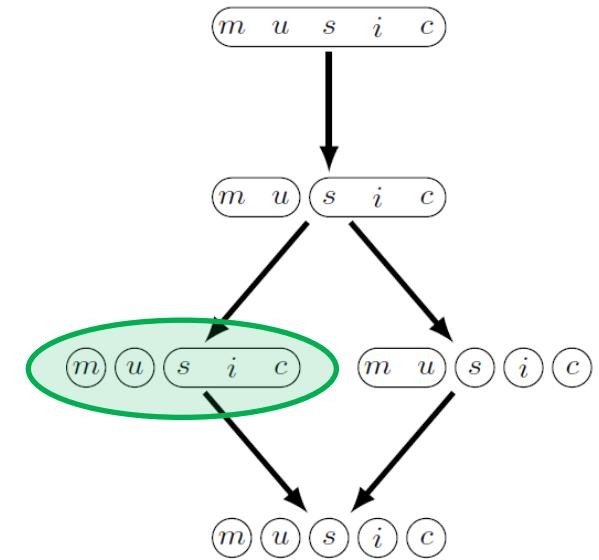
Parties admissibles

(nœuds de la hiérarchie)



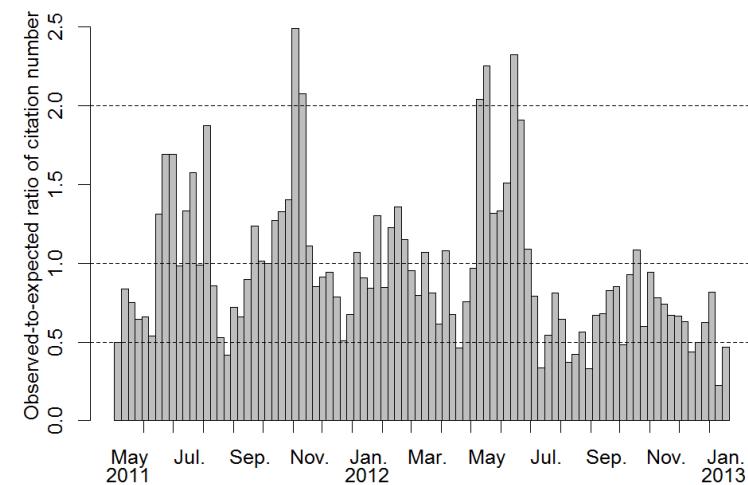
Partitions admissibles

(coupe dans la hiérarchie)

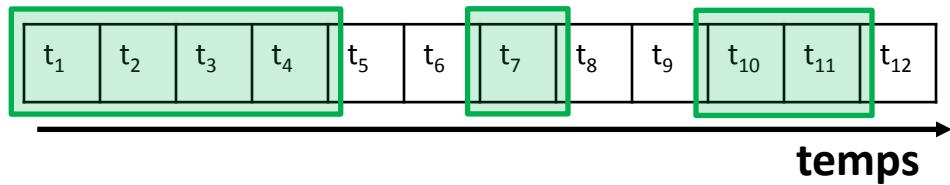


Conserver l'ordre du temps

Parties admissibles :
intervalles de temps



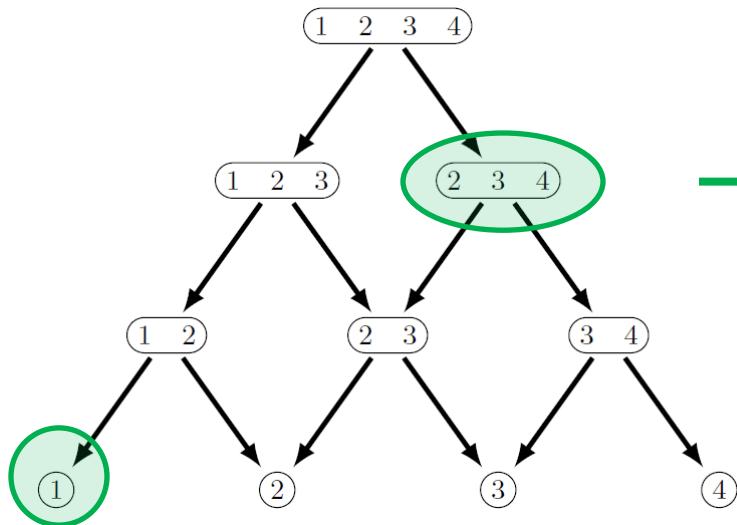
Exemples de parties admissibles



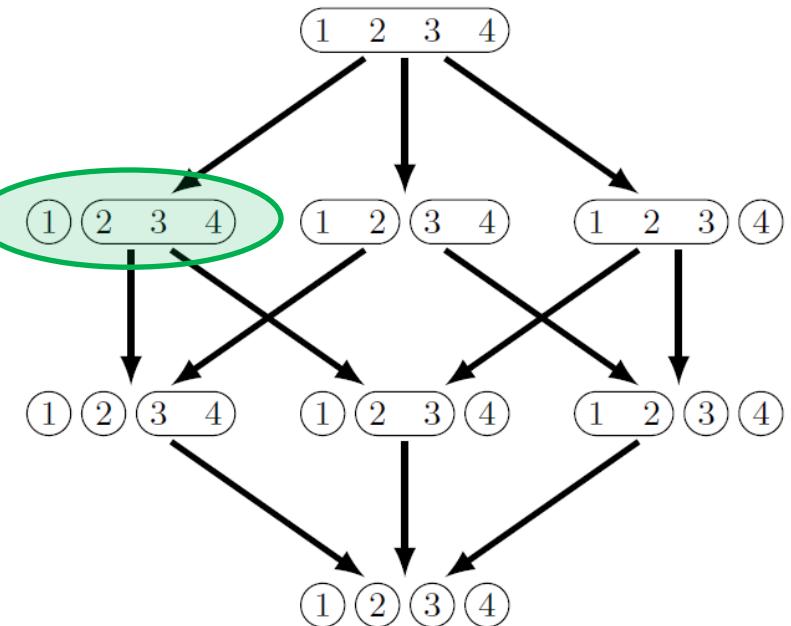
Agrégation selon un ordre total

Parties admissibles

(intervalles de temps)

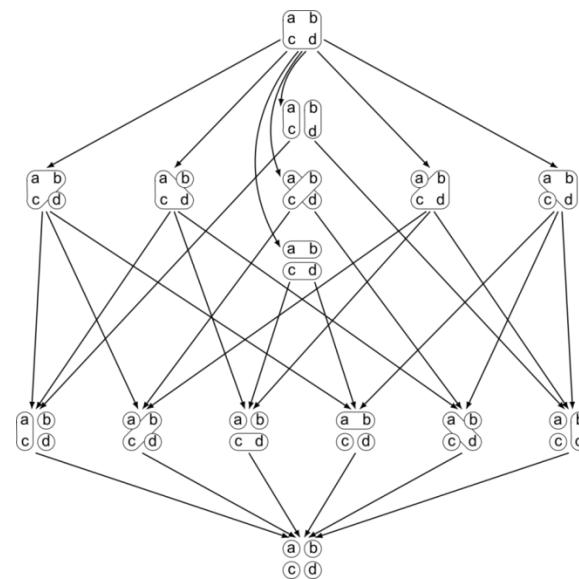


Partitions admissibles (séquences d'intervalles)

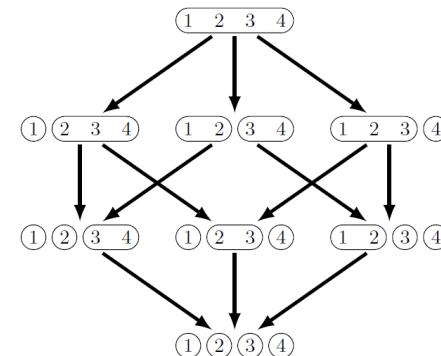


Complexité des structures algébriques

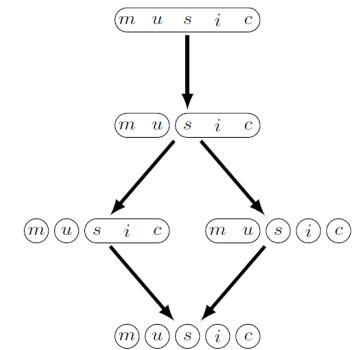
Partitions
non-constraintes



Partitions admissibles
selon un **ordre total**



Partitions admissibles
selon une **hiérarchie**



Moins constraint
Plus complexe

Plus constraint
Moins complexe

Mon approche

P0 Caractériser le processus d'agrégation

→ Algèbre des partitions possibles

P1 Conserver la sémantique du système

→ Contraindre le partitionnement

Agréger selon plusieurs dimensions

→ Des contraintes adaptées aux topologies du système

P2 Évaluer et comparer les représentations

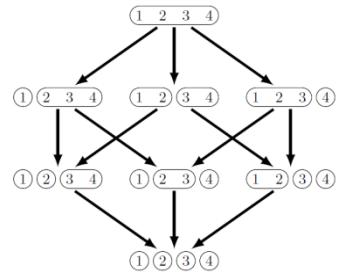
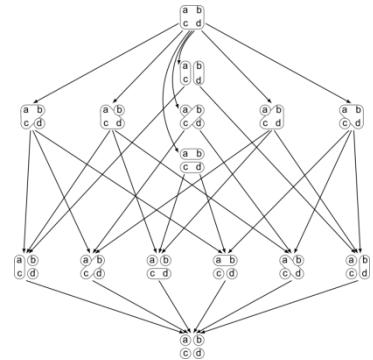
→ Mesurer la complexité et l'information

Proposer plusieurs niveau de granularité

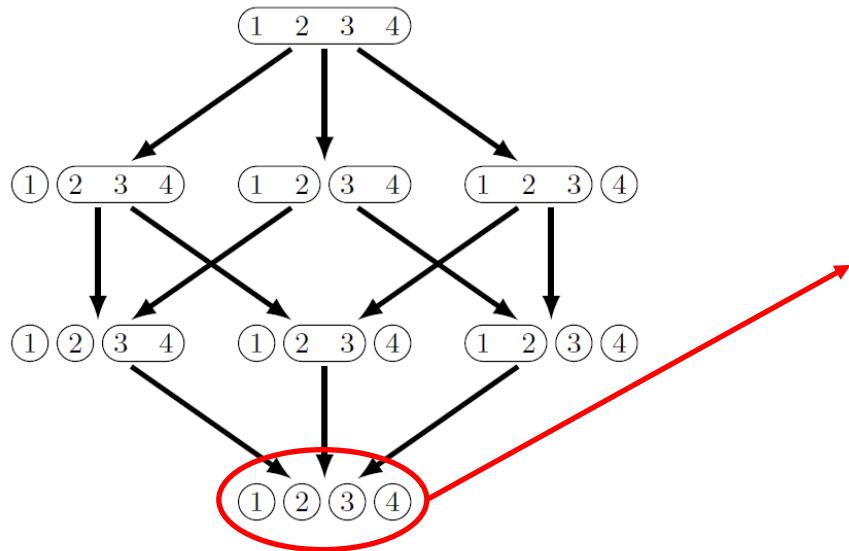
→ Optimiser un compromis de qualité

P3 Calculer les meilleures représentations

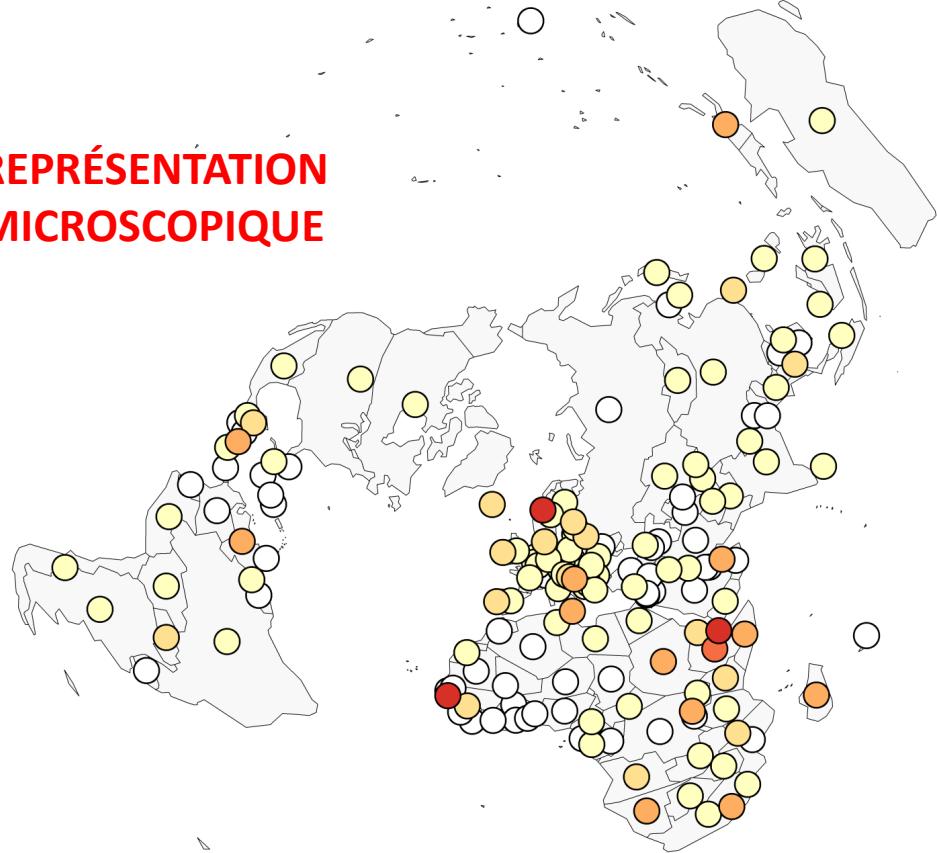
→ Algorithme générique d'optimisation sous contraintes



Objectifs et difficultés



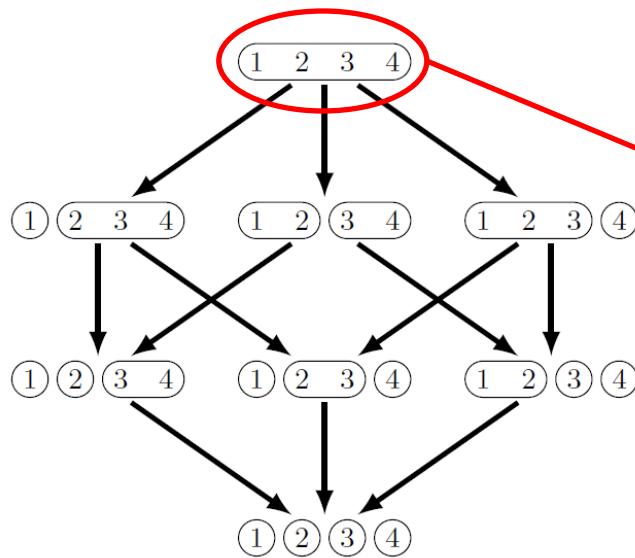
**REPRÉSENTATION
MICROSCOPIQUE**



**Quelle partition admissible
est la meilleure pour un jeu
de données particulier ?**

→ TROP COMPLEXE POUR
PASSER À L'ÉCHELLE

Objectifs et difficultés



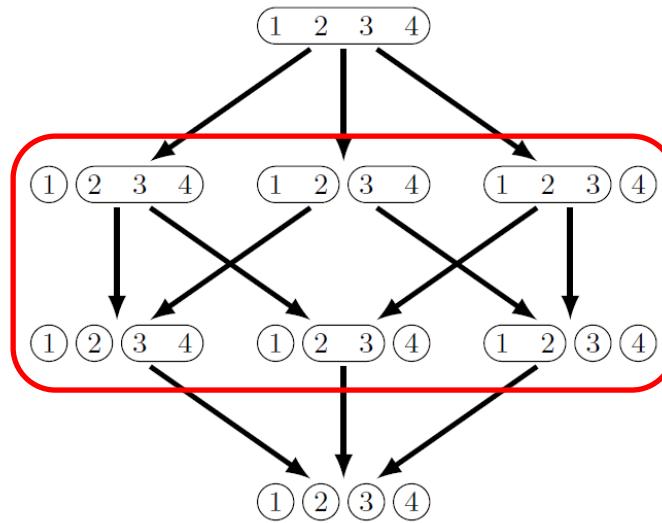
REPRÉSENTATION
ENTIÈREMENT AGRÉGÉE



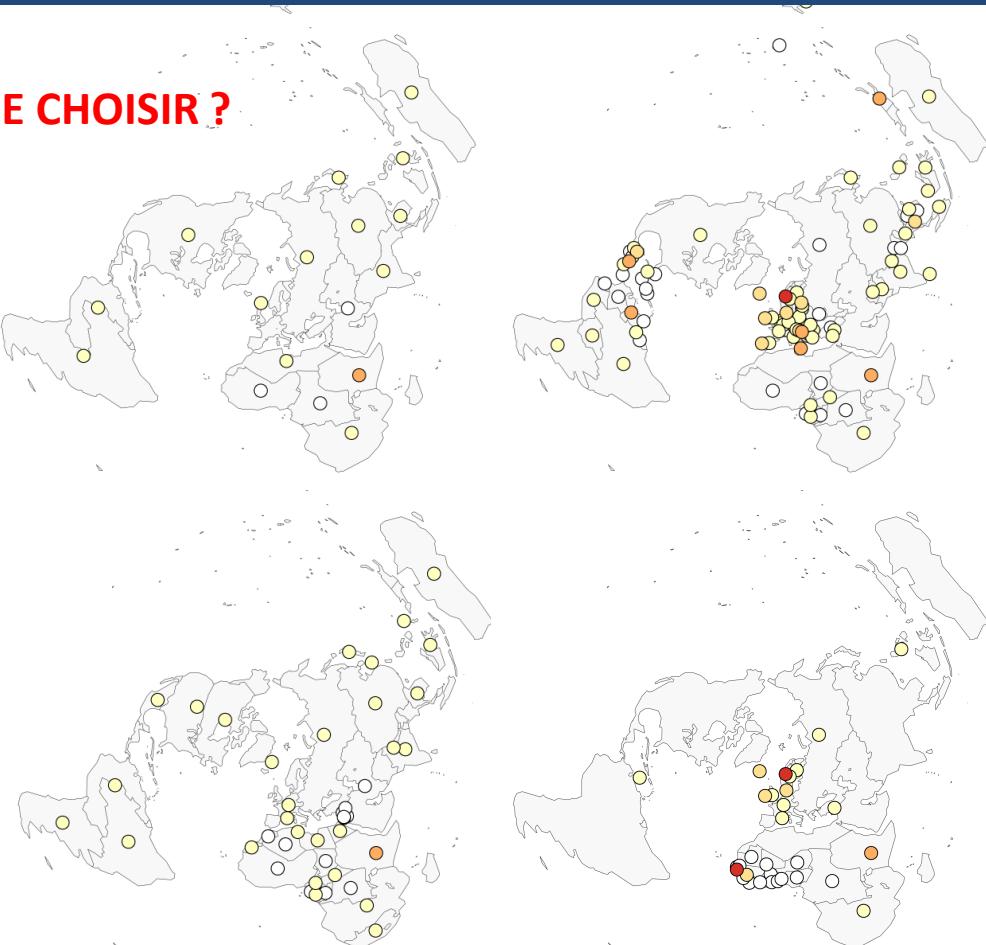
Quelle partition admissible
est la meilleure pour un jeu
de données particulier ?

→ NE DONNE QUE TRÈS
PEU D'INFORMATION

Objectifs et difficultés

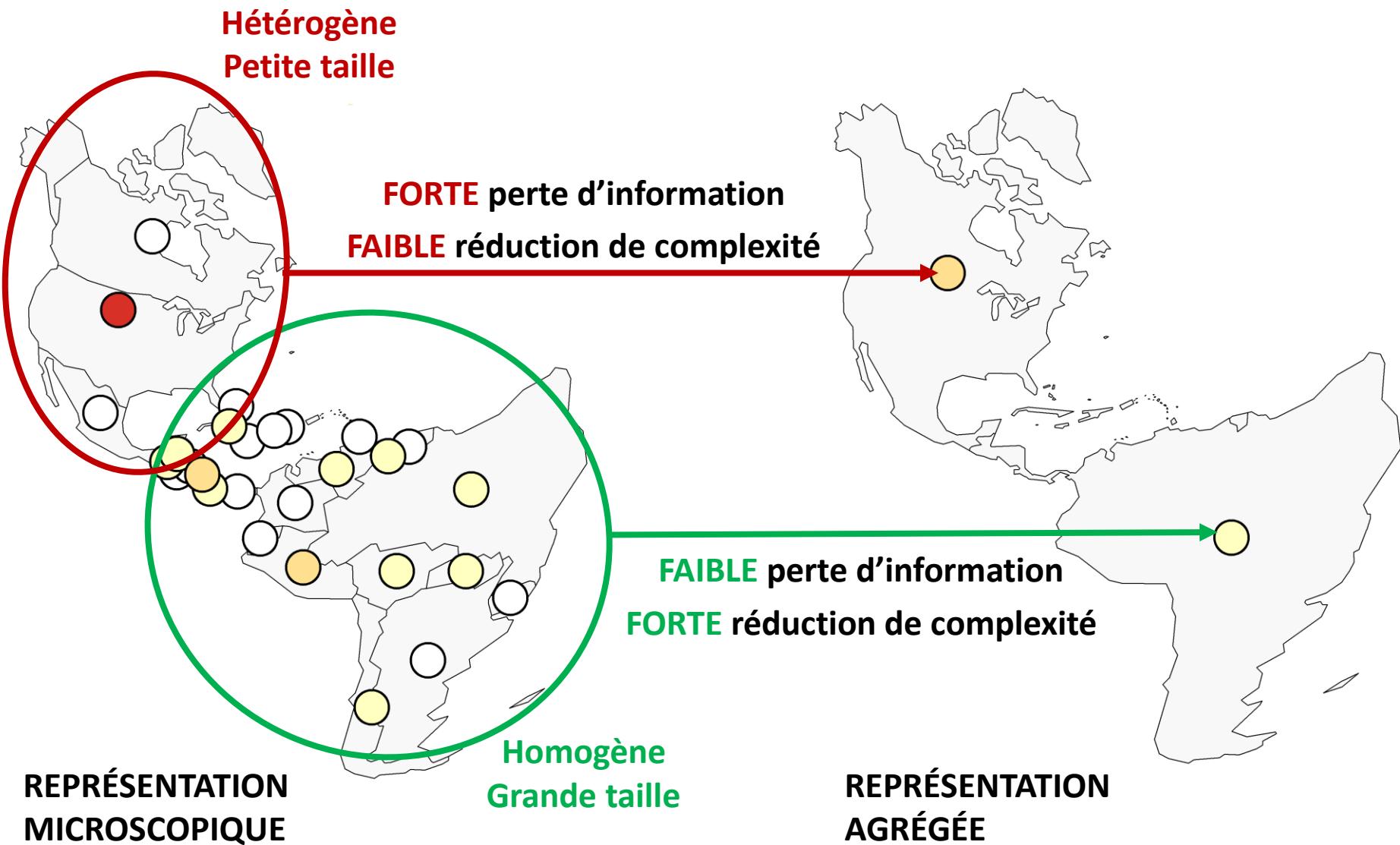


QUE CHOISIR ?

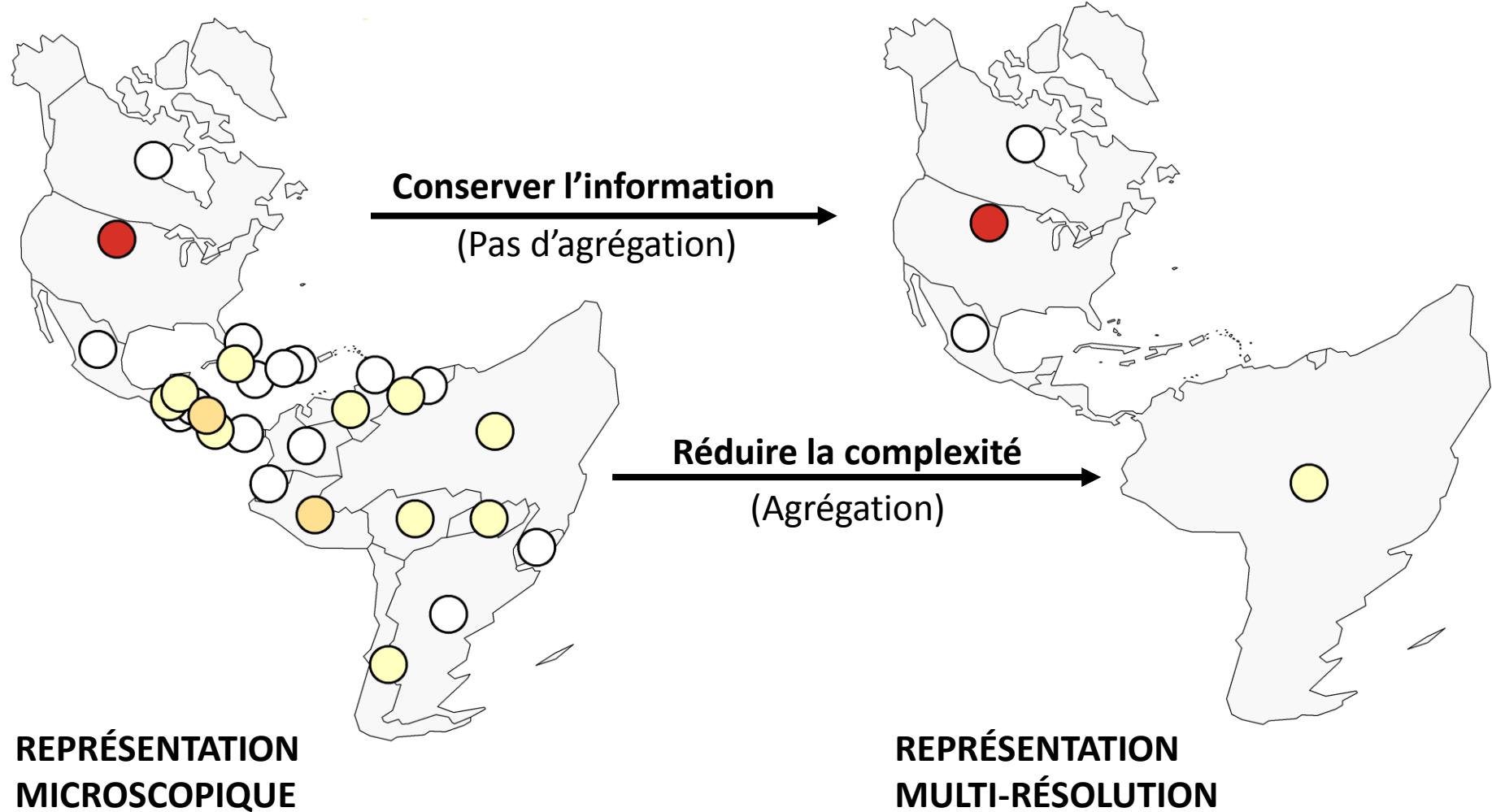


Quelle partition admissible
est la meilleure pour un jeu
de données particulier ?

Complexité et information



Complexité et information



Mesures de qualité

[Lamarche-Perrin et al., ECCS 2012]

La **complexité** dépend de la **tâche à accomplir** et des **outils de description** disponibles

[Bonabeau et Dessalles, 1997]

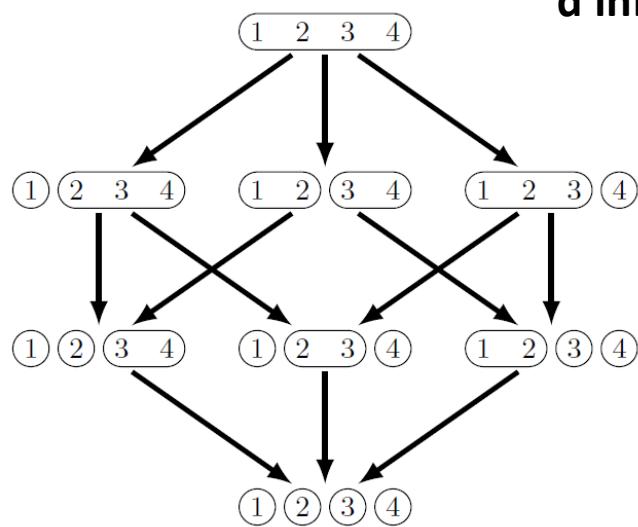
La **perte d'information** est mesurée par la **divergence** entre deux distributions de probabilité

[Kullback et Leibler, 1951]

Nombre d'agrégats représentés :

$$T(\mathcal{X}) = |\mathcal{X}|$$

Complexité



Perte
d'information

Divergence de Kullback-Leibler :

$$D(\mathcal{X}) = \sum_{X \in \mathcal{X}} \sum_{x \in X} \frac{v(x)}{v(\Omega)} \log_2 \left(\frac{v(x) |X|}{v(X)} \right)$$

Décomposabilité des mesures

Nombre d'agrégats représentés :

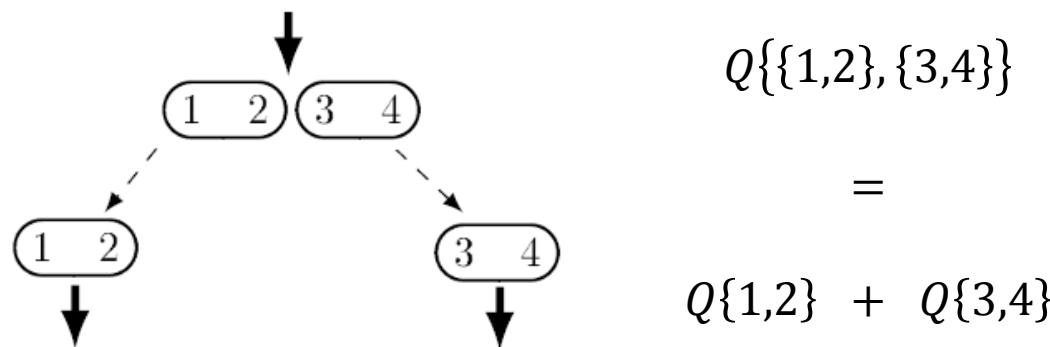
$$T(\mathcal{X}) = |\mathcal{X}|$$

Divergence de Kullback-Leibler :

$$D(\mathcal{X}) = \sum_{X \in \mathcal{X}} \sum_{x \in X} \frac{v(x)}{v(\Omega)} \log_2 \left(\frac{v(x) |X|}{v(X)} \right)$$

Décomposabilité additive : La qualité d'une partition est la somme des qualités de ses parties

[Jackson *et al.*, 2005] [Csiszár, 2008]

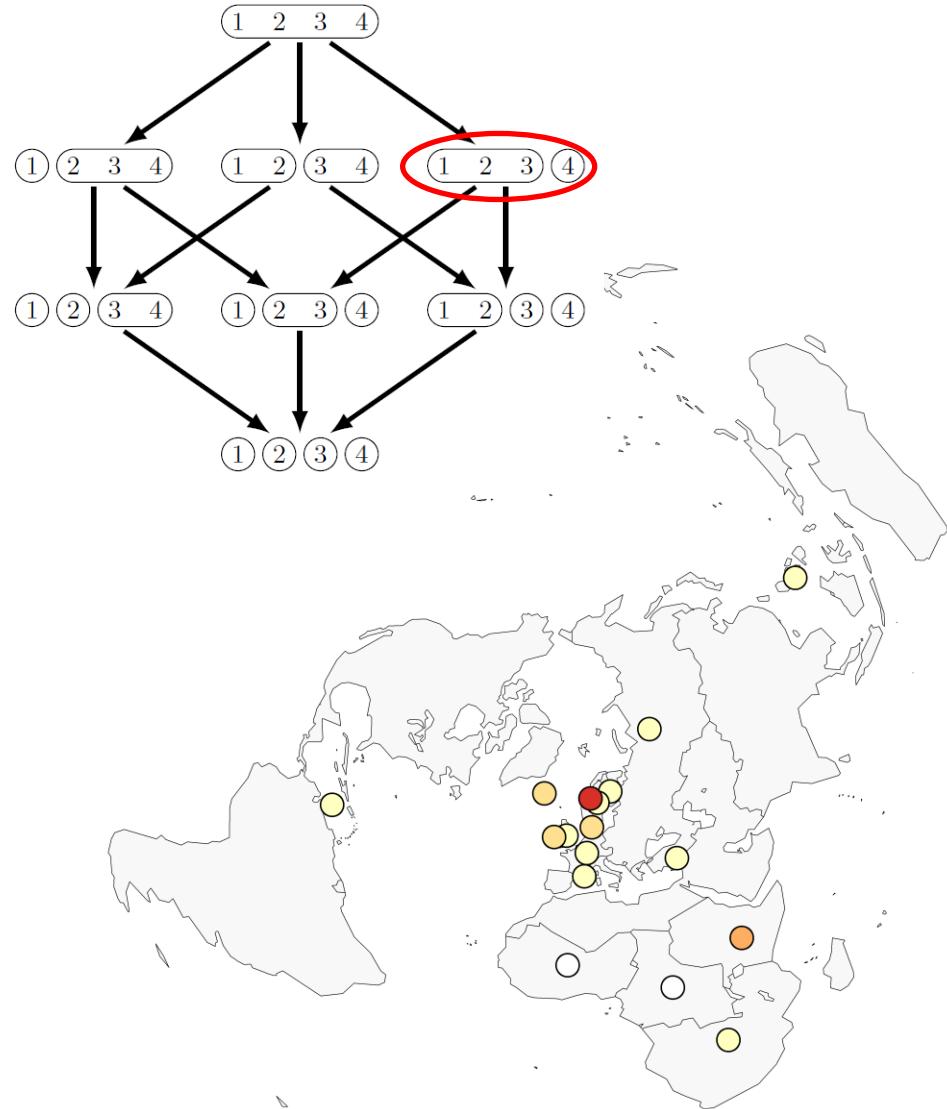
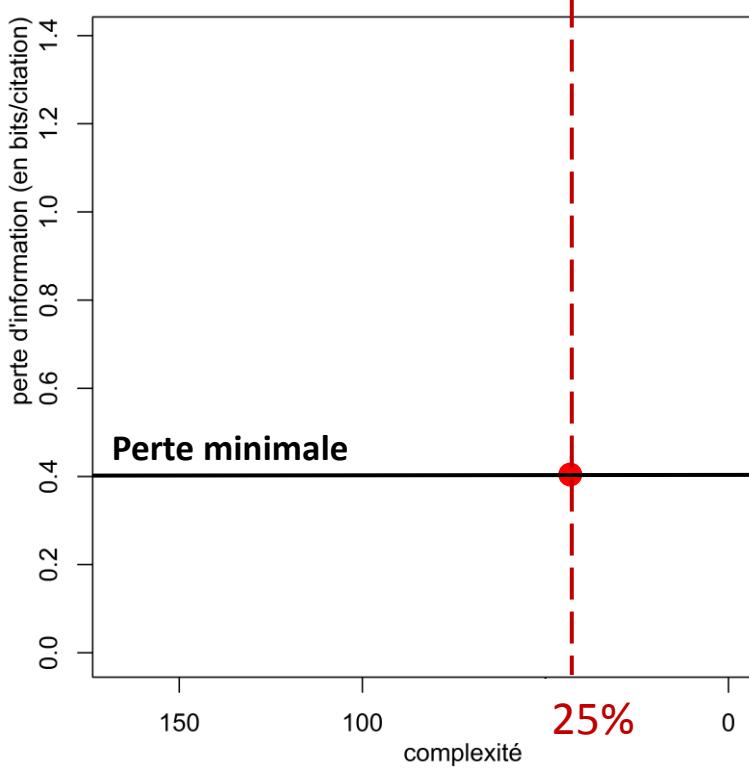


Optimisation des mesures de qualité

Deux critères d'évaluation indépendants

Compromis de qualité :

$$CQL_\alpha = \alpha \frac{\Delta T}{\Delta T_{\max}} - (1 - \alpha) \frac{D}{D_{\max}}$$

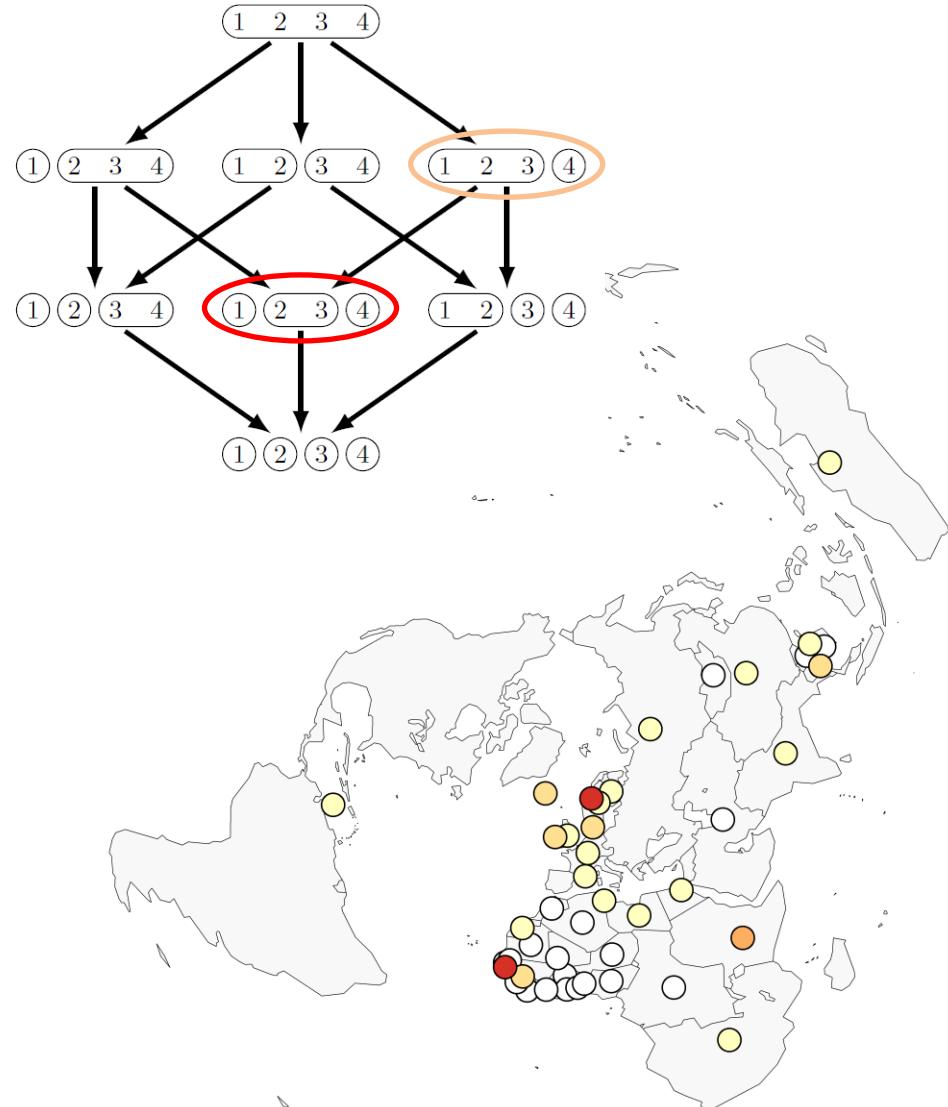
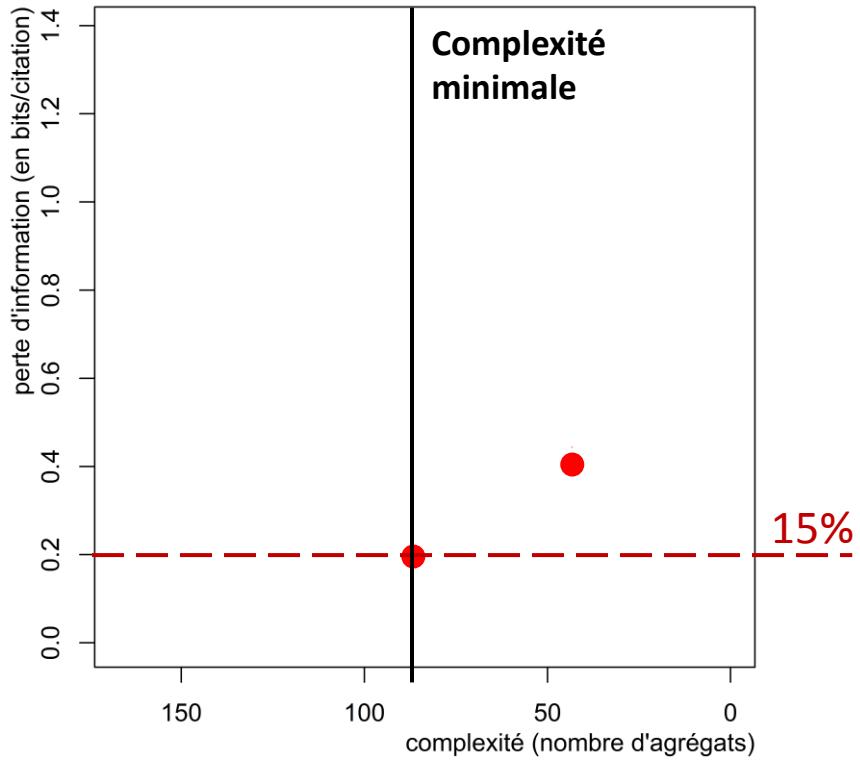


Optimisation des mesures de qualité

Deux critères d'évaluation indépendants

Compromis de qualité :

$$CQL_\alpha = \alpha \frac{\Delta T}{\Delta T_{\max}} - (1 - \alpha) \frac{D}{D_{\max}}$$

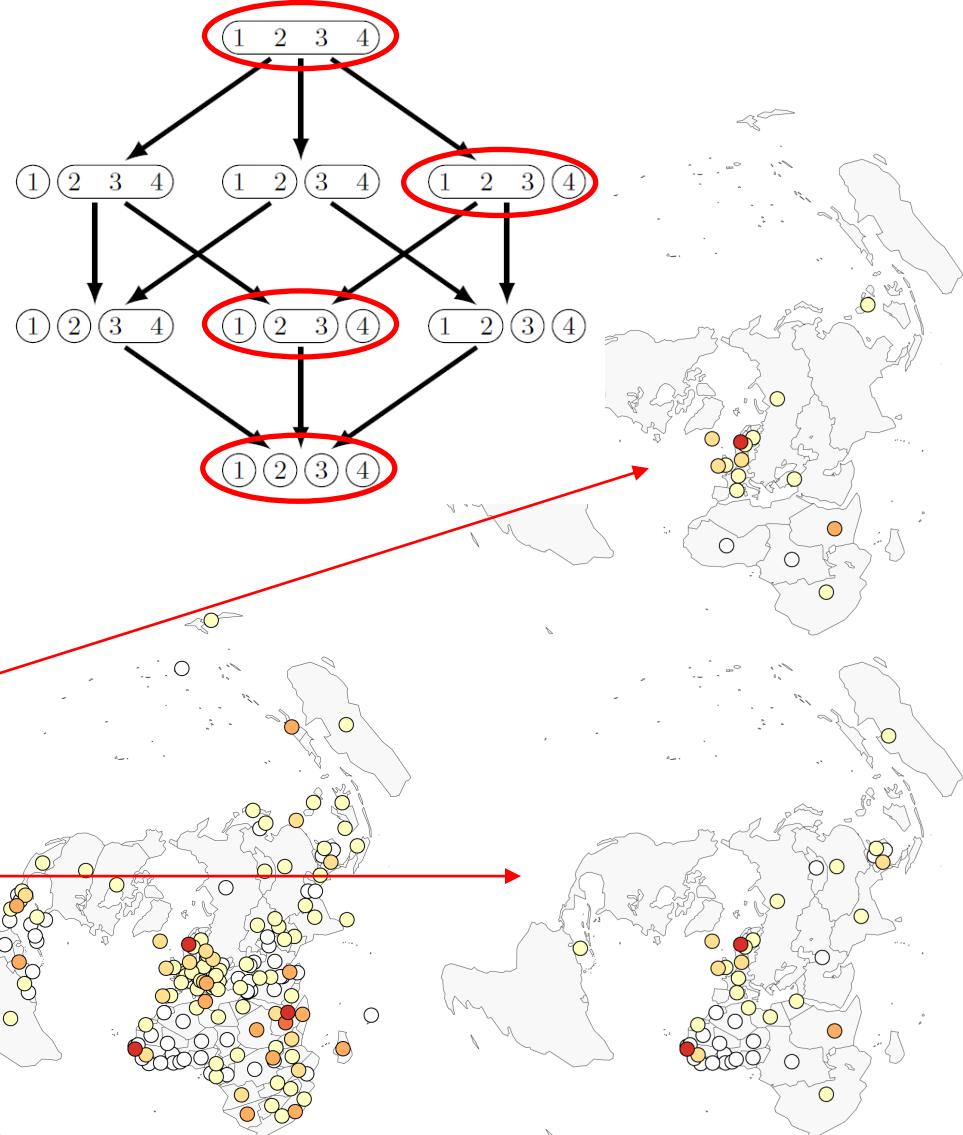
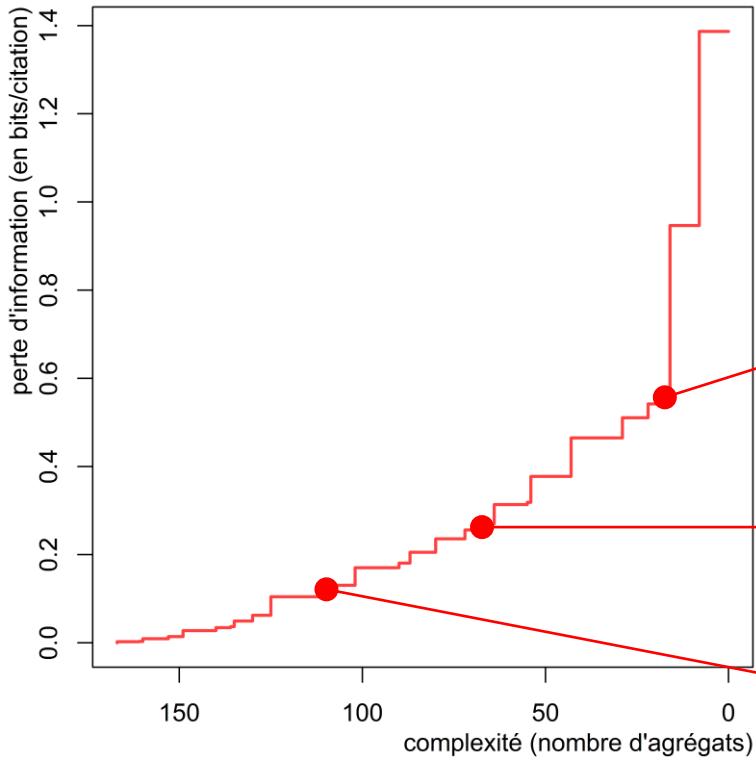


Optimisation des mesures de qualité

Deux critères d'évaluation indépendants

Compromis de qualité :

$$CQL_\alpha = \alpha \frac{\Delta T}{\Delta T_{\max}} - (1 - \alpha) \frac{D}{D_{\max}}$$



Mon approche

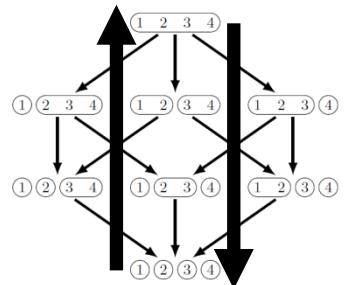
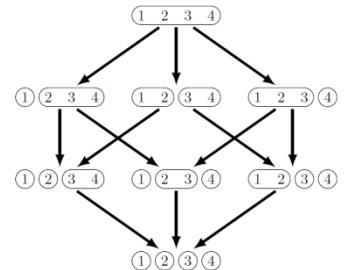
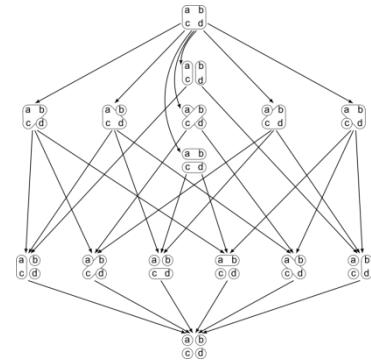
P0 Caractériser le processus d'agrégation
→ Algèbre des partitions possibles

P1 Conserver la sémantique du système
→ Contraindre le partitionnement

Agréger selon plusieurs dimensions
→ Des contraintes adaptées aux topologies du système

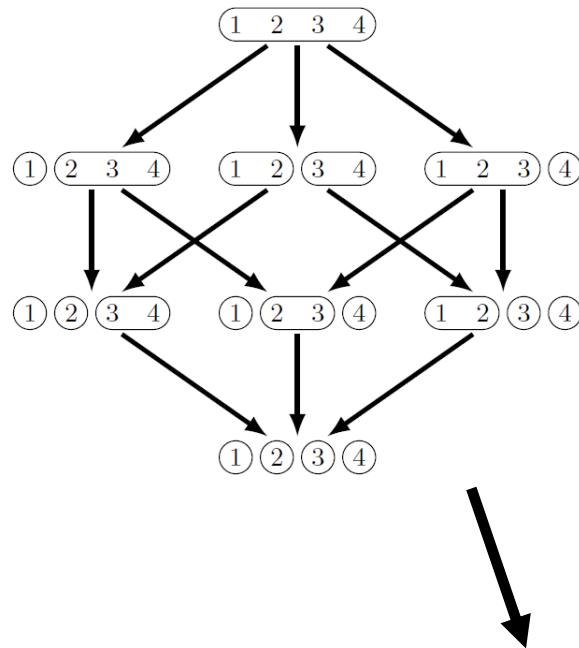
P2 Évaluer et comparer les représentations
→ Mesurer la complexité et l'information
Proposer plusieurs niveau de granularité
→ Optimiser un compromis de qualité

P3 Calculer les meilleures représentations
→ Algorithme générique d'optimisation sous contraintes



Objectifs

Contraintes d'admissibilité



Mesures de qualité

$$\Delta T(\mathcal{X}) = |\Omega| - |\mathcal{X}|$$

$$D(\mathcal{X}) = \sum_{X \in \mathcal{X}} \sum_{x \in X} \frac{v(x)}{v(\Omega)} \log_2 \left(\frac{v(x) |X|}{v(X)} \right)$$

$$CQL_\alpha = \alpha \frac{\Delta T}{\Delta T_{\max}} - (1 - \alpha) \frac{D}{D_{\max}}$$

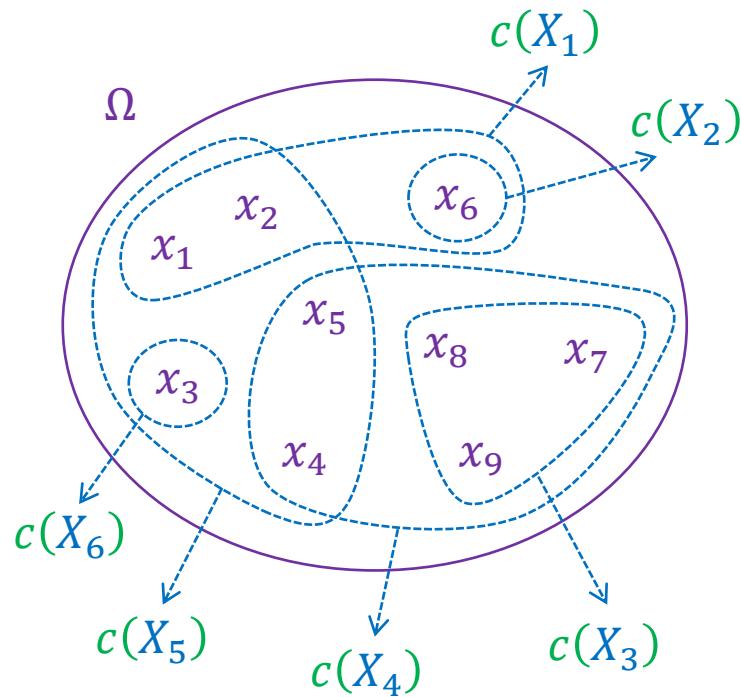


Problème des partitions admissibles optimales : quelle est la partition admissible qui optimise un compromis de qualité donné ?
 → problème d'optimisation sous contraintes

The Set Partitioning Problem

Given:

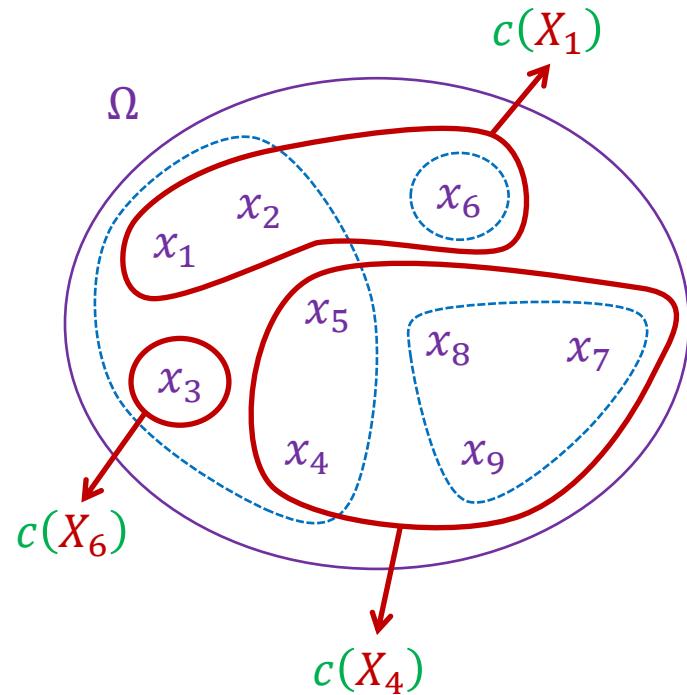
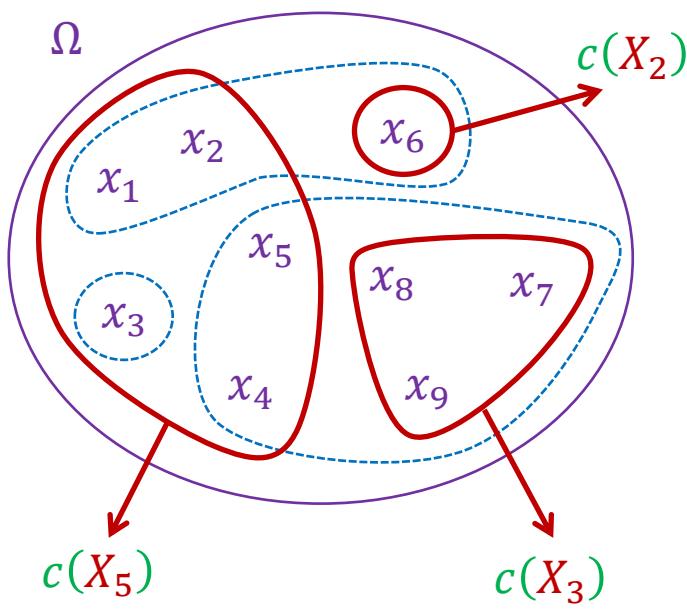
- a set of individuals $\Omega = \{x_1, \dots, x_n\}$
- a set of admissible parts $\mathcal{P} = \{X_1, \dots, X_m\} \subset 2^\Omega$
- a cost function $c : \mathcal{P} \rightarrow \mathbb{R}$



The Set Partitioning Problem

Given:

- a set of individuals $\Omega = \{x_1, \dots, x_n\}$
- a set of admissible parts $\mathcal{P} = \{X_1, \dots, X_m\} \subset 2^\Omega$
- a cost function $c : \mathcal{P} \rightarrow \mathbb{R}$
- the corresponding set of admissible partitions $\mathfrak{P} = \{\mathcal{X} \subset \mathcal{P} \text{ such that } \mathcal{X} \text{ is a partition of } \Omega\}$

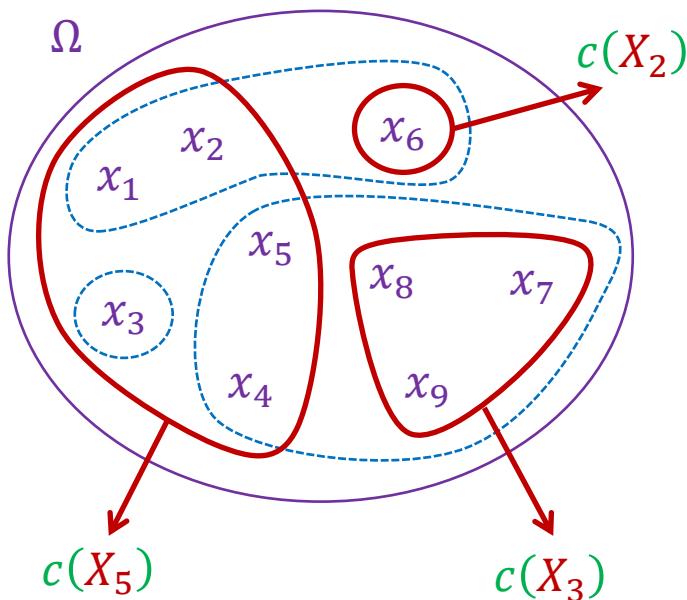


The Set Partitioning Problem

Given:

- a set of individuals $\Omega = \{x_1, \dots, x_n\}$
- a set of admissible parts $\mathcal{P} = \{X_1, \dots, X_m\} \subset 2^\Omega$
- a cost function $c : \mathcal{P} \rightarrow \mathbb{R}$
- the corresponding set of admissible partitions $\mathfrak{P} = \{\mathcal{X} \subset \mathcal{P} \text{ such that } \mathcal{X} \text{ is a partition of } \Omega\}$

Additional assumptions



Problem: Find an admissible partition that minimizes the cost function:

$$\mathcal{X}^* = \arg \min_{\mathcal{X} \in \mathfrak{P}} \left(\sum_{X \in \mathcal{X}} c(X) \right)$$

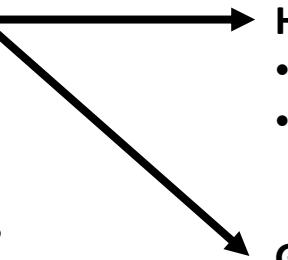
→ NP-complete!

Applications

Special Versions

Multilevel Geographical Analysis

- Ω = territorial units
- \mathcal{P} = admissible aggregates
- c = compression rate
- \mathfrak{P} = aggregated representations

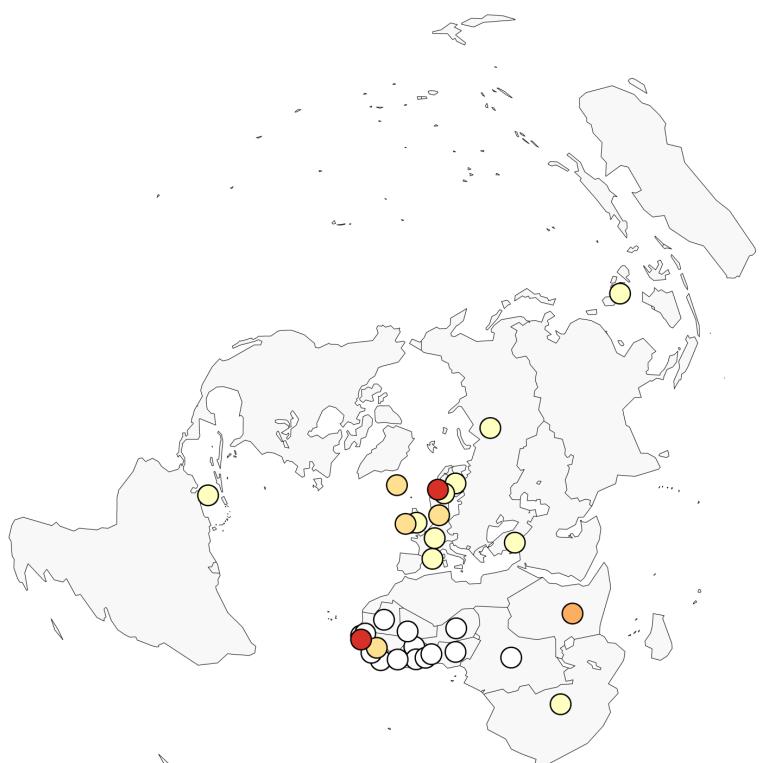
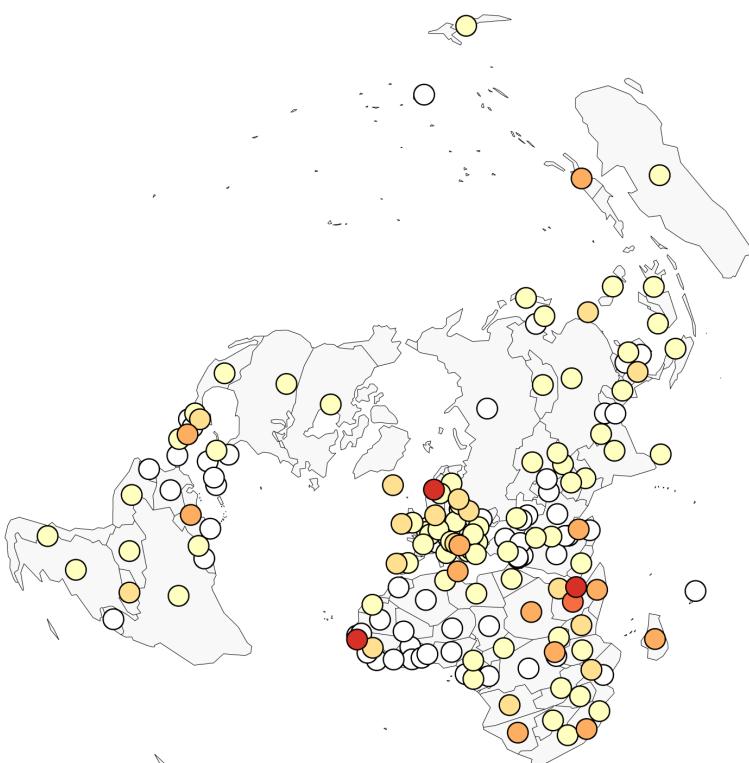


Hierarchical SPP

- Assumption: \mathcal{P} forms a hierarchy
- Result: $\mathcal{O}(n)$ depth-first search
[Pons *et al.*, 2011] [Lamarche-Perrin *et al.*, 2014]

Graph SPP

- Assumption: \mathcal{P} are connected parts of a graph
- Result: NP-complete [Becker *et al.*, 1998]



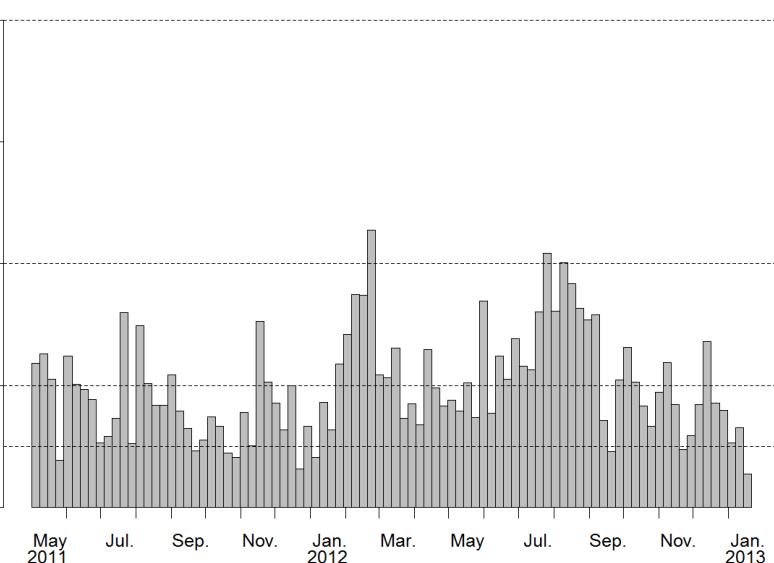
Applications

Special Versions

Multilevel Geographical Analysis

Time Series Analysis

- Ω = ordered data points
- \mathcal{P} = time intervals
- c = compression rate
- \mathfrak{P} = aggregated time series



Hierarchical SPP

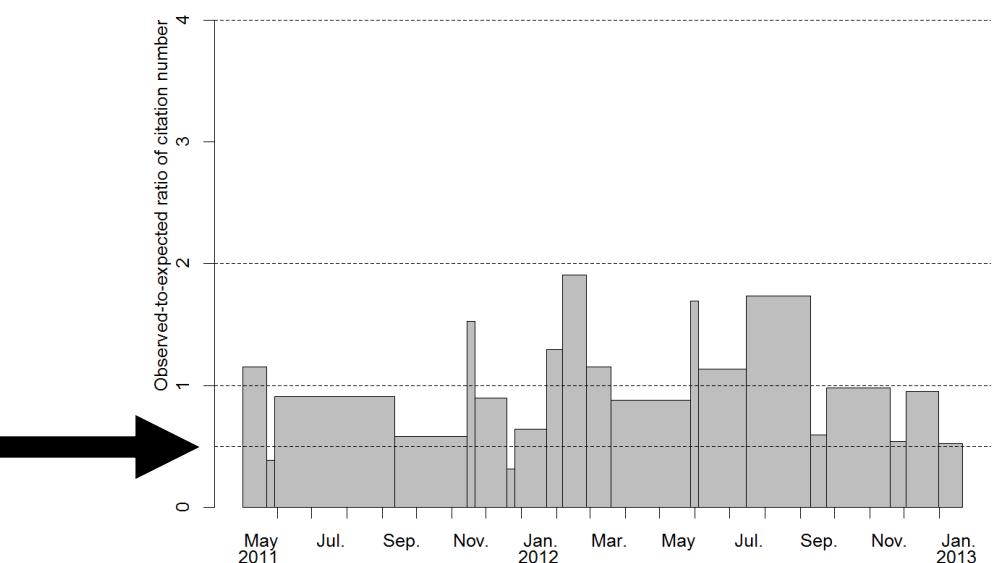
- Assumption: \mathcal{P} forms a hierarchy
- Result: $\mathcal{O}(n)$ depth-first search
[Pons *et al.*, 2011] [Lamarche-Perrin *et al.*, 2014]

Graph SPP

- Assumption: \mathcal{P} are connected parts of a graph
- Result: NP-complete [Becker *et al.*, 1998]

Ordered SPP

- Assumption: \mathcal{P} are intervals
- Result: $\mathcal{O}(n^2)$ dynamic programming
[Anily *et al.*, 1991] [Jackson *et al.*, 2005]



Applications

Special Versions

Multilevel Geographical Analysis

Time Series Analysis

Coalition Structure Generation

Community Detection

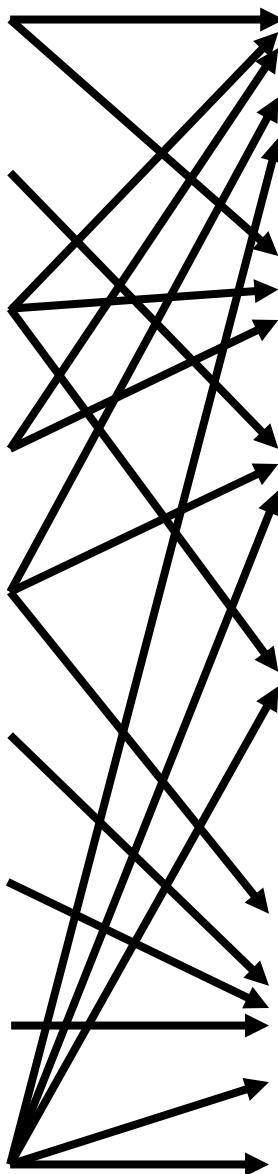
Distributed System Monitoring

Load Balancing Problem

Database Optimization

Image Processing

Combinatorial Auctions



Hierarchical SPP

- Assumption: \mathcal{P} forms a hierarchy
- Result: $\mathcal{O}(n)$ depth-first search
[Pons *et al.*, 2011] [Lamarche-Perrin *et al.*, 2014]

Graph SPP

- Assumption: \mathcal{P} are connected parts of a graph
- Result: NP-complete [Becker *et al.*, 1998]

Ordered SPP

- Assumption: \mathcal{P} are intervals
- Result: $\mathcal{O}(n^2)$ dynamic programming
[Anily *et al.*, 1991] [Jackson *et al.*, 2005]

Complete SPP

- Assumption: \mathcal{P} contains all parts
- Result: $\mathcal{O}(3^n)$ dynamic programming
[Yeh, 1986] [Lehmann *et al.*, 2006]

Ordered x Hierarchical SPP [Dosimont *et al.*, 2014]

Array SPP [Muthukrishnan *et al.*, 2005]

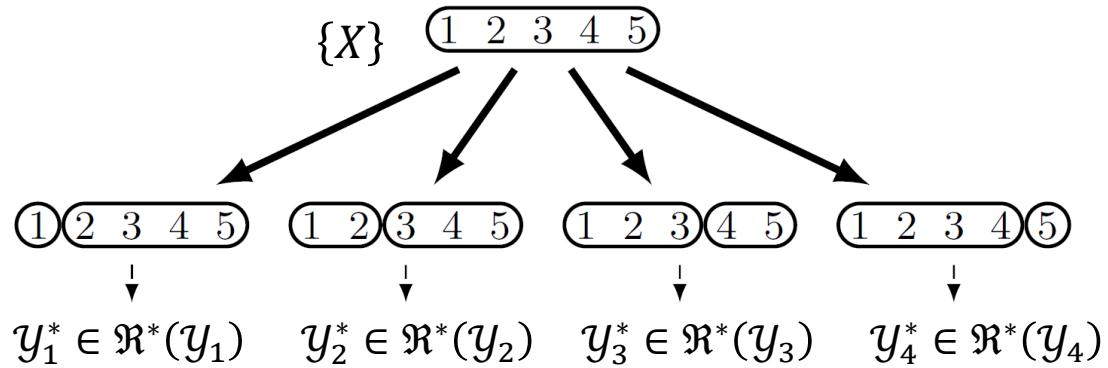
SPP with Size Bounds [Rothkopf *et al.*, 1998]

Cyclic SPP [Rothkopf *et al.*, 1998]

Divide and Conquer

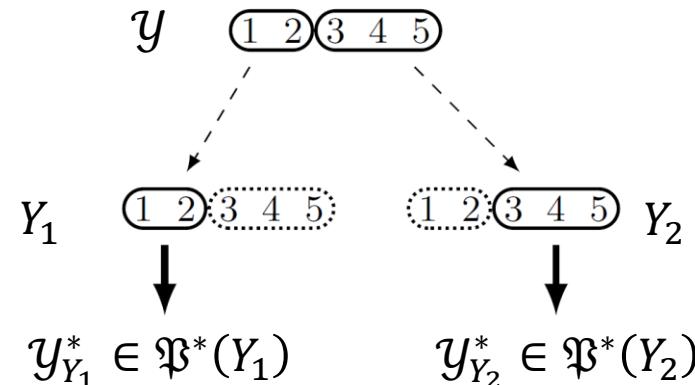
DIVIDE...

Branching the state space
according to the covering
relation



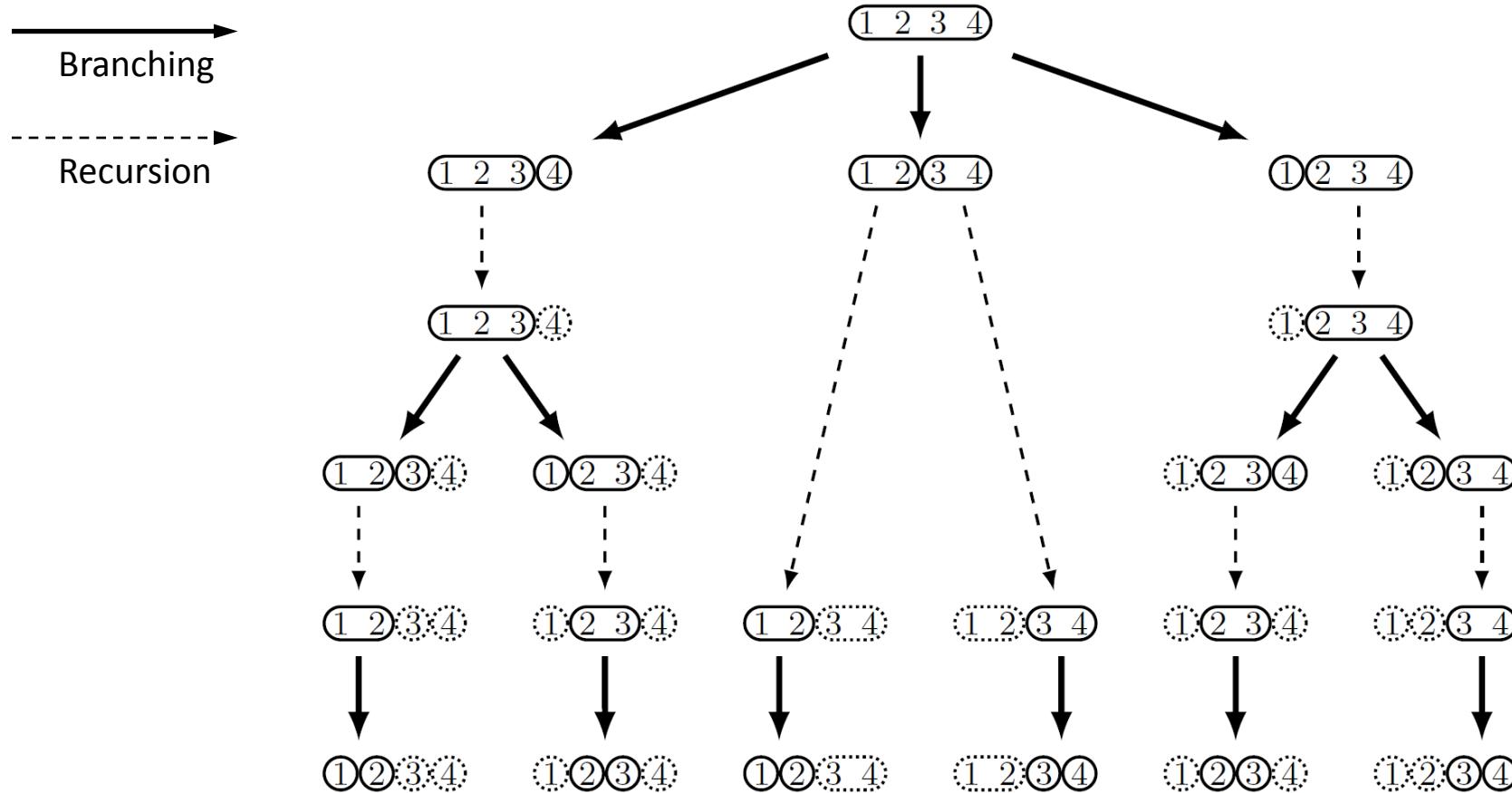
...AND CONQUER

Recursion according to the
principle of optimality

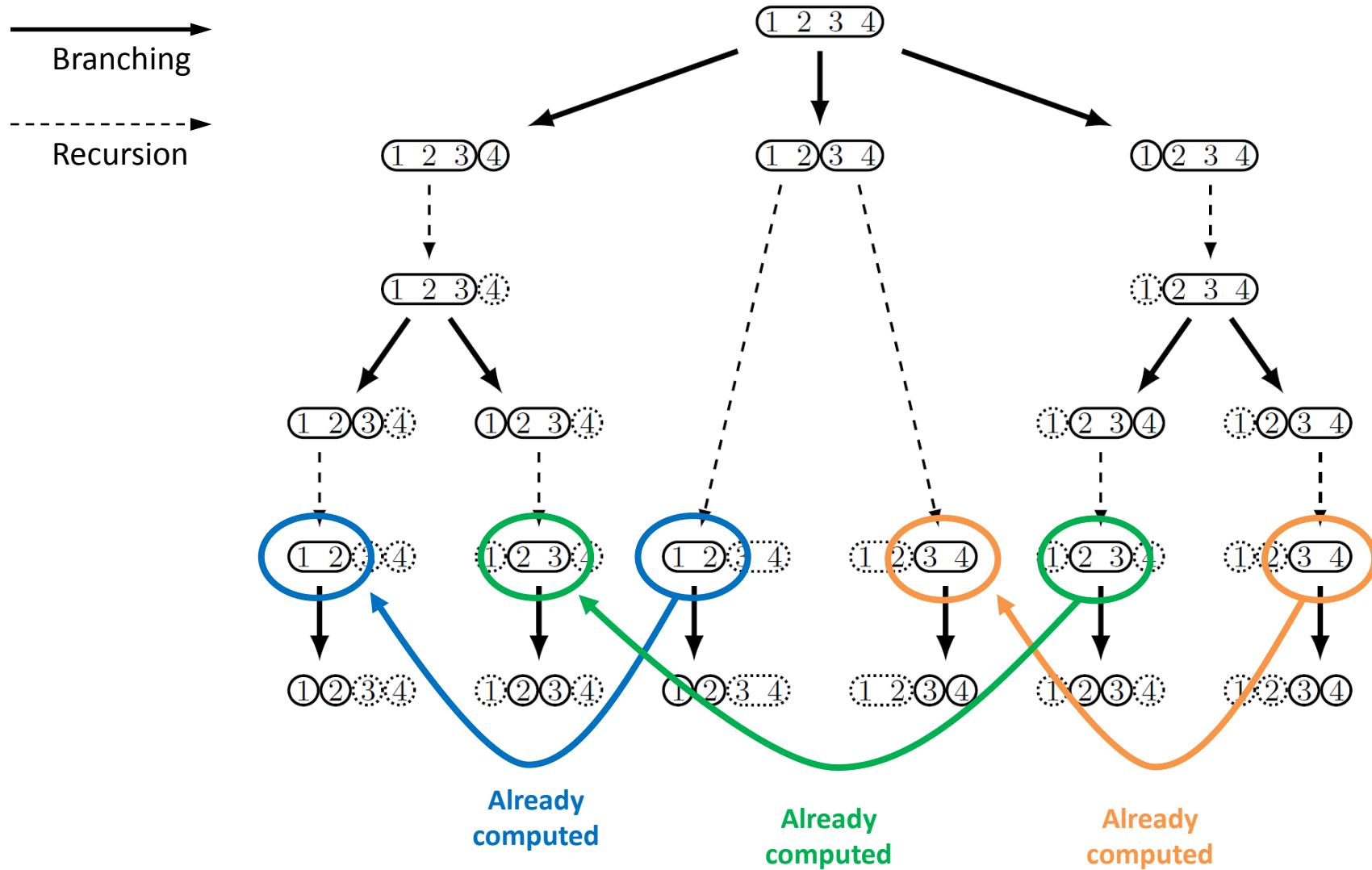


Execution of the Generic Algorithm

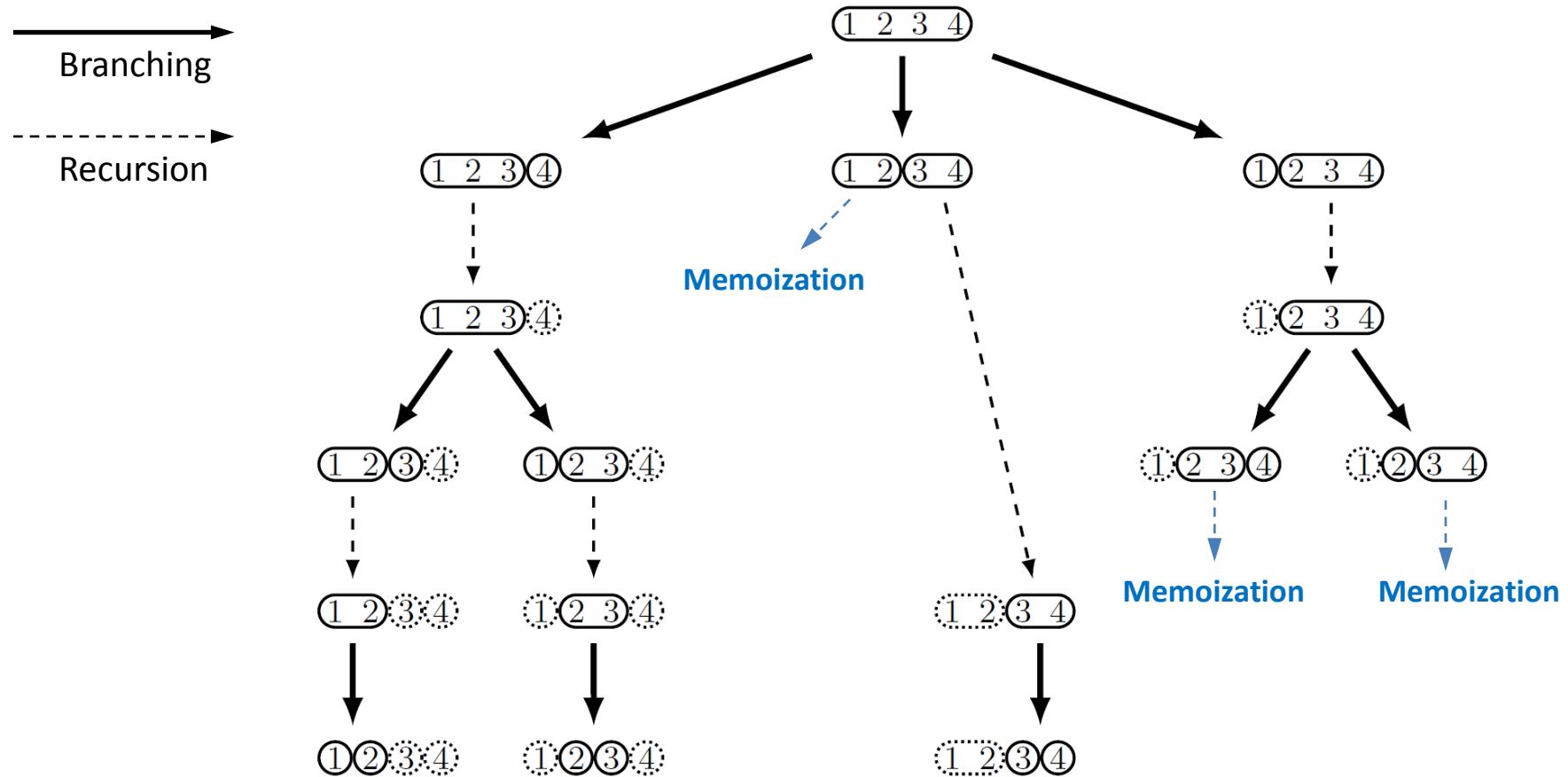
Ordered SPP on a Population of Size 4



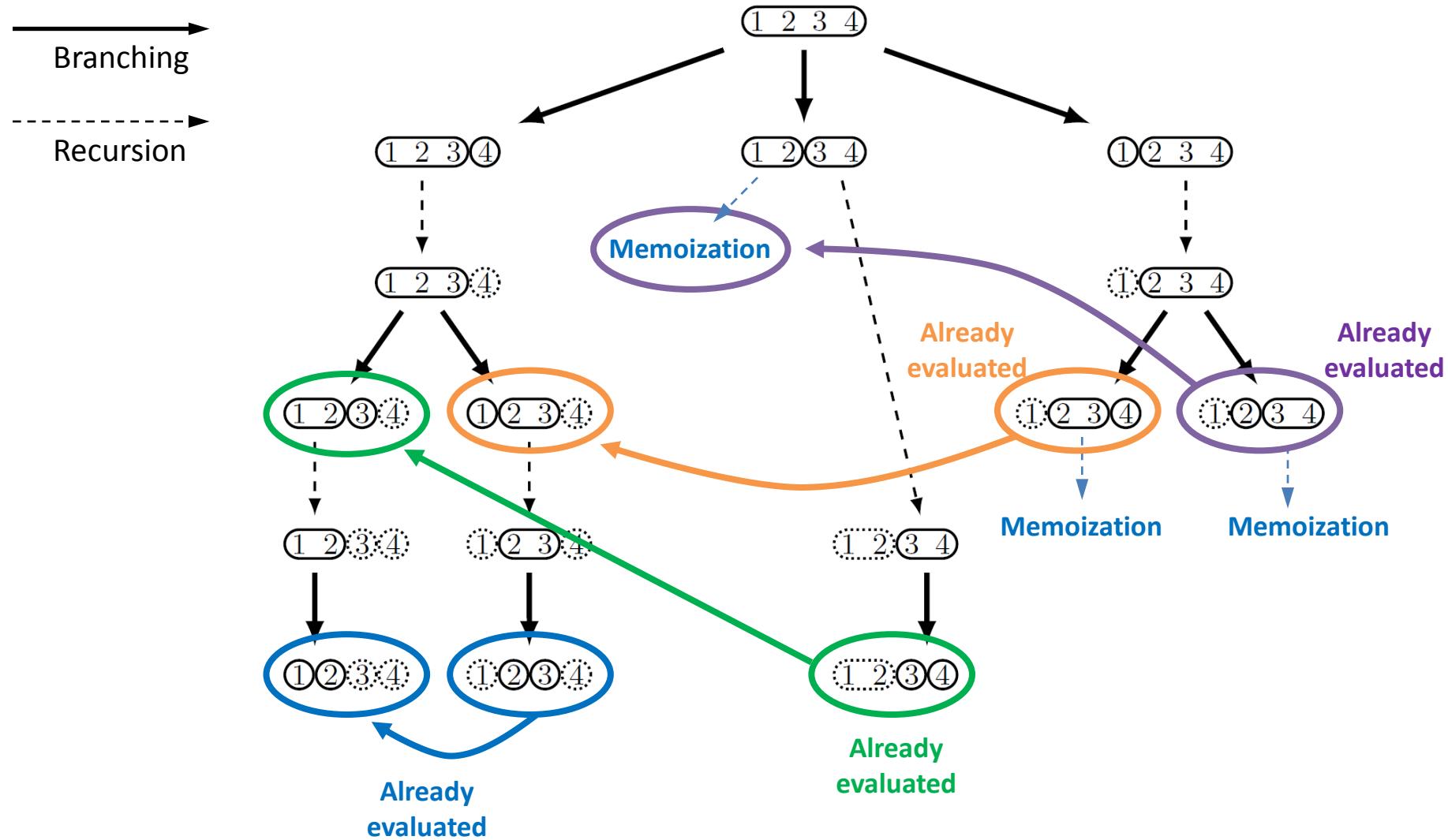
Memoization



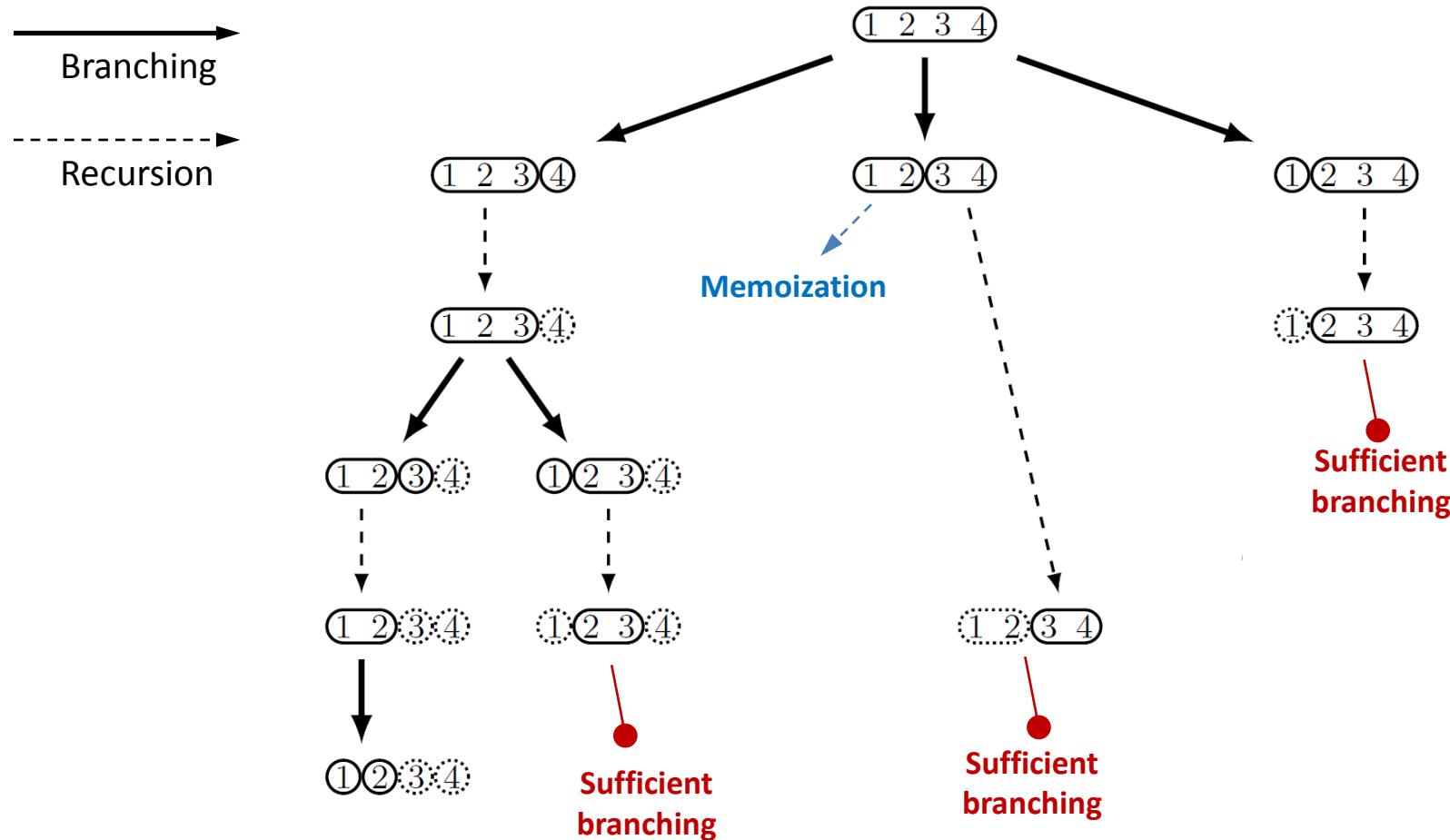
Memoization



Non-redundant Branching



Non-redundant Branching



The Generic Algorithm

A Generic Algorithm to Solve the SPP

Global Inputs:

- c a cost function;
- \mathcal{P} a set of admissible parts defining admissible partitions;
- \mathcal{L} a set of locally-optimal admissible partitions of parts on which the algorithm has already been applied.

Local Inputs:

- X an admissible part;
- $\bar{\mathcal{X}}$ the complementary partition of X inherited from the “higher” call ($\bar{\mathcal{X}}$ is a partition of $\Omega \setminus X$);
- \mathfrak{D} the set of admissible partitions which refinements have already been evaluated during “higher” calls.

Output:

- \mathcal{X}^* a locally-optimal admissible partition of X .

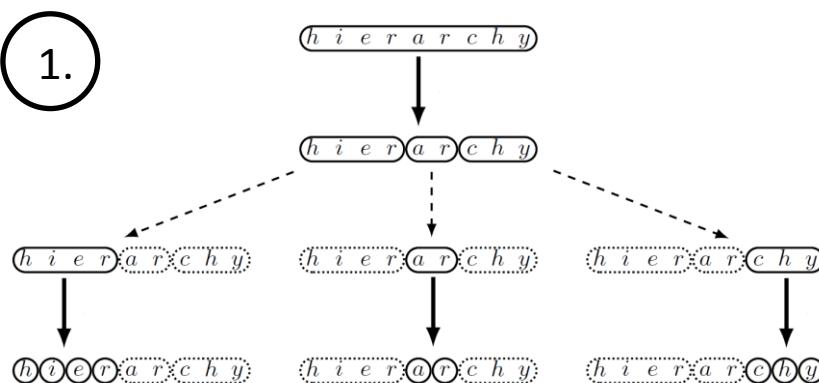
-
- If the algorithm has already been applied to part X , return the locally-optimal partition recorded in \mathcal{L} .
 - Initialization: $\mathcal{X}^* \leftarrow \{\{X\}\}$ and $\mathfrak{D}' \leftarrow \mathfrak{D}$.
 - For each $\mathcal{Y} \in \mathfrak{C}(\{X\})$ such that $\bar{\mathcal{X}} \cup \mathcal{Y}$ does not refine any partition in \mathfrak{D} , do the following:
 - For each part $Y \in \mathcal{Y}$, call the algorithm with local inputs $X \leftarrow Y$, $\bar{\mathcal{X}} \leftarrow \bar{\mathcal{X}} \cup \mathcal{Y} \setminus \{Y\}$, and $\mathfrak{D} \leftarrow \mathfrak{D}'$ to compute a locally-optimal partition $\mathcal{Y}_Y^* \in \mathfrak{P}^*(Y)$.
 - $\mathcal{Y}^* \leftarrow \bigcup_{Y \in \mathcal{Y}} \mathcal{Y}_Y^*$.
 - If $c(\mathcal{Y}^*) > c(\mathcal{X}^*)$, then $\mathcal{X}^* \leftarrow \mathcal{Y}^*$.
 - $\mathfrak{D}' \leftarrow \mathfrak{D}' \cup \{\mathcal{Y}\}$.
 - Return \mathcal{X}^* and record this result in \mathcal{L} .
-

Generic: solve any instance of the SPP
 → but inefficient for special versions

Designing dedicated implementations:

- ① Analysing the generic execution
- ② Building appropriate data structures
- ③ Deriving a specialized algorithm

Application to the Hierarchical SPP



2.

Data Structure

- Set of parts: rooted tree
- Optimal partition: cut of the tree
- Algorithm: depth-first search

3.

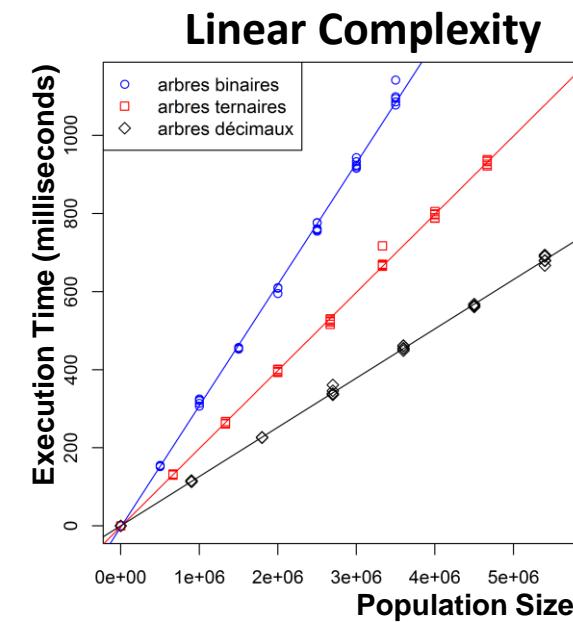
Algorithm 1 for the HSPP

Require: A tree with a label *cost* on each node representing the cost of the corresponding admissible part.

Ensure: Each node of the tree has a Boolean label *optimalCut* representing an optimal partition (see above).

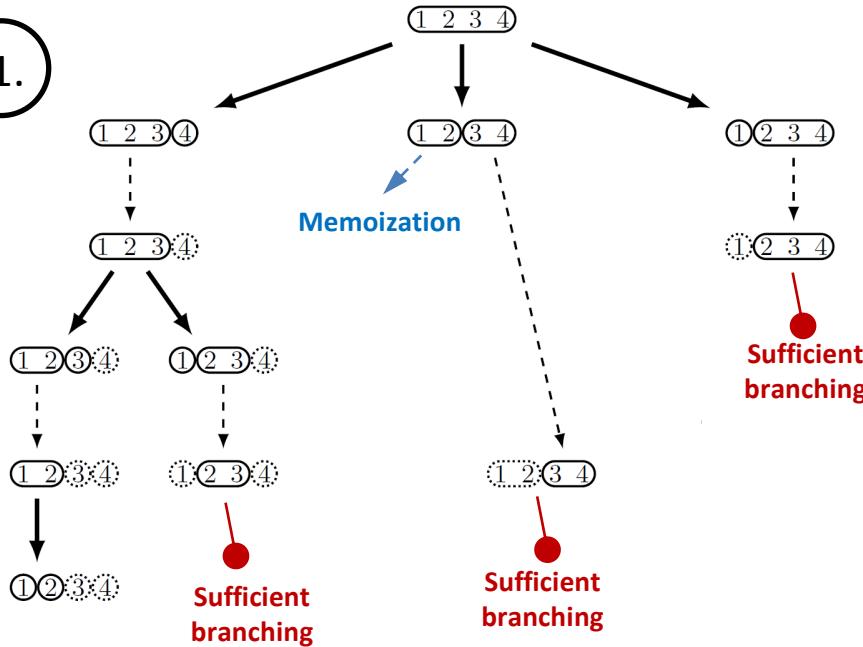
```

procedure SOLVEHSPP(node)
  if node has no child then
    node.optimalCost ← node.cost
    node.optimalCut ← true
  else
    MCost ← node.cost
    μCost ← 0
    for each child of node do
      SOLVEHSPP(child)
      μCost ← μCost + child.optimalCost
    node.optimalCost ← max(μCost, MCost)
    node.optimalCut ← (μCost < MCost)
  end if
end procedure
  
```



Application to the Ordered SPP

1.



2.

Data Structure

- Set of parts: triangular matrix
- Optimal partition: array of cuts
- Algorithm: dynamic programming

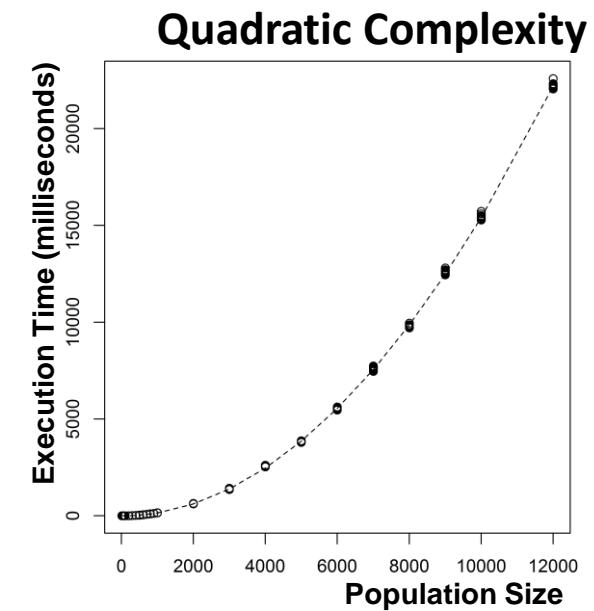
3.

Algorithm 2 for the OSPP

Require: A matrix $cost$ recording the costs of intervals.
Ensure: The vector $optimalCut$ represents an optimal partition (see text above).

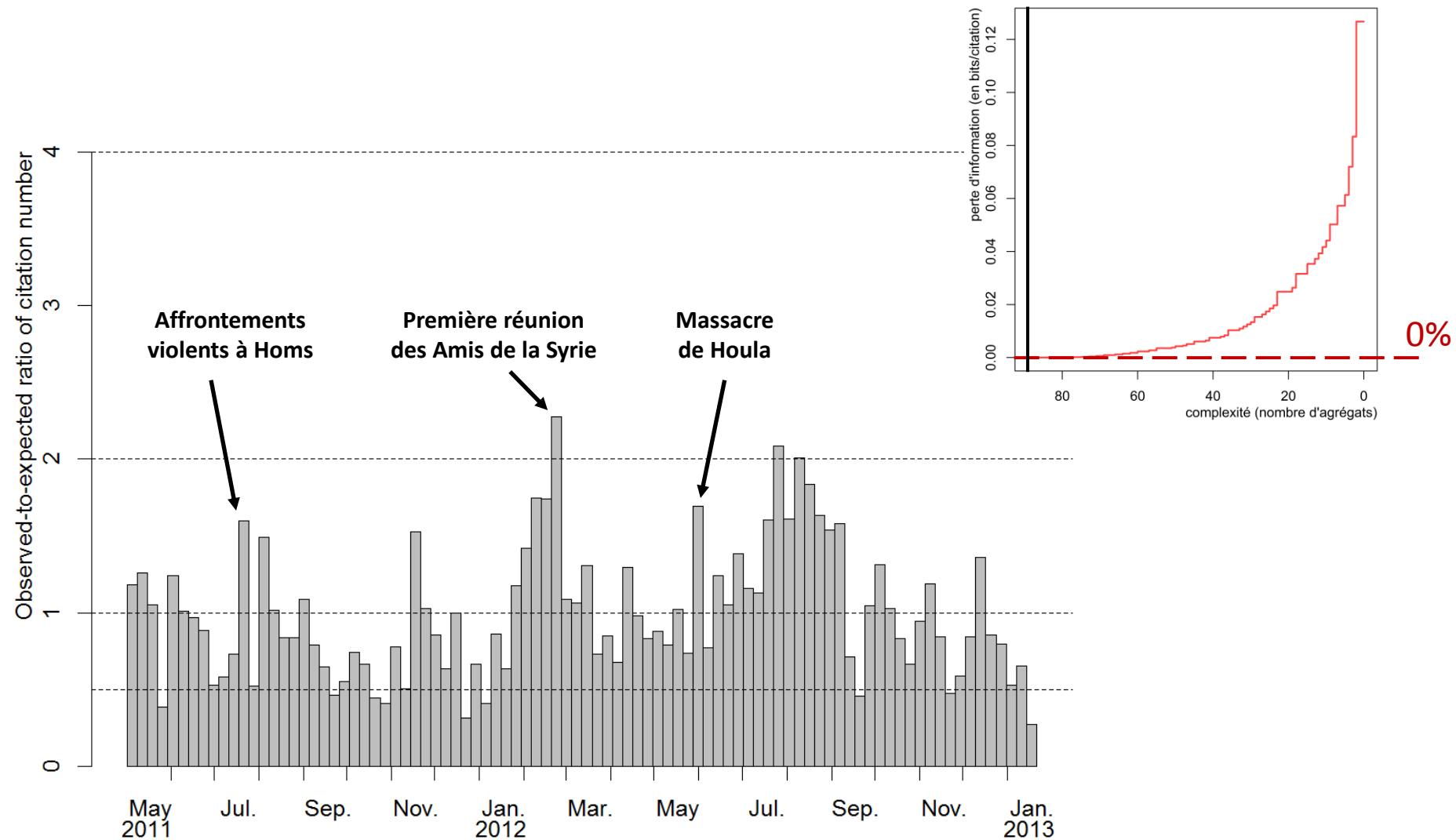
```

for  $j \in [1, n]$  do
     $optimalCost[j] \leftarrow cost[1, j]$ 
     $optimalCut[j] \leftarrow 1$ 
    for  $cut \in [2, j]$  do
         $\muCost \leftarrow optimalCost[cut - 1] + cost[cut, j]$ 
        if  $\muCost > optimalCost[j]$  then
             $optimalCost[j] \leftarrow \muCost$ 
             $optimalCut[j] \leftarrow cut$ 
```



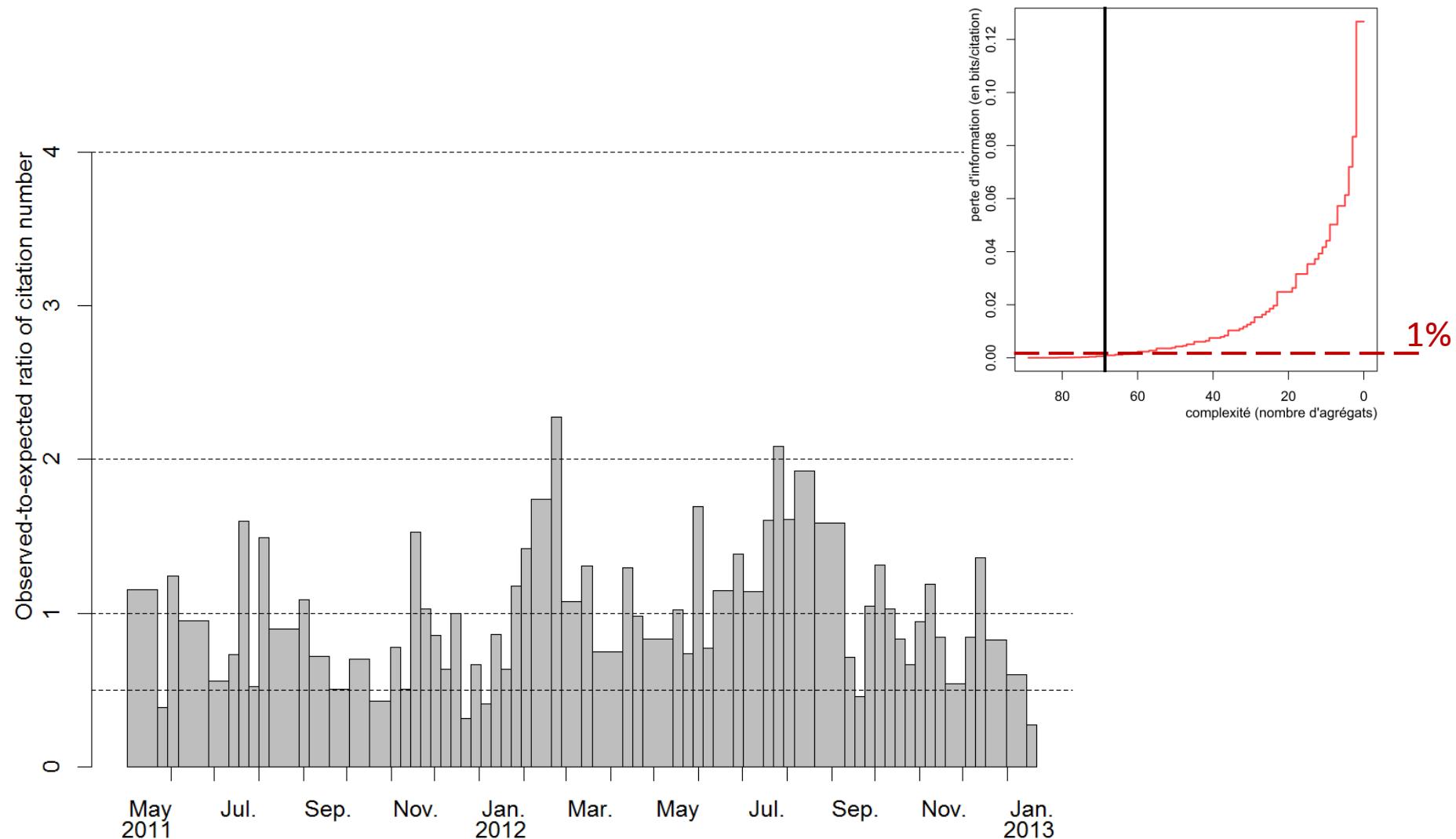
La Syrie vue par LE MONDE

[Giraud, Grasland, Lamarche-Perrin *et al.*, ECTQG 2013]



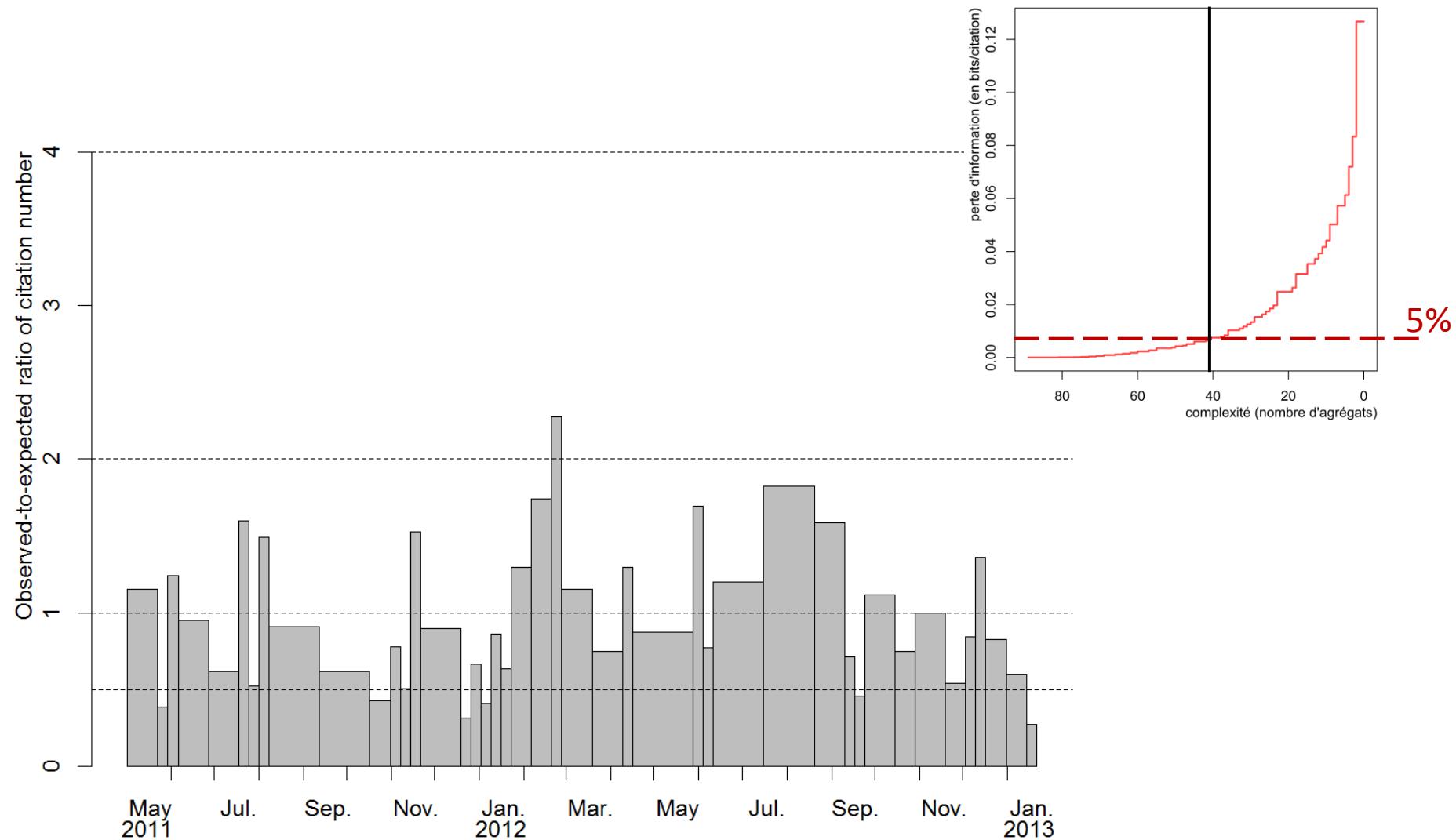
La Syrie vue par LE MONDE

[Graud, Grasland, Lamarche-Perrin *et al.*, ECTQG 2013]



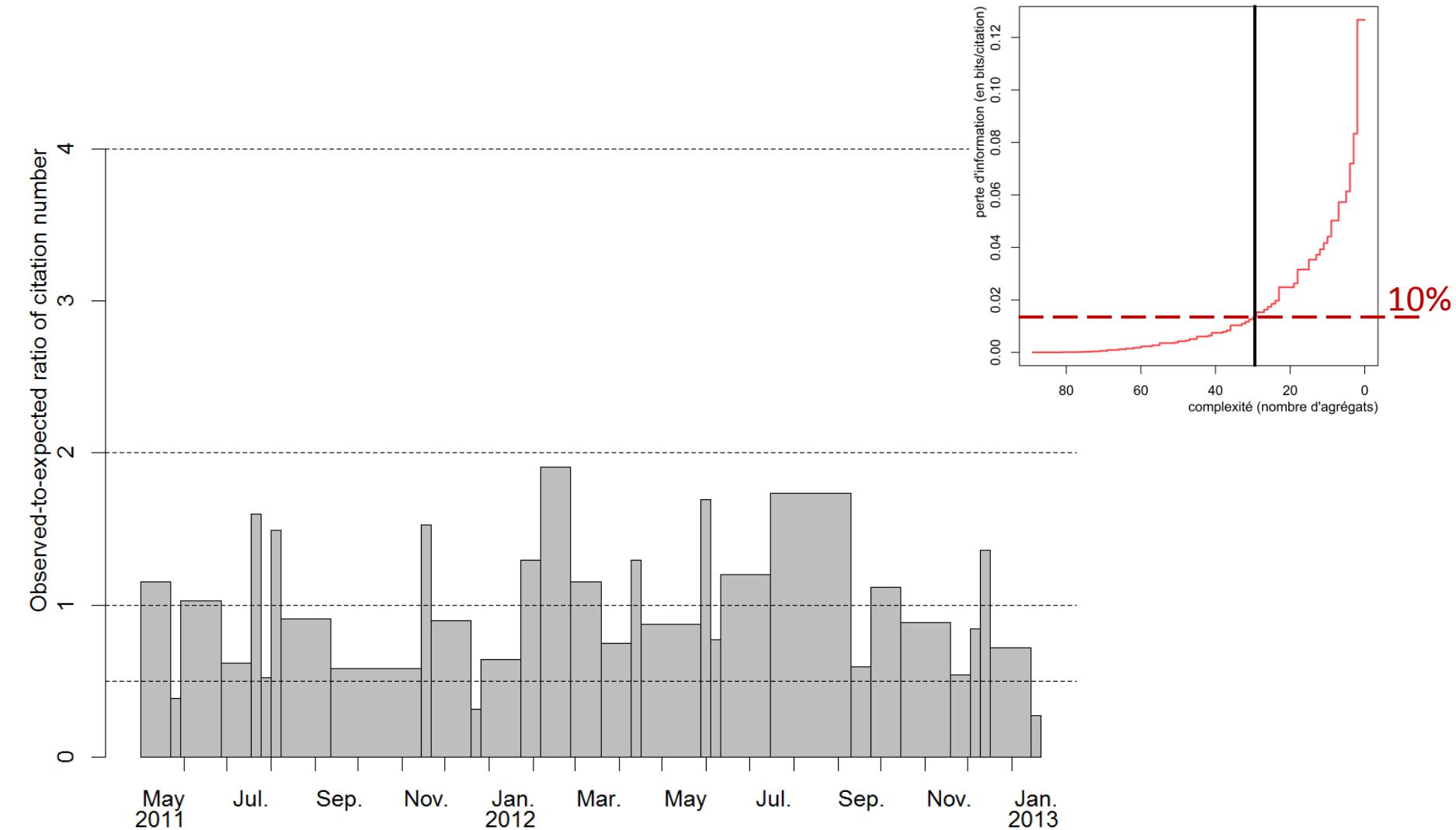
La Syrie vue par LE MONDE

[Graud, Grasland, Lamarche-Perrin *et al.*, ECTQG 2013]



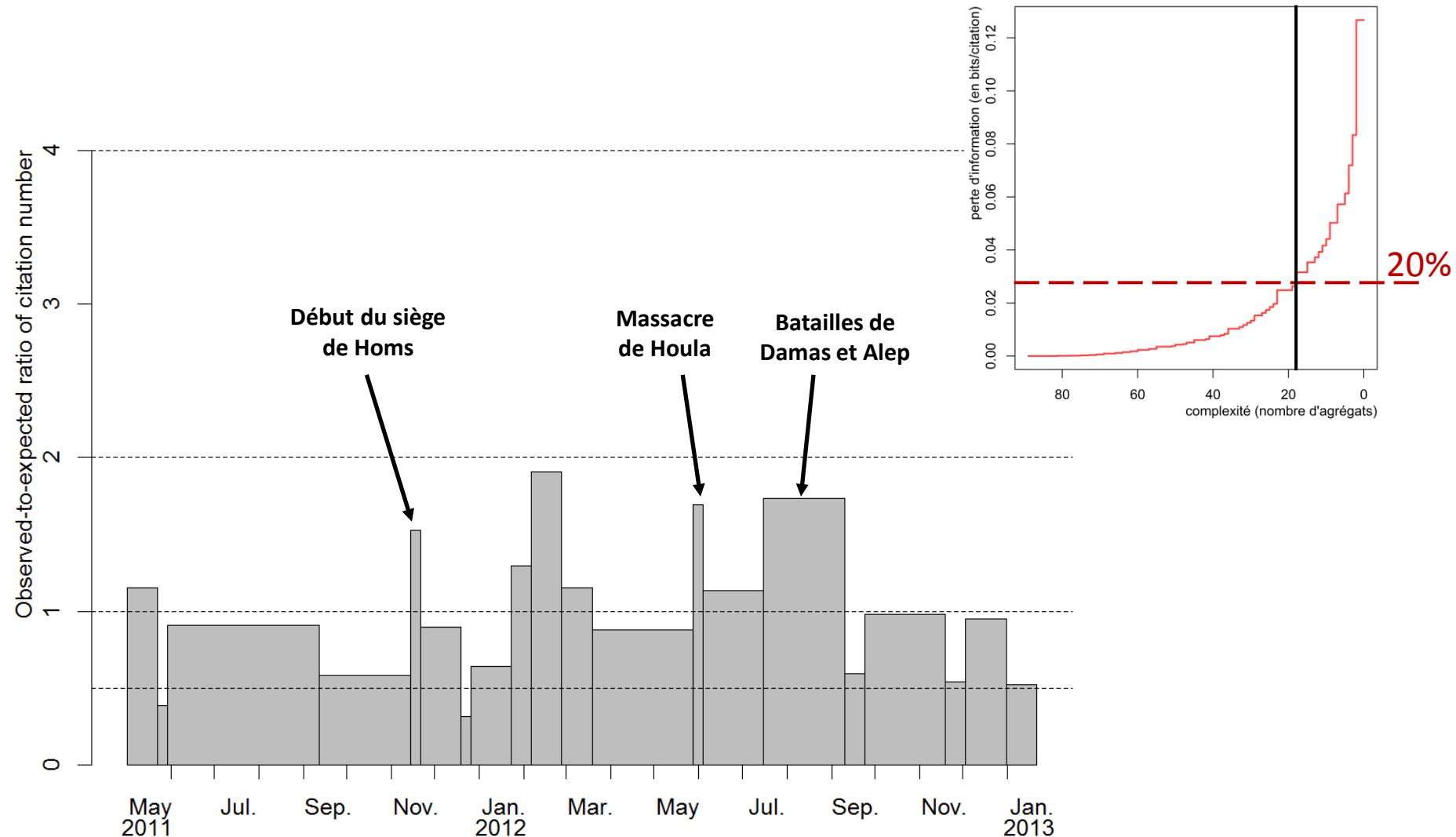
La Syrie vue par LE MONDE

[Graud, Grasland, Lamarche-Perrin *et al.*, ECTQG 2013]



La Syrie vue par LE MONDE

[Giraud, Grasland, Lamarche-Perrin *et al.*, ECTQG 2013]

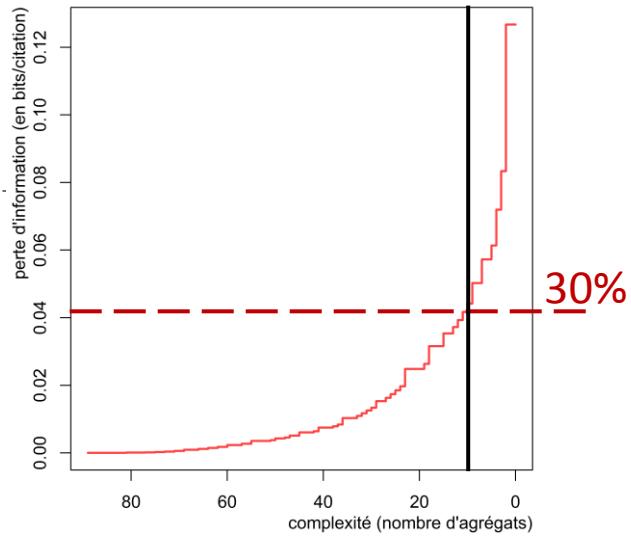
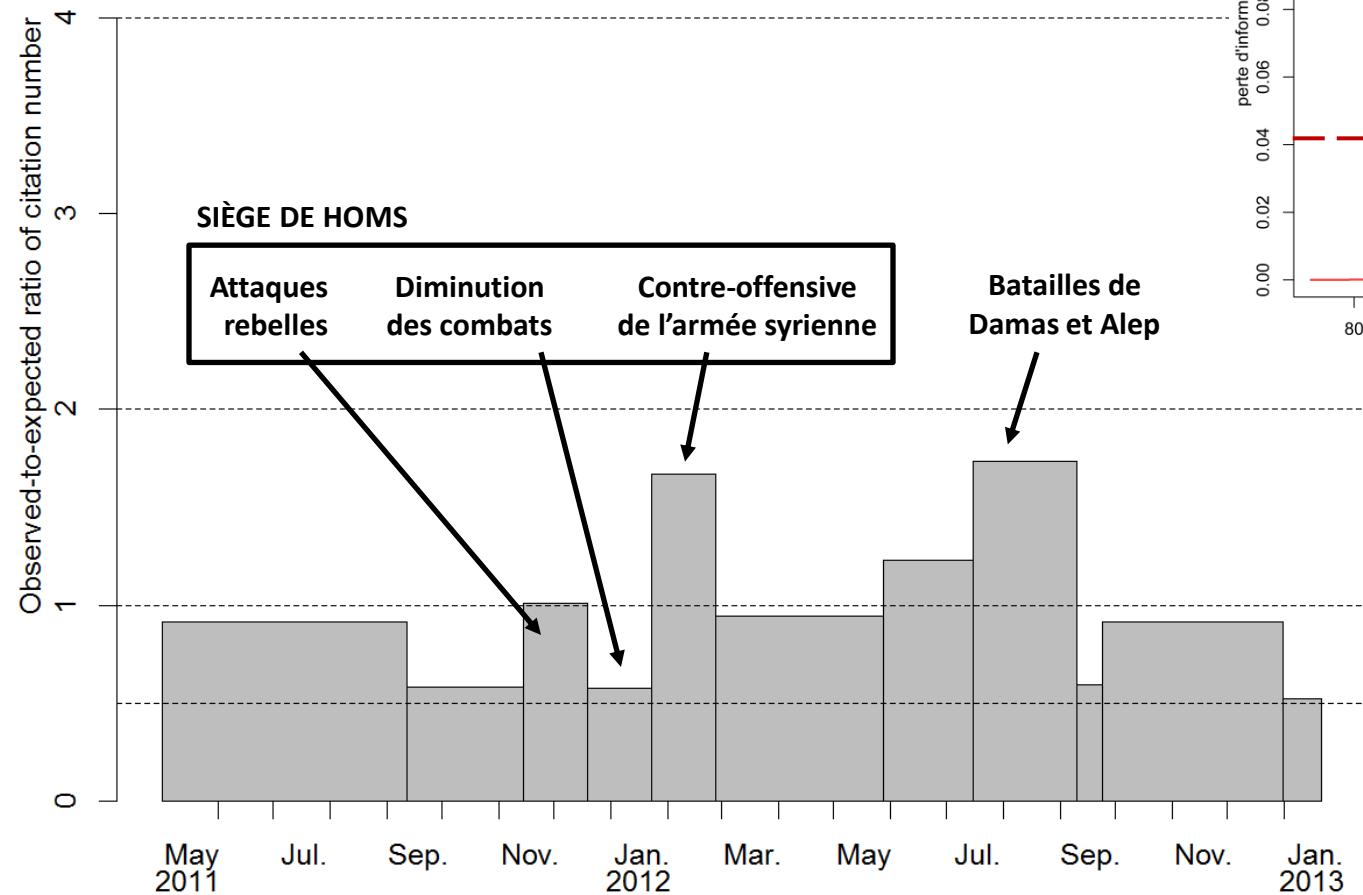


La Syrie vue par LE MONDE

[Graud, Grasland, Lamarche-Perrin *et al.*, ECTQG 2013]

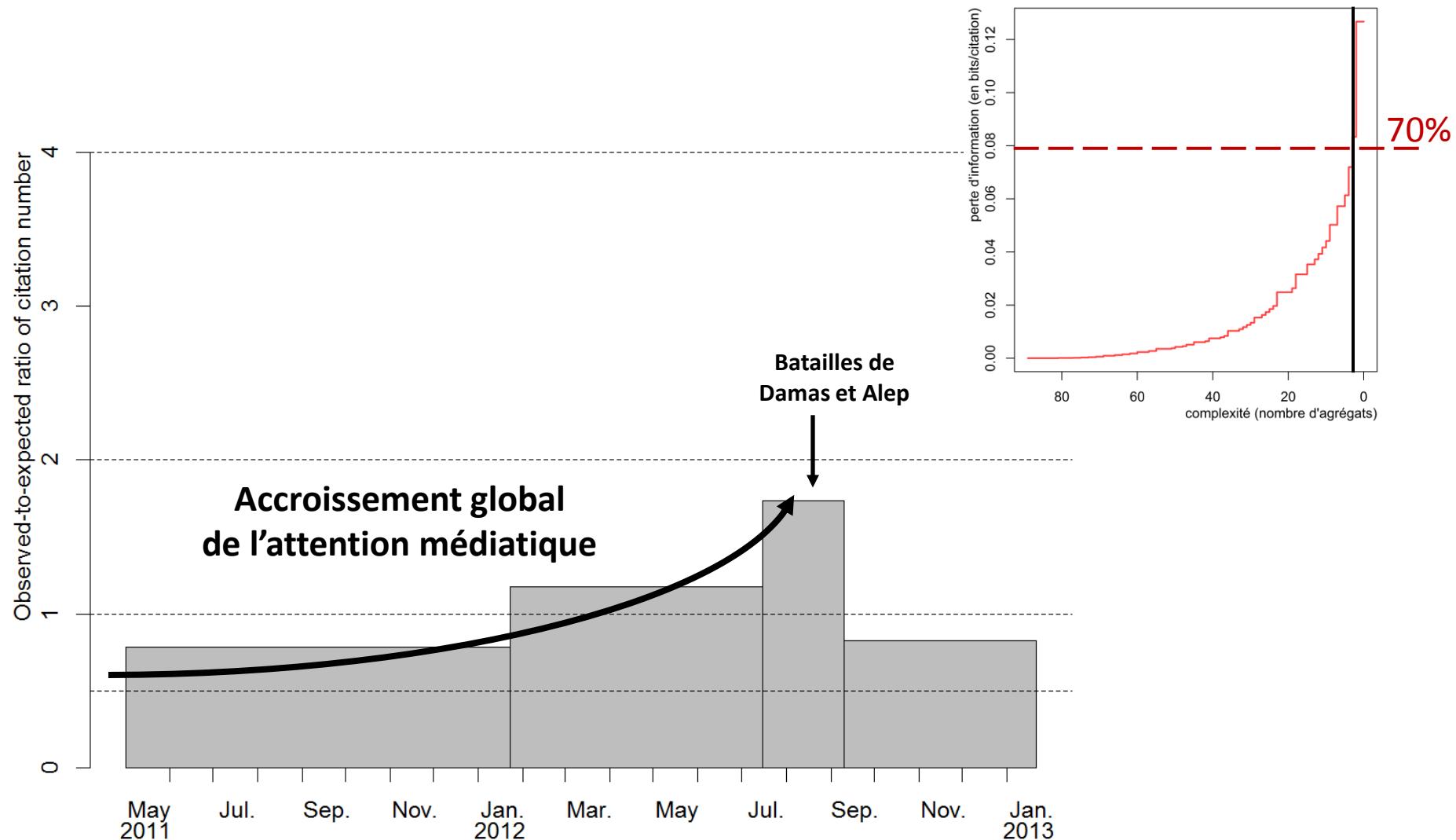
Source : Wikipedia

Timeline of the Syrian civil war
Siege of Homs



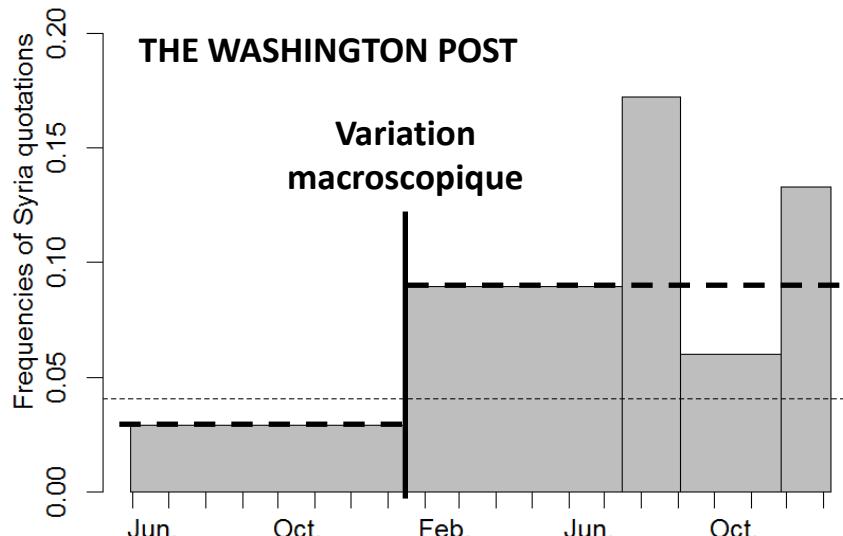
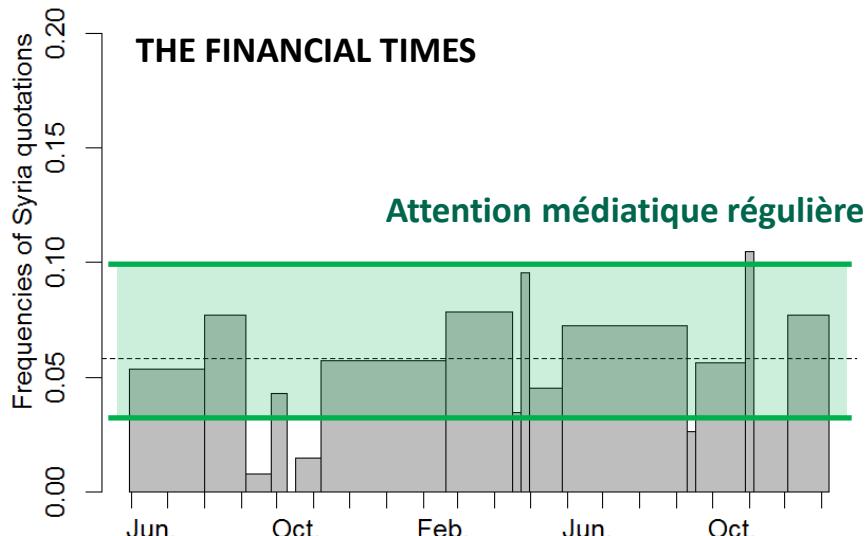
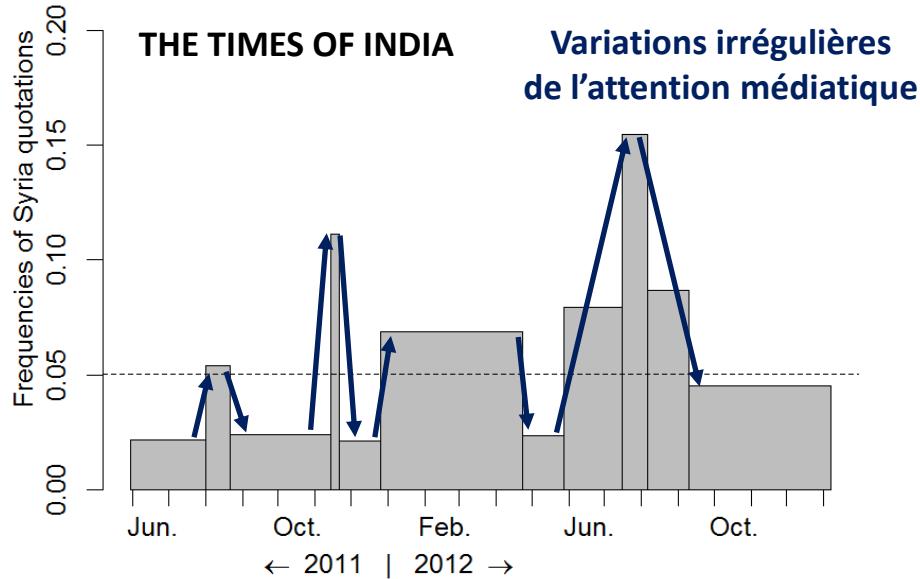
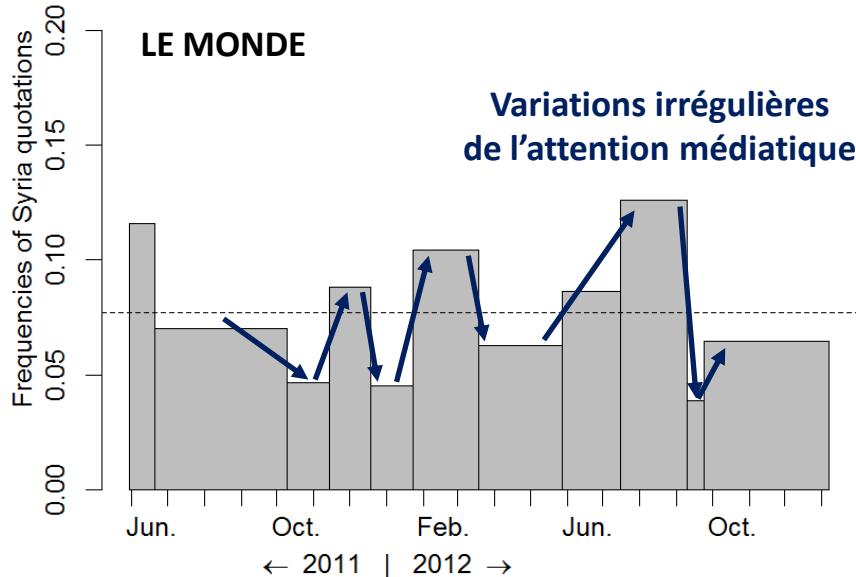
La Syrie vue par LE MONDE

[Giraud, Grasland, Lamarche-Perrin *et al.*, ECTQG 2013]

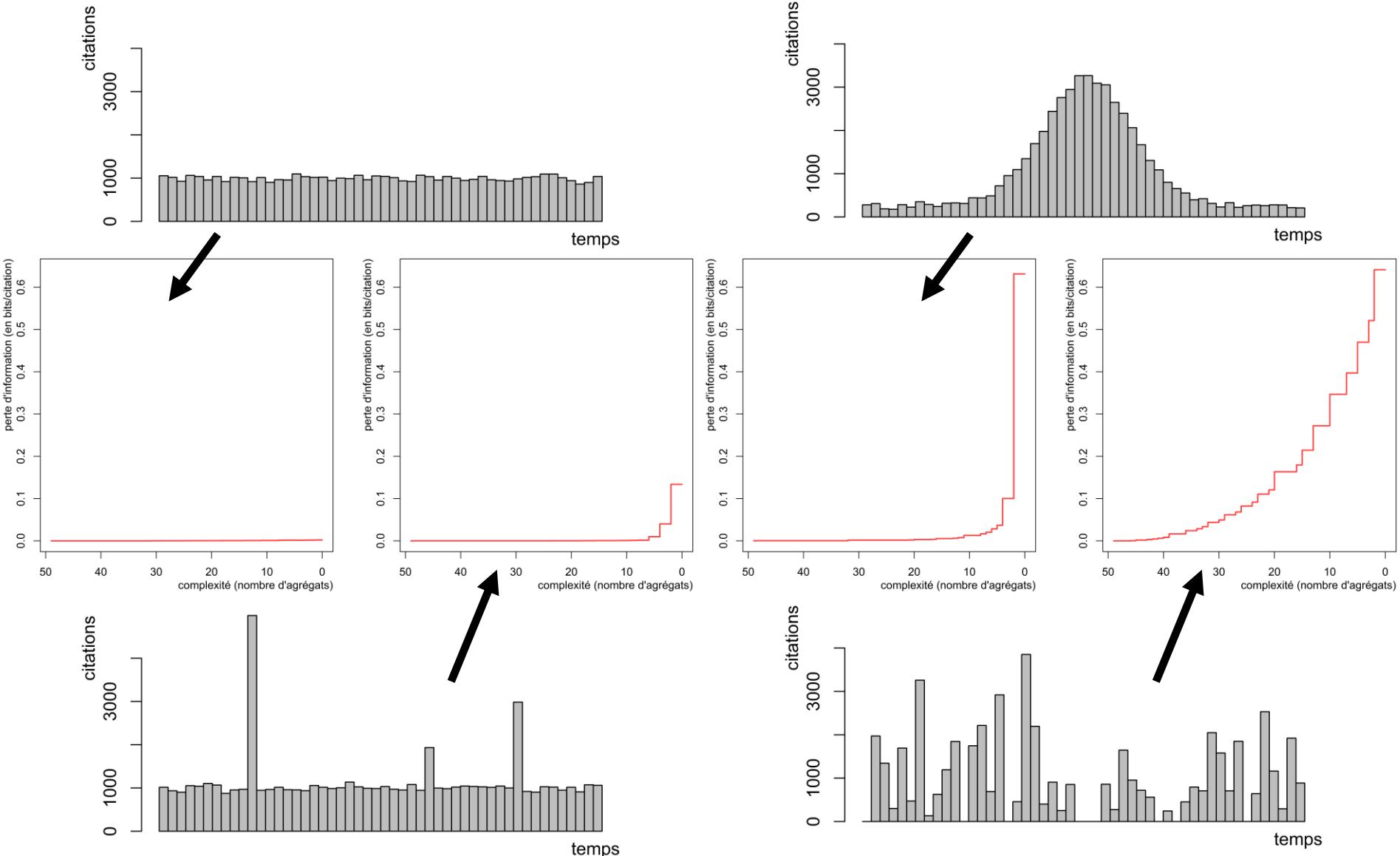


La Syrie vue par 4 journaux

[Giraud, Grasland, Lamarche-Perrin *et al.*, ECTQG 2013]



Caractériser un jeu de données

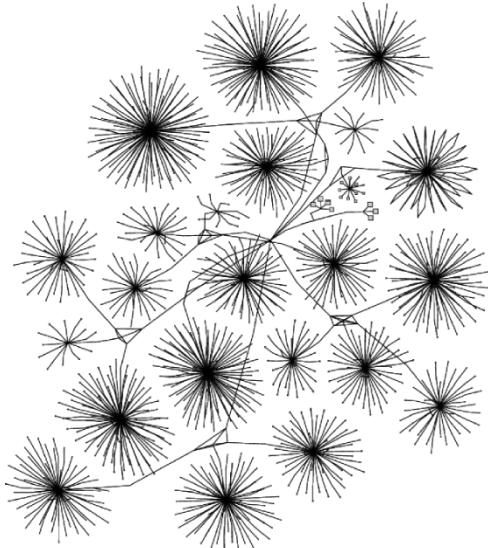


APPLICATION À L'ANALYSE DE TRACES D'EXECUTION

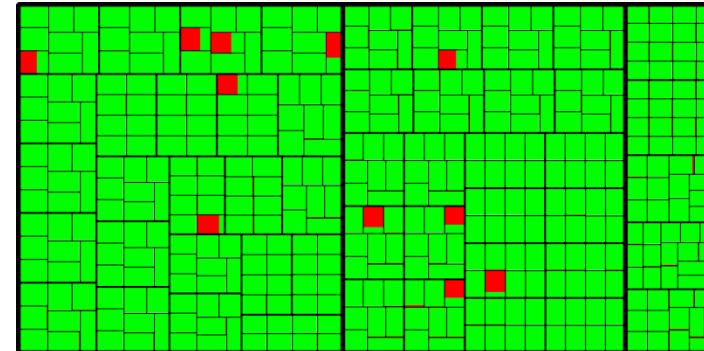
Agrégation de traces d'exécution

[Lamarche-Perrin, Schnorr *et al.*, TSI 2013]

Structure hiérarchique
de la grille de calcul
[Schnorr *et al.*, 2013]

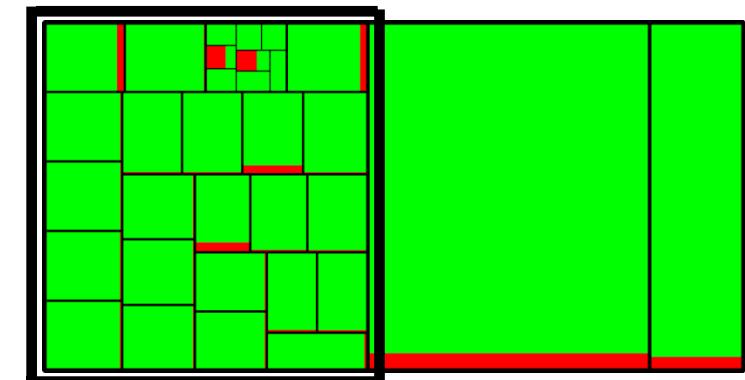
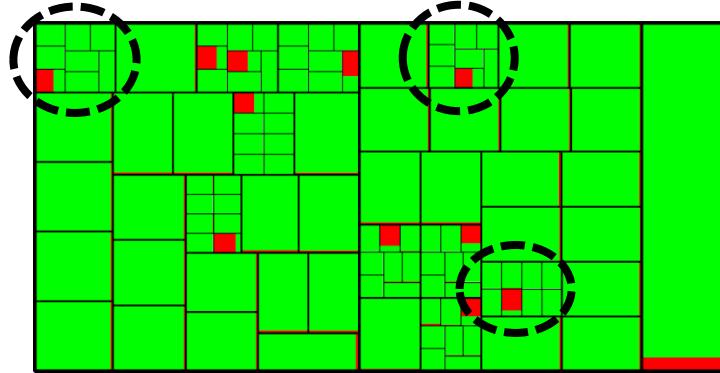


Représentation treemap
microscopique



Représentations treemap agrégées

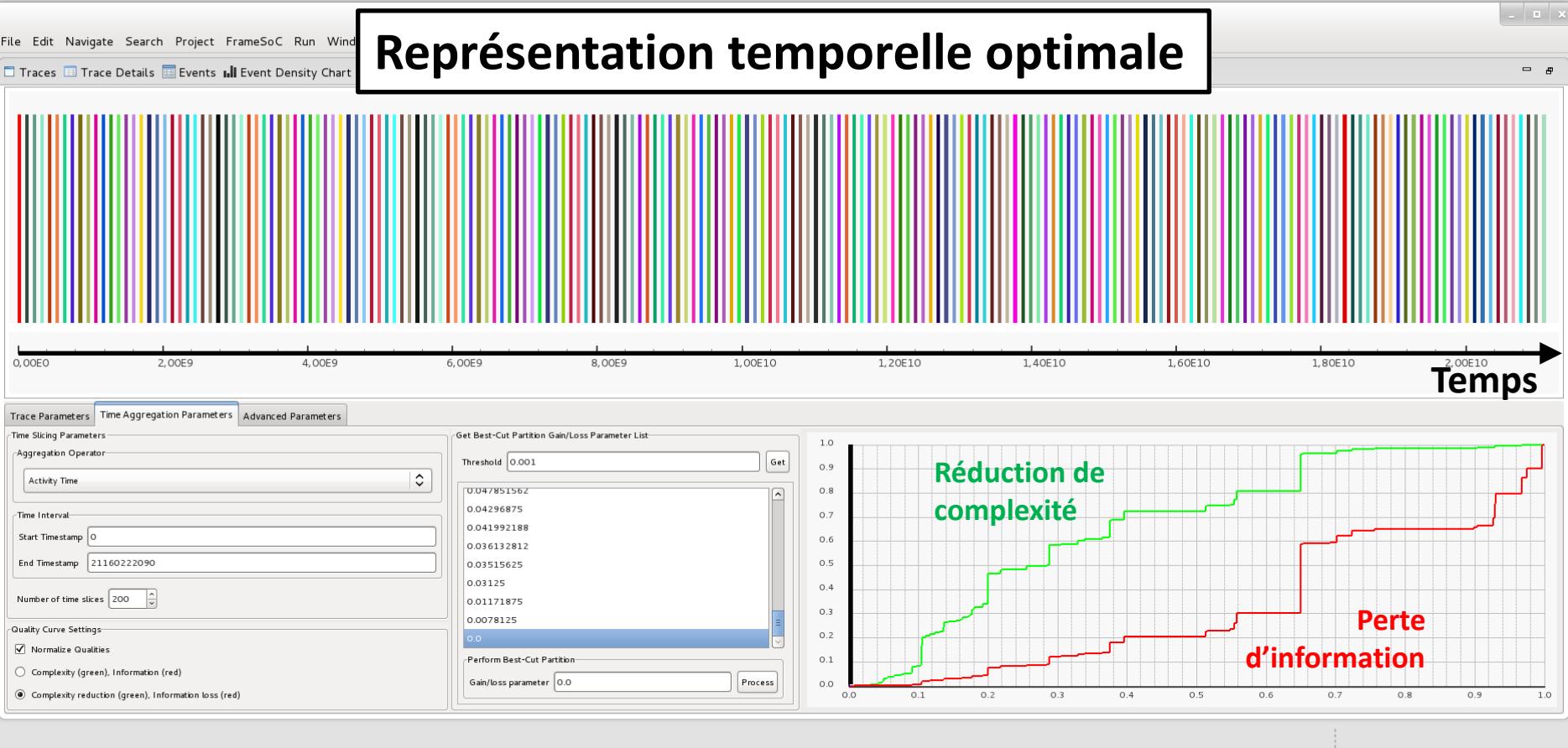
Détection
d'anomalies
multi-échelle



Agrégation de traces d'exécution

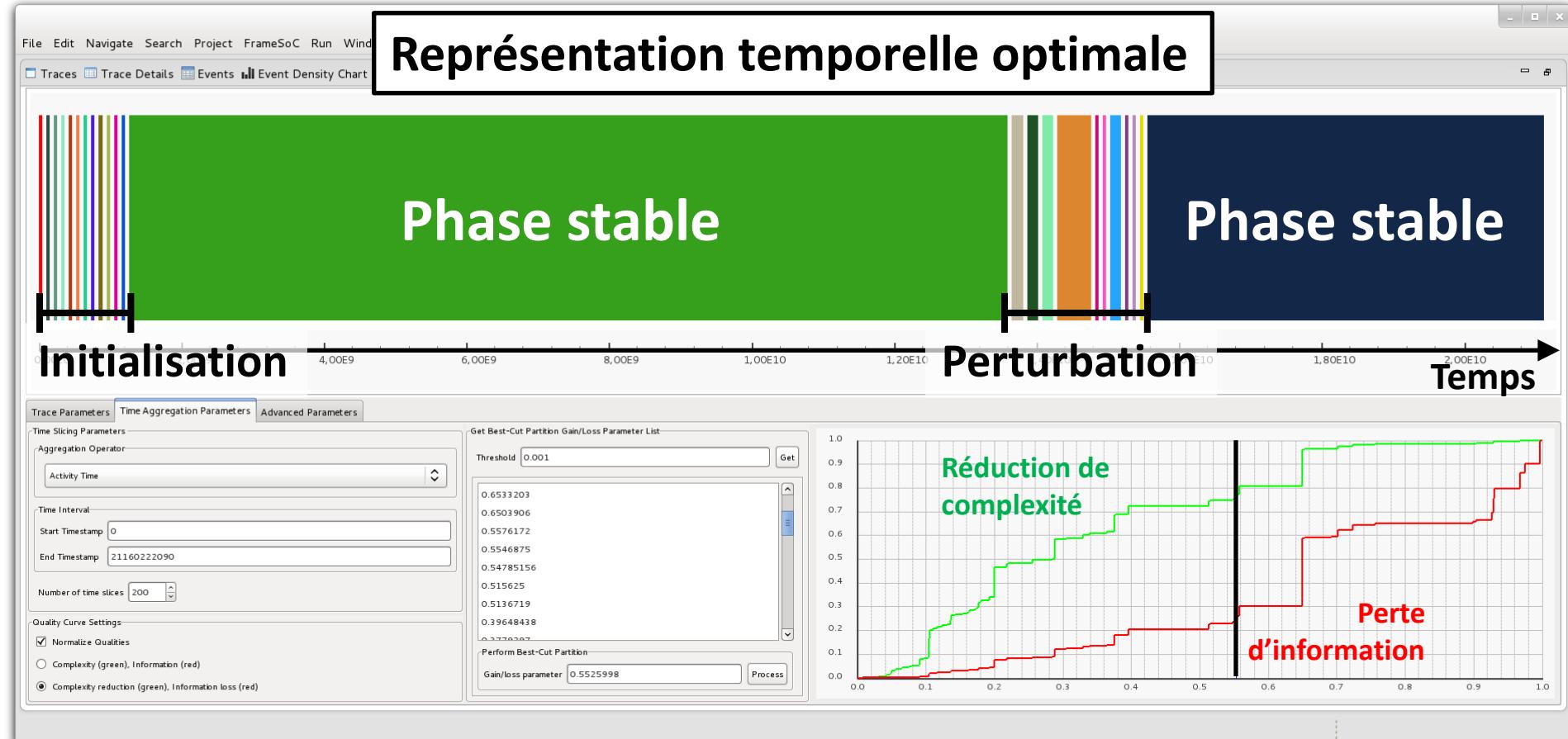
[Pagano, Dosimont et al., 2013]

Représentation temporelle optimale



Agrégation de traces d'exécution

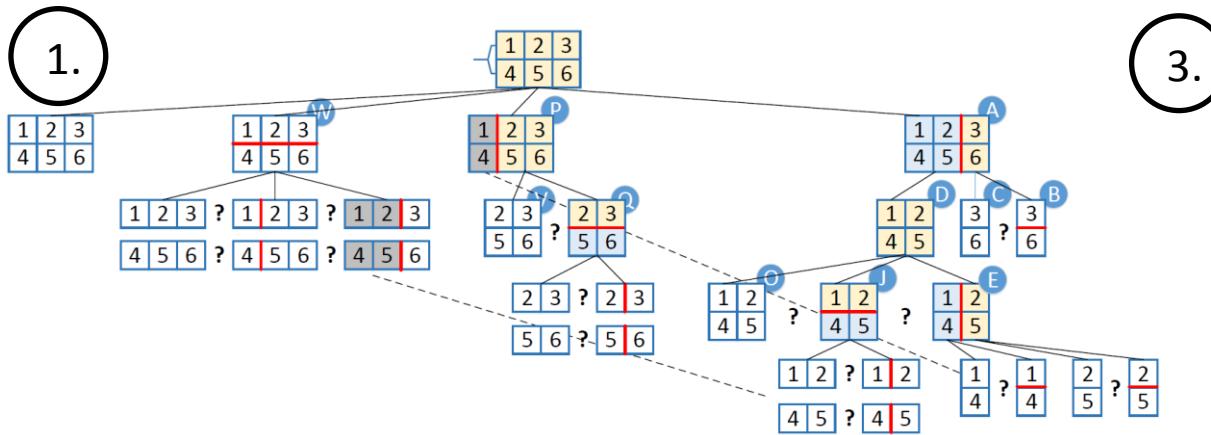
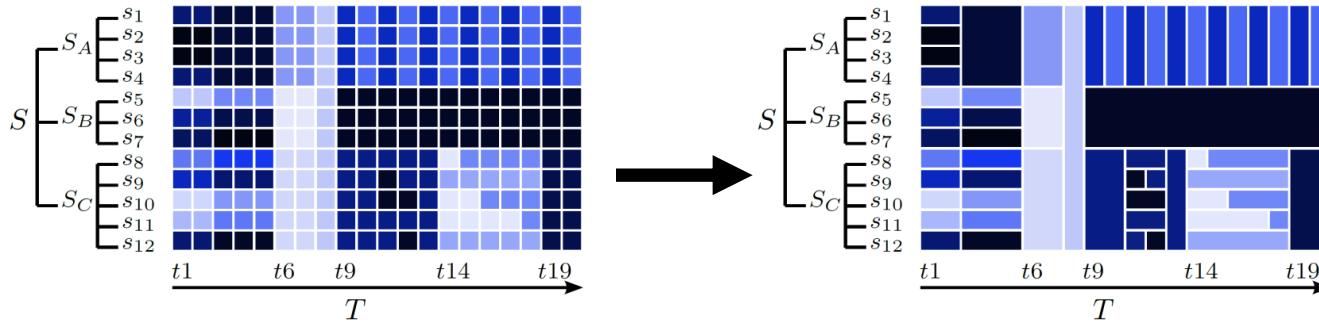
[Pagano, Dosimont *et al.*, 2013]



Application to a Multidimensional SPP

78

[Dosimont *et al.*, CLUSTER 2014]



2. Data Structure

- Set of parts: rooted tree of triangular matrices
- Optimal partition: cut of the tree and arrays of cuts
- Algorithm: depth-first search and dynamic programming

Algorithm 1 computes a hierarchy-and-order-consistent partition that maximizes the parametrized information criterion

```

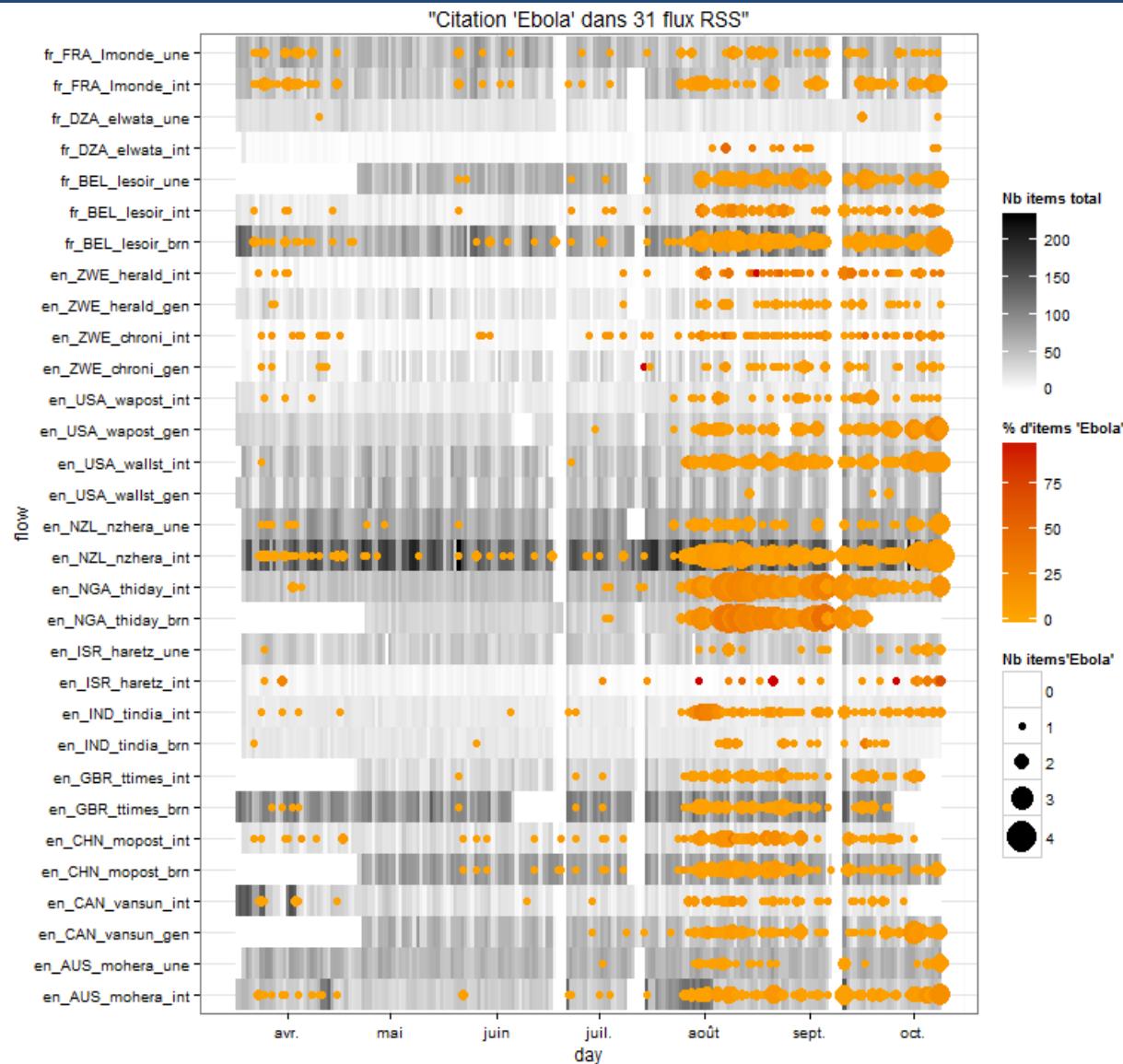
procedure node.COMPUTEOPTIMALPARTITION(p)
    for each child do                                ▷ Recursion
        child.COMPUTEOPTIMALPARTITION(p)

    for i = |T| - 1, ..., 0 do                      ▷ Iteration
        for j = i, ..., |T| - 1 do
            cut[i, j] ← j
            pIC[i, j] ← p.gain[i, j] - (1 - p).loss[i, j]

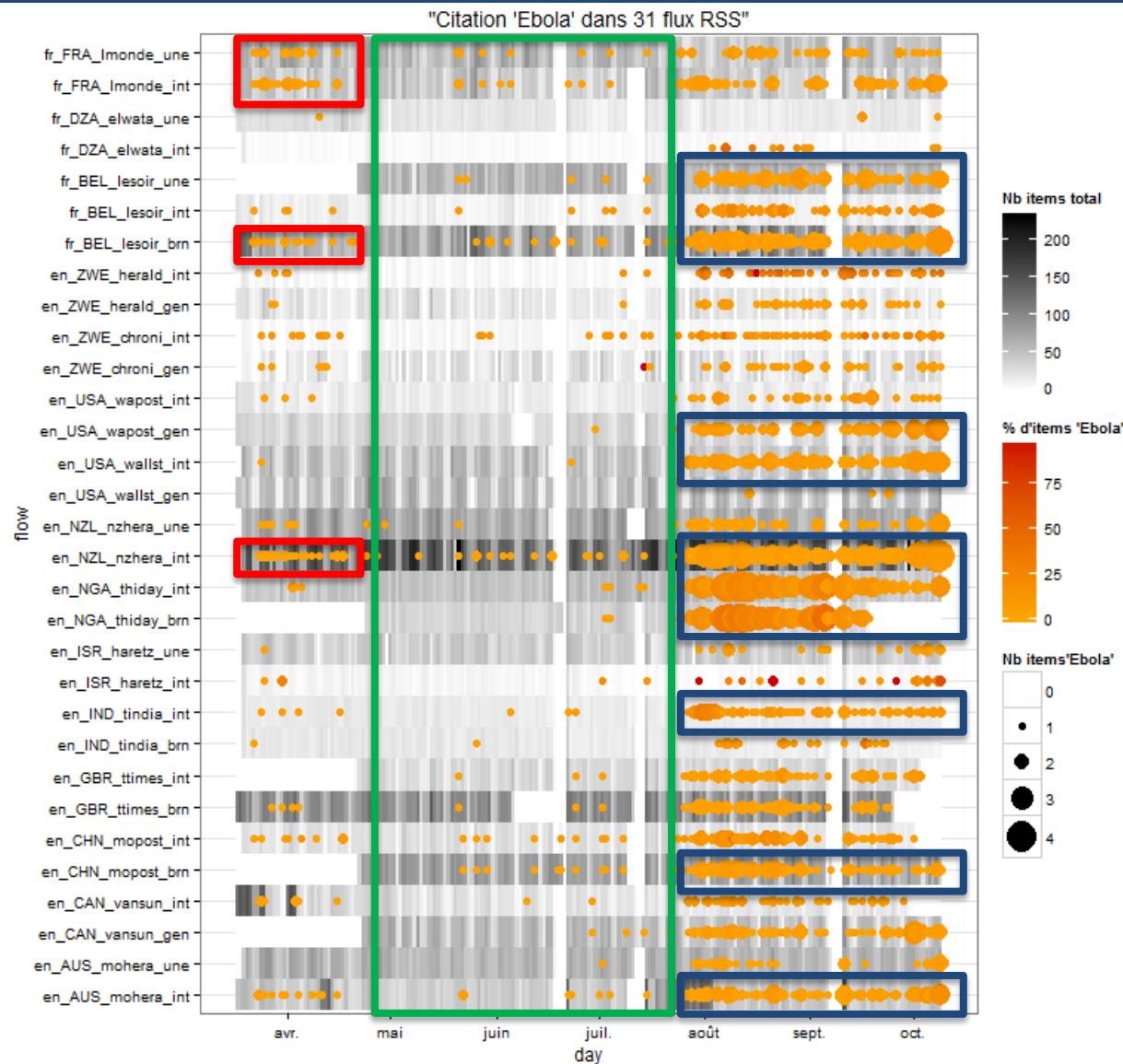
            if has children then                      ▷ Spatial cut?
                pICs ← 0
                for each child do
                    pICs ← pICs + child.pIC[i, j]
                if pICs > pIC[i, j] then
                    cut[i, j] ← -1
                    pIC[i, j] ← pICs

            for cutt = i, ..., j - 1 do           ▷ Temporal cut?
                pICt ← pIC[i, cut] + pIC[cut + 1, j]
                if pICt > pIC[i, j] then
                    cut[i, j] ← cutt
                    pIC[i, j] ← pICt
```

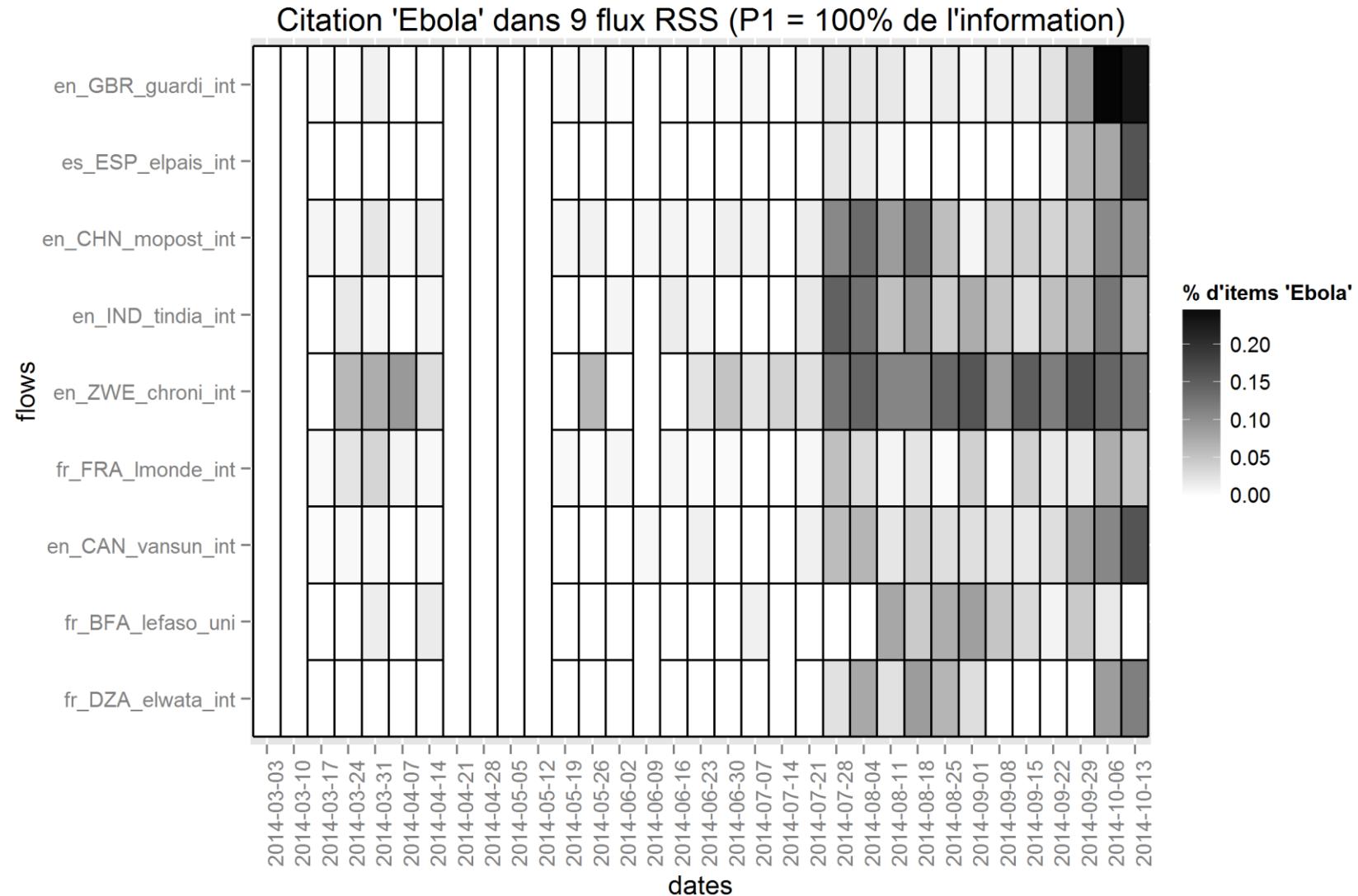
Agrégation médiatique



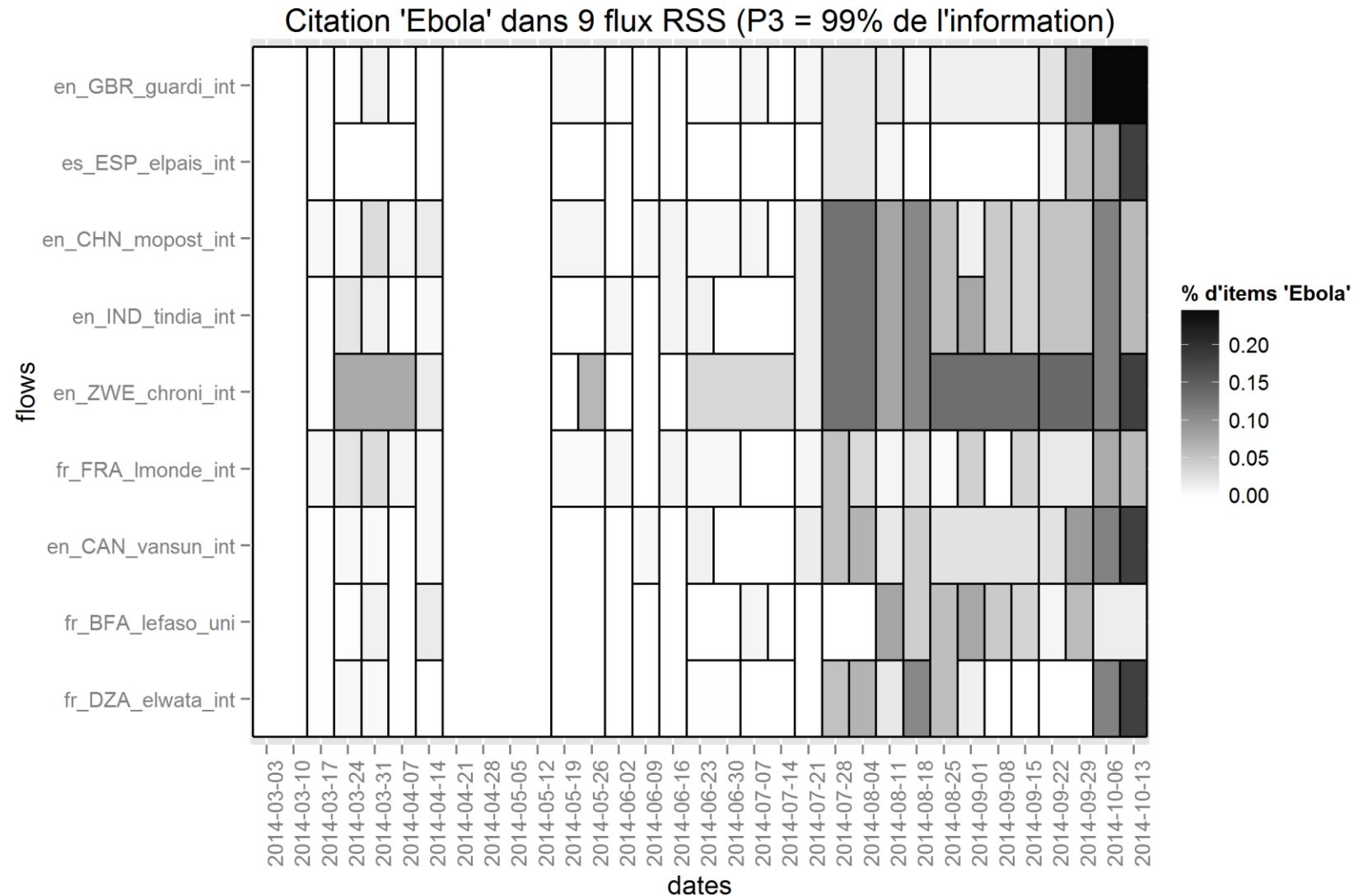
Agrégation médiatique



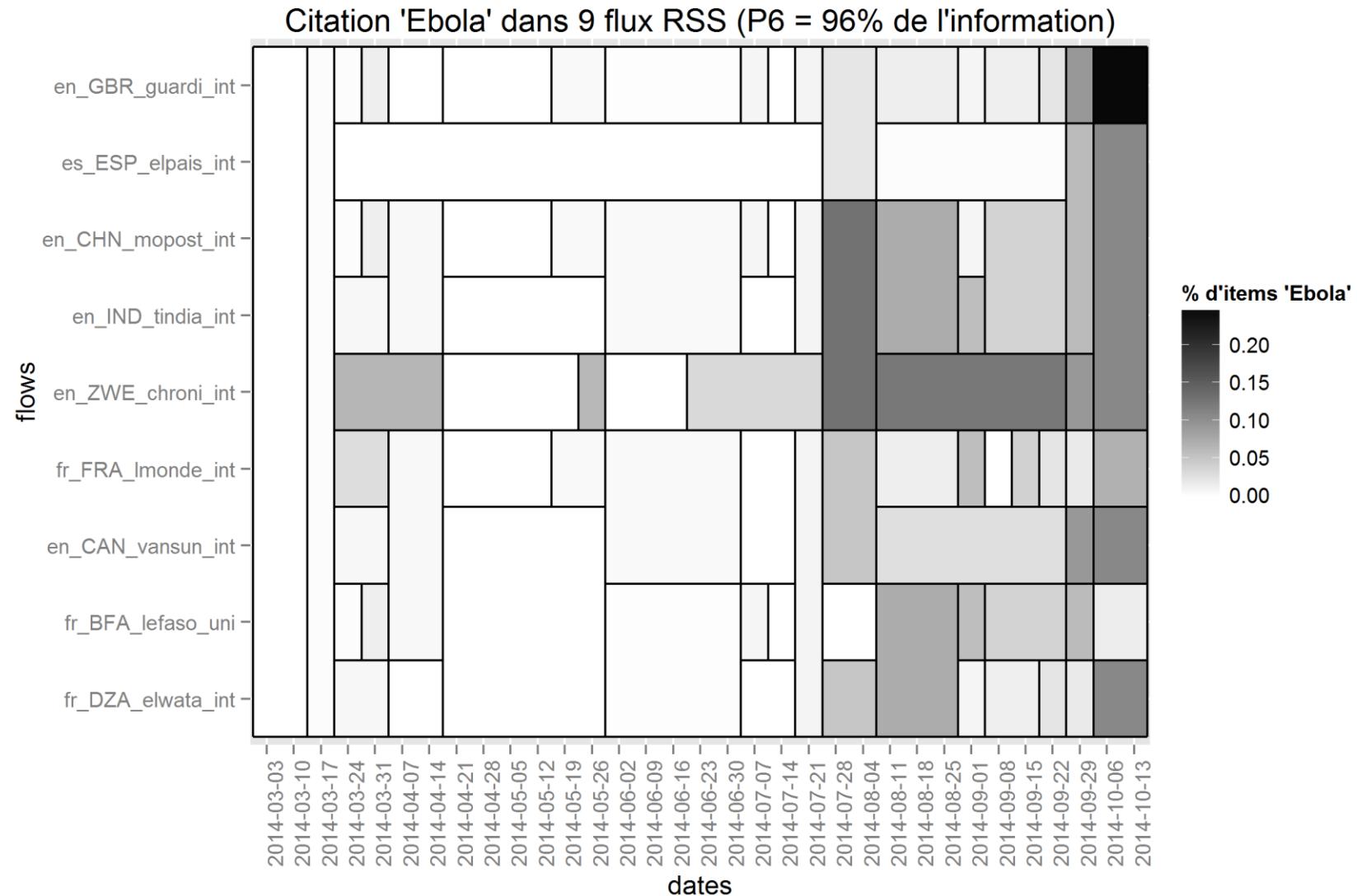
Agrégation médiatique



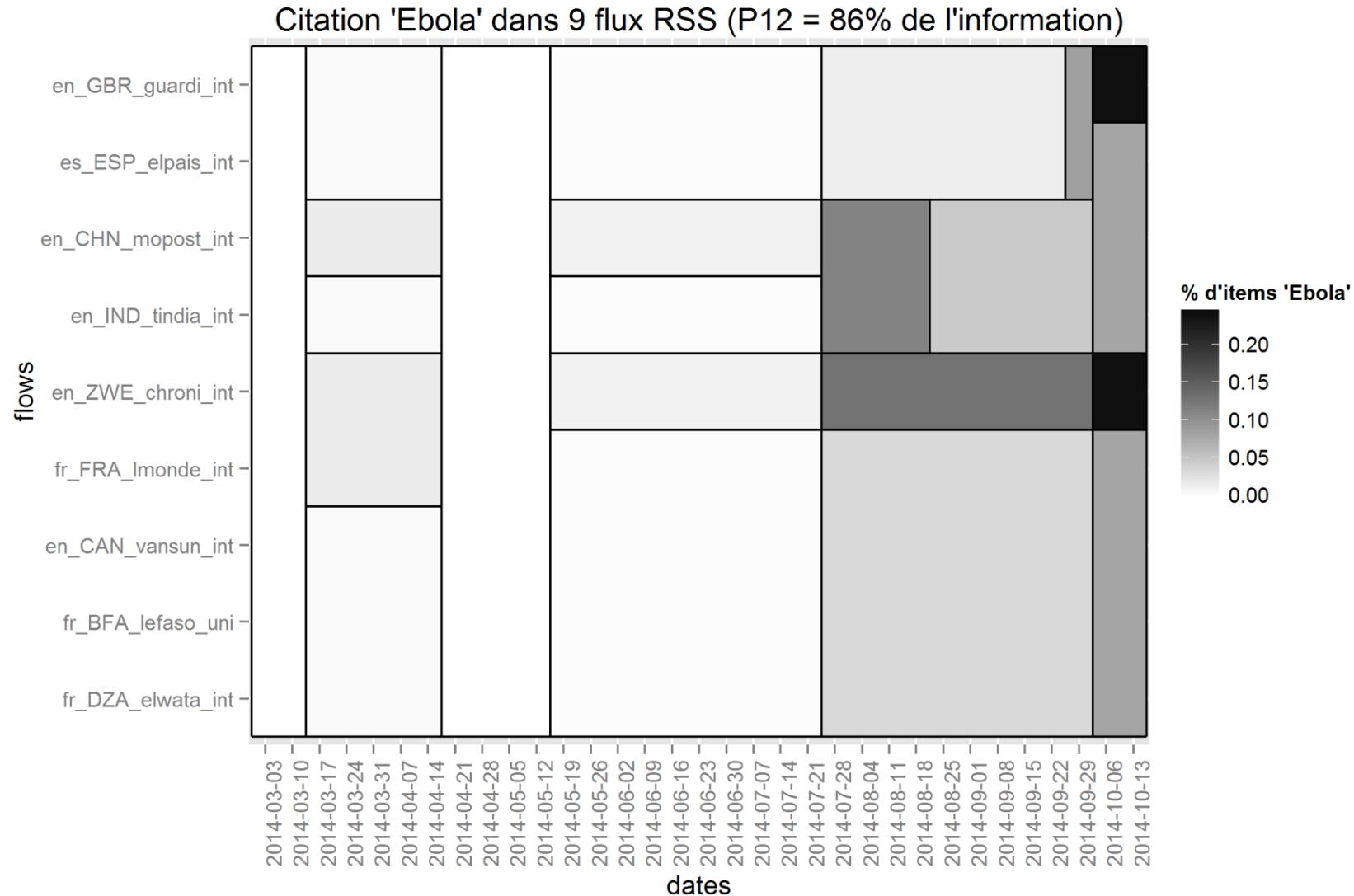
Agrégation médiatique



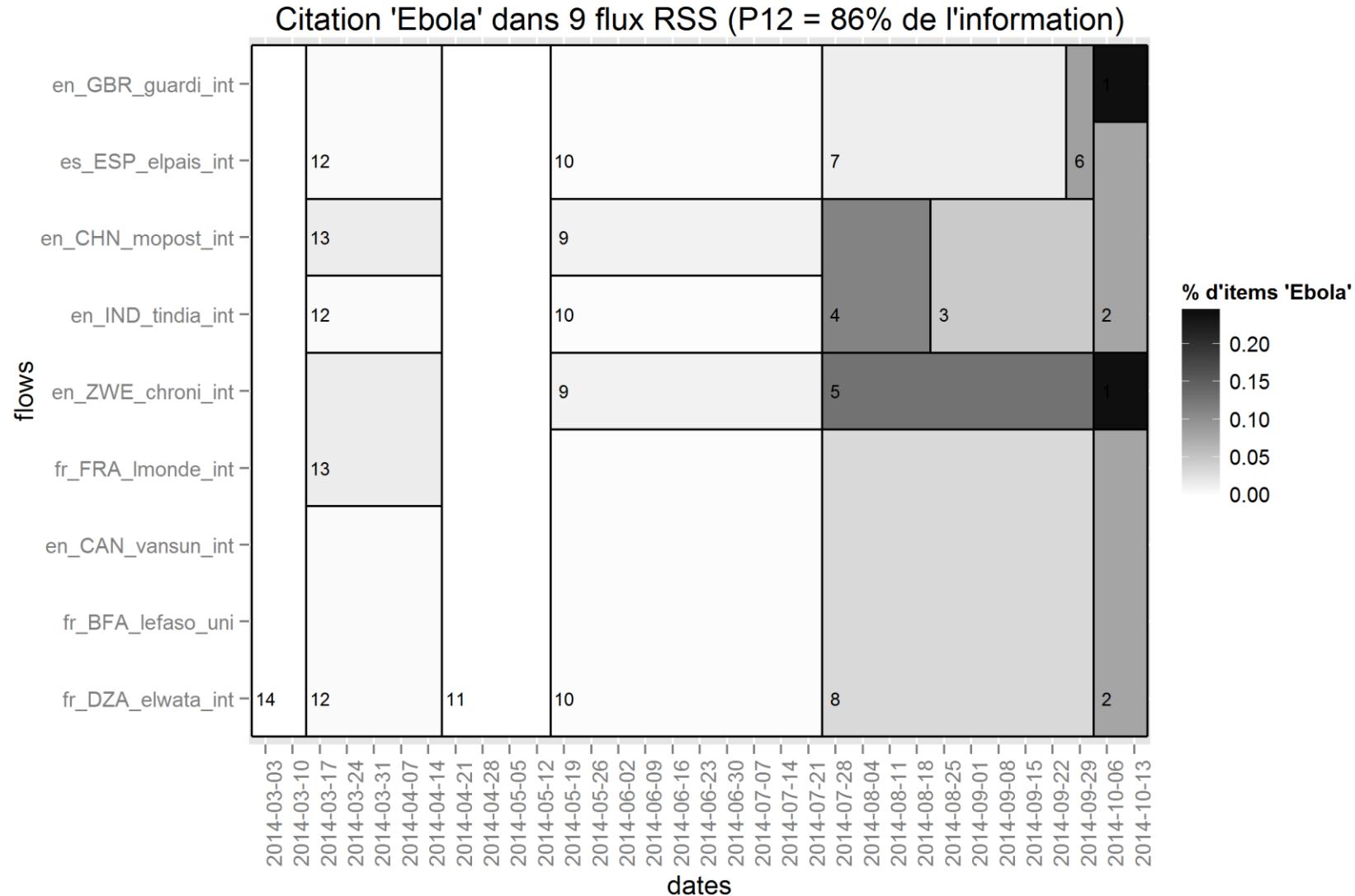
Agrégation médiatique



Agrégation médiatique



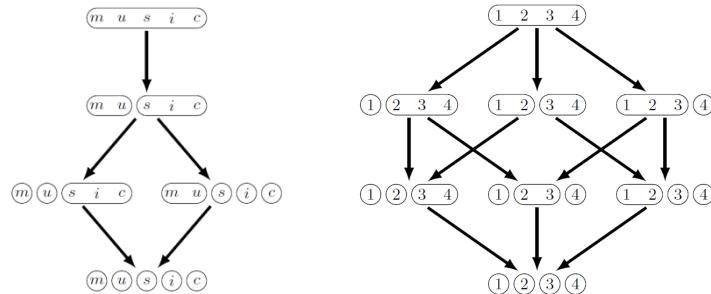
Agrégation médiatique



BILAN ET PERSPECTIVES

Bilan des contributions

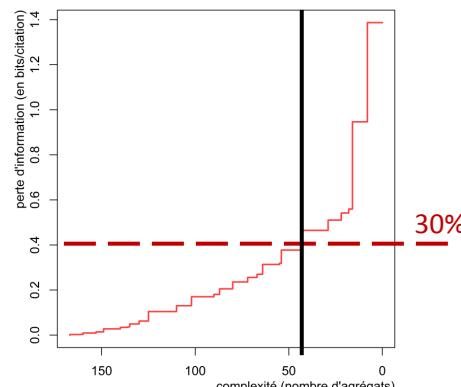
P1 Des structures algébriques exprimant la **sémantique du système**



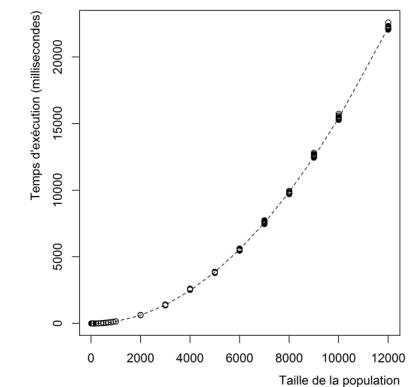
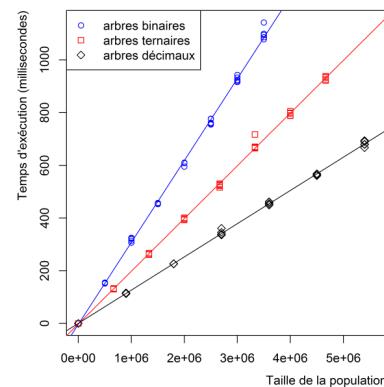
P2 Un compromis de qualité pour engendrer des **représentations multi-résolution**

$$CQL_\alpha = \alpha \frac{\Delta T}{\Delta T_{\max}} - (1 - \alpha) \frac{D}{D_{\max}}$$

P2 Des graphes de qualité pour **choisir la granularité** des représentations

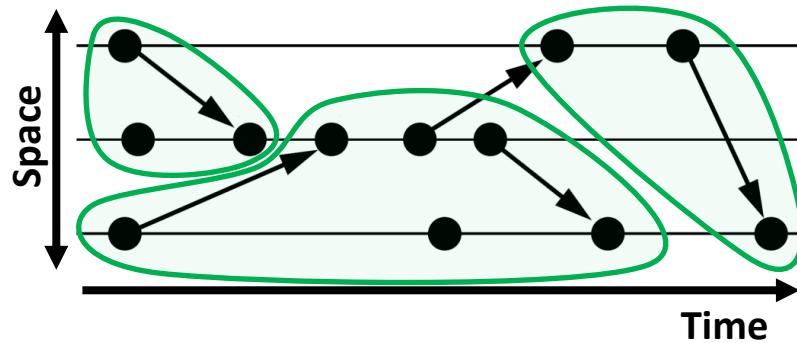


P3 Un algorithme d'agrégation générique de **complexité polynomiale**



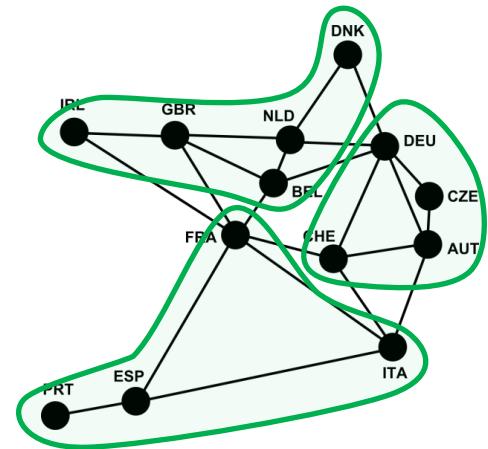
Application Perspectives

Partitioning of Interaction Diagrams



[Mattern, 1989]

Partitioning of Graphs

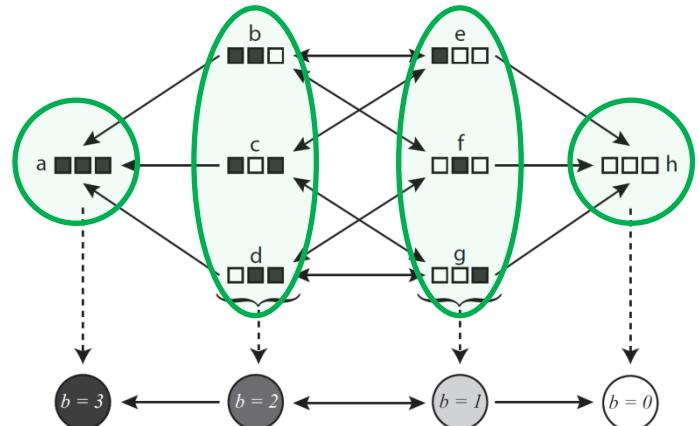


Partitioning of Interaction Matrices

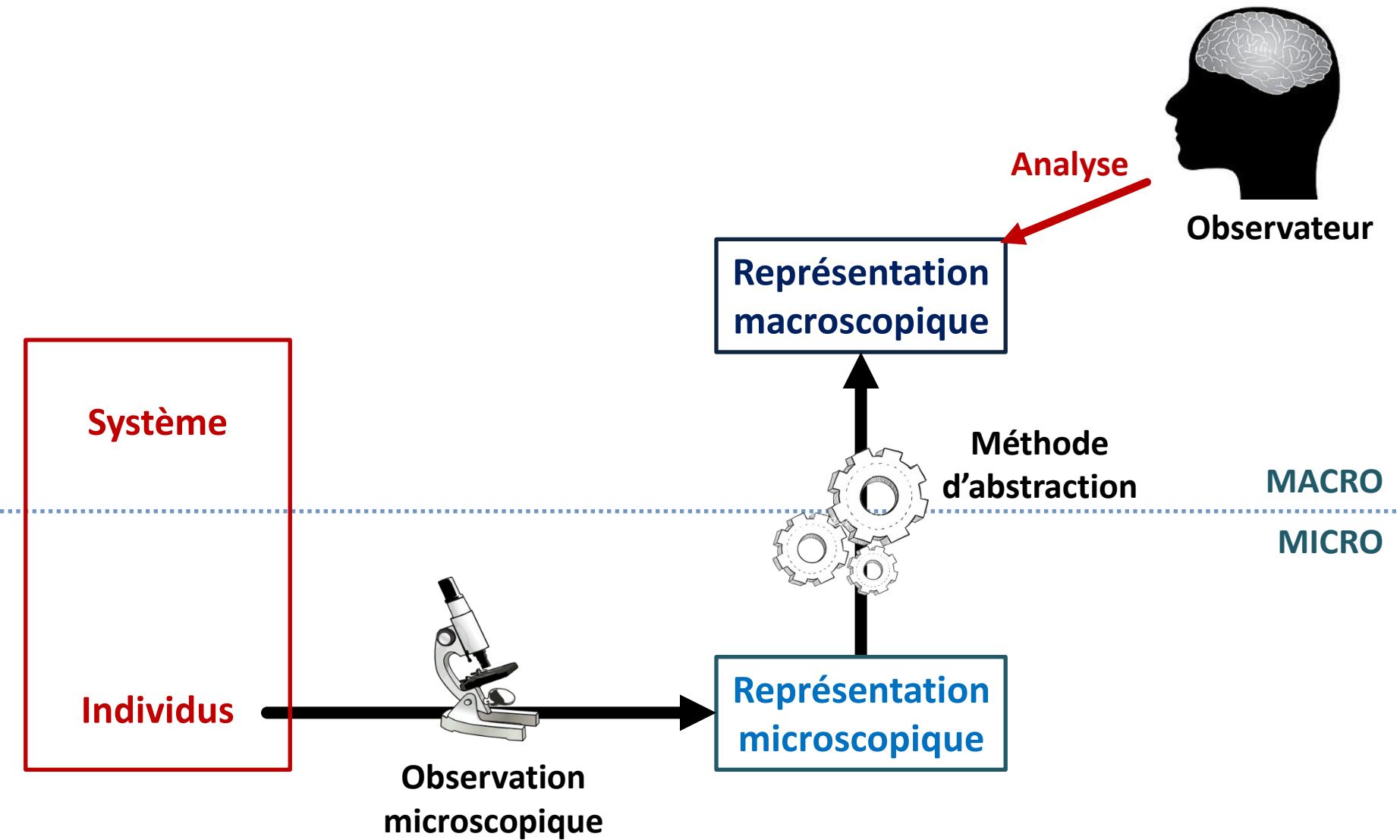
	ESP	FRA	GBR	BEL	CHE
ESP	X	12	11	10	4
FRA	14	X	12	12	5
GBR	20	11	X	6	9
BEL	15	9	6	X	5
CHE	10	16	17	9	X

Partitioning the State Space of Dynamical Systems

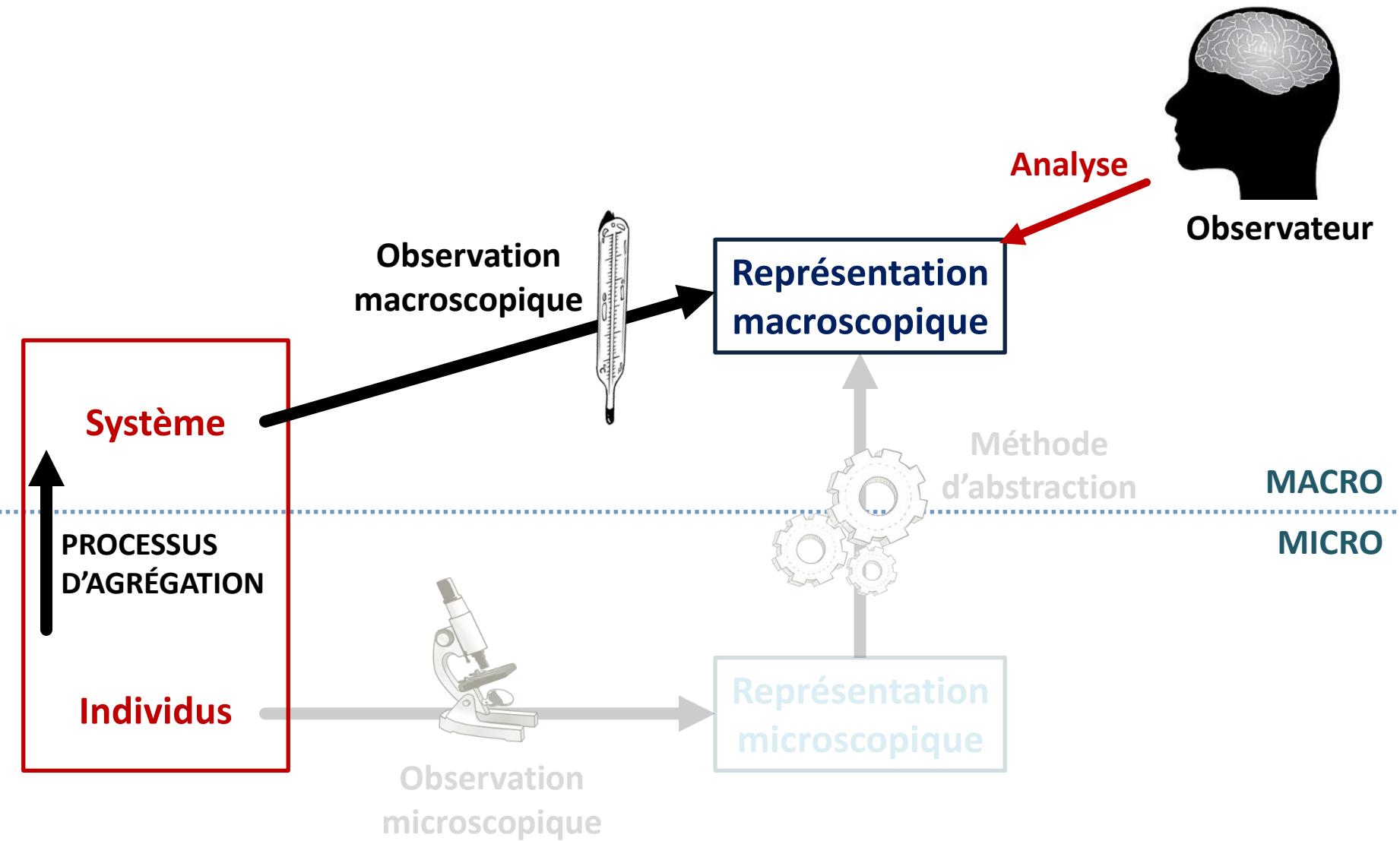
[Banisch *et al.*, 2013]



Observation macroscopique



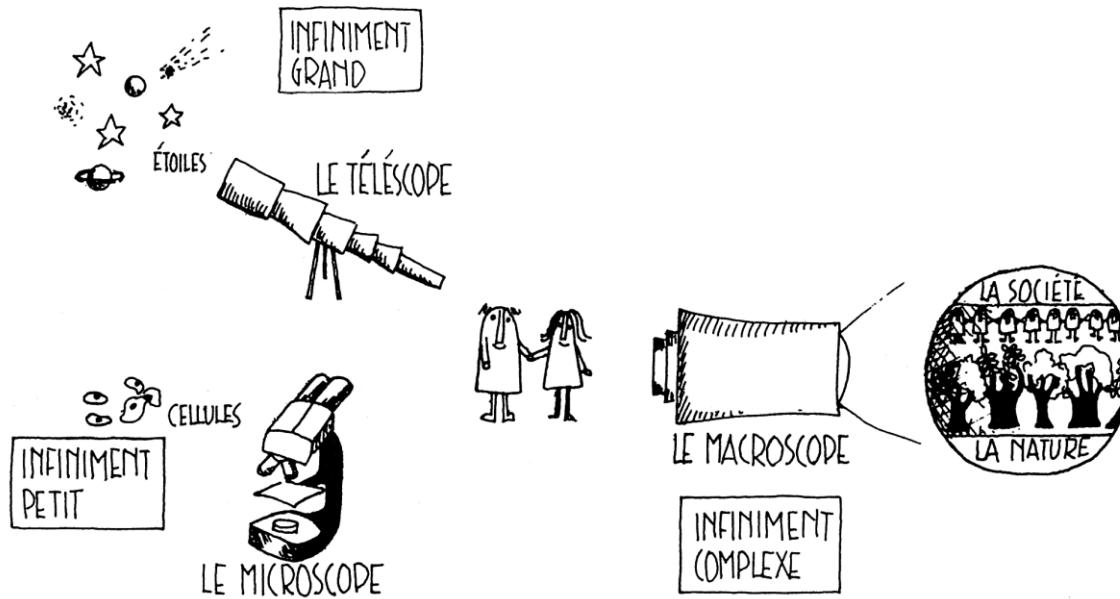
Observation macroscopique



MERCI POUR VOTRE ATTENTION

Courriel: Robin.Lamarche-Perrin@mis.mpg.de

Web: www.mis.mpg.de/jjost/members/robin-lamarche-perrin.html



« Aujourd’hui nous sommes confrontés à un autre infini : l’infiniment complexe. Mais cette fois, plus d’instrument. »

Joël de Rosnay, *Le macroscope*, 1975