

MPI MIS – WV Seminar
Leipzig, 12th November 2013

Macroscopic Analysis of Large-scale Systems

Robin Lamarche-Perrin

Reviewers: Bernard Moulin and Éric Fleury

Examiners: Salima Hassas and Brigitte Plateau

Supervisors: Yves Demazeau and Jean-Marc Vincent



The Analysis of Large-scale Systems

System

Analysis



Observer

The Analysis of Large-scale Systems

System

Large-scale
Distributed

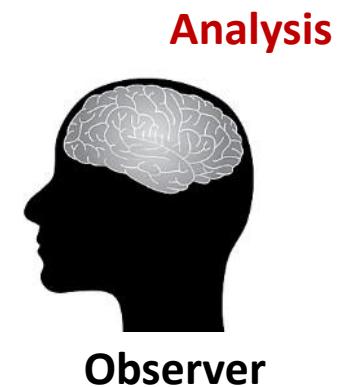
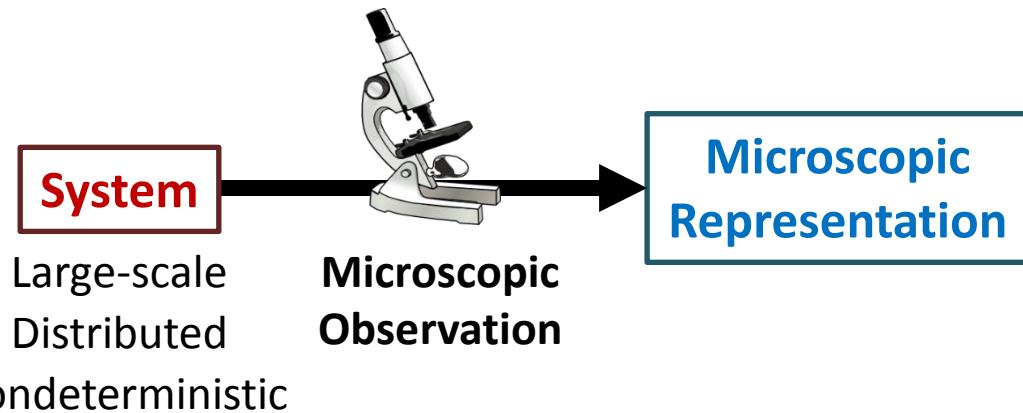
Nondeterministic

Analysis

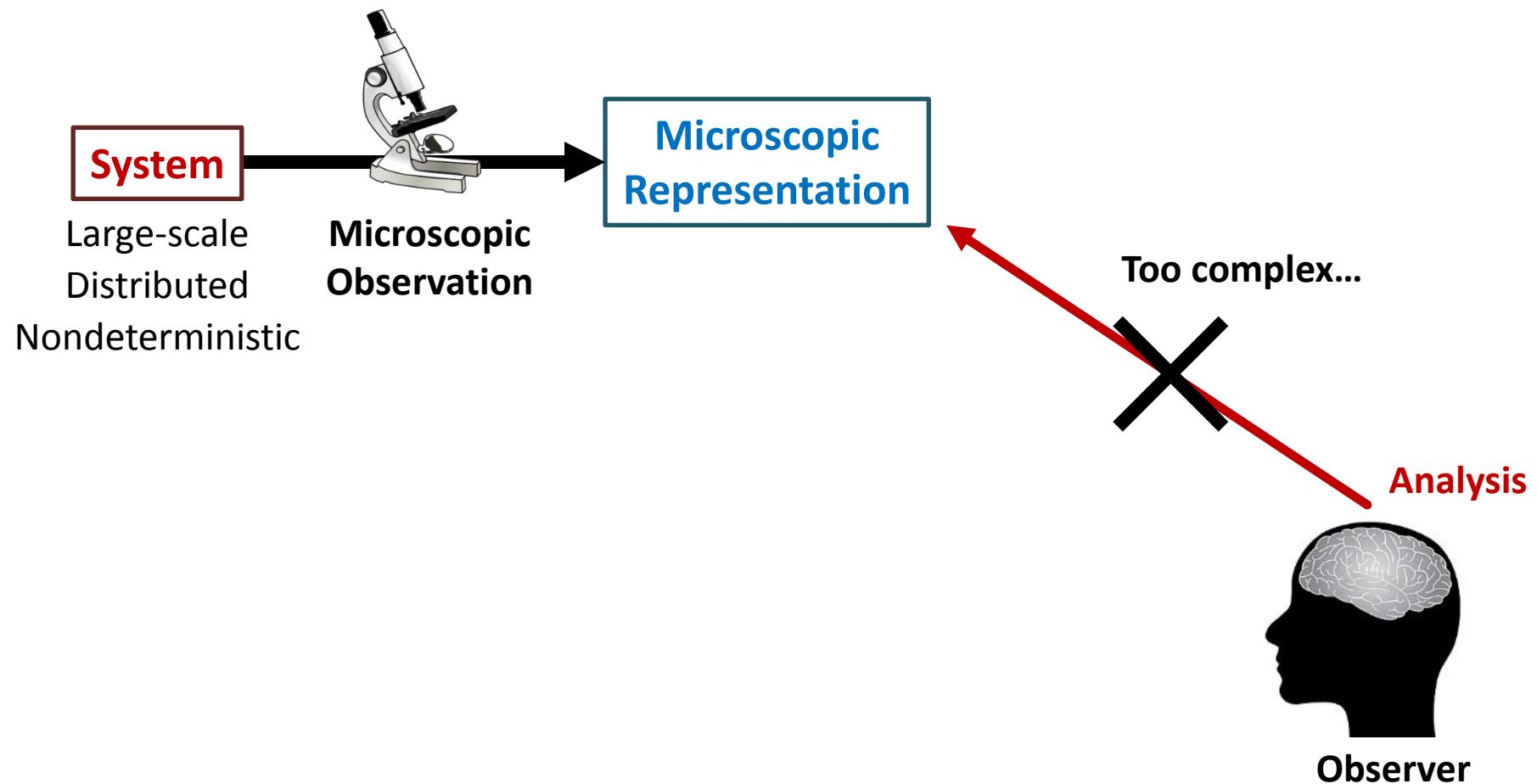


Observer

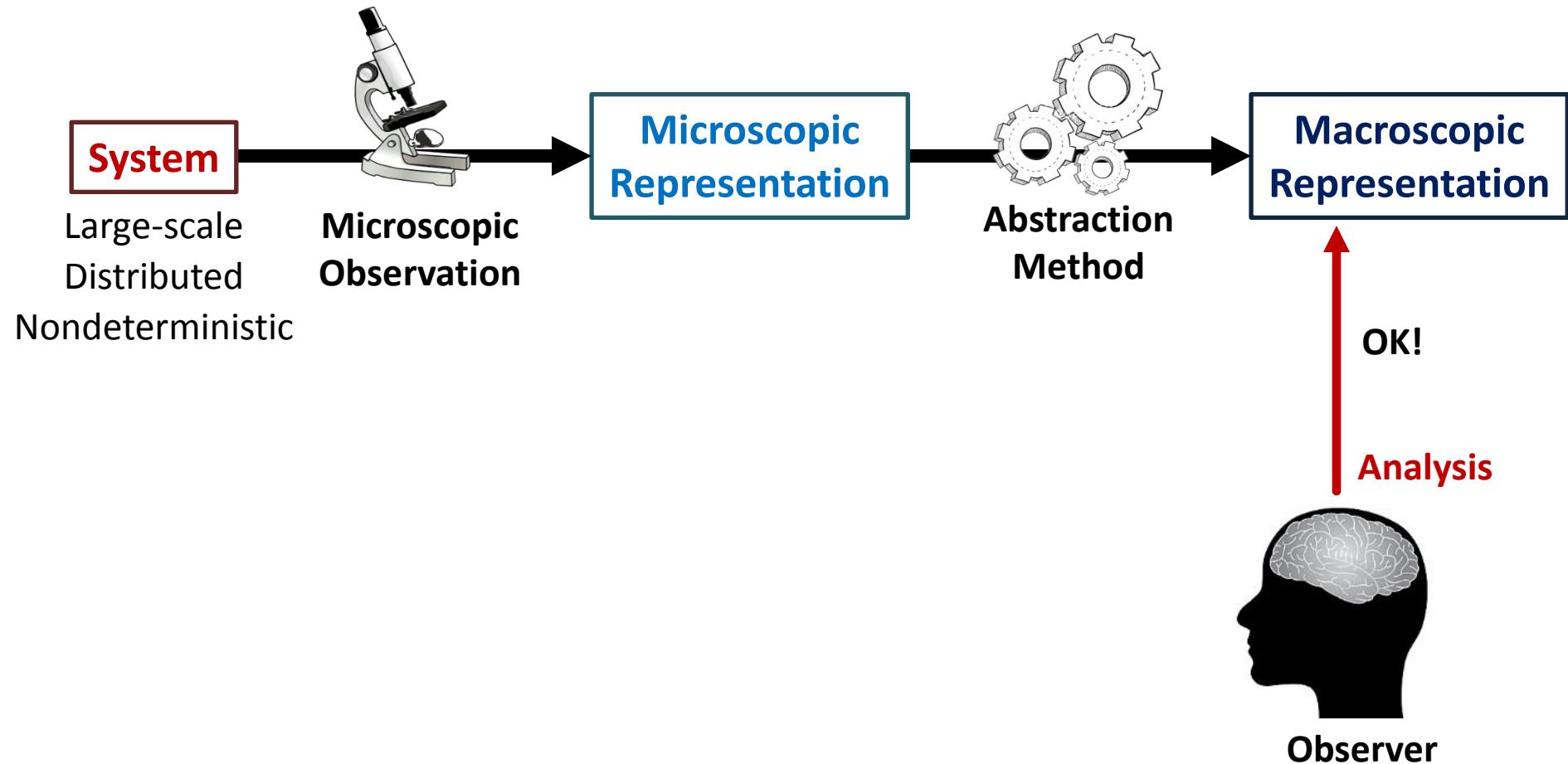
The Analysis of Large-scale Systems



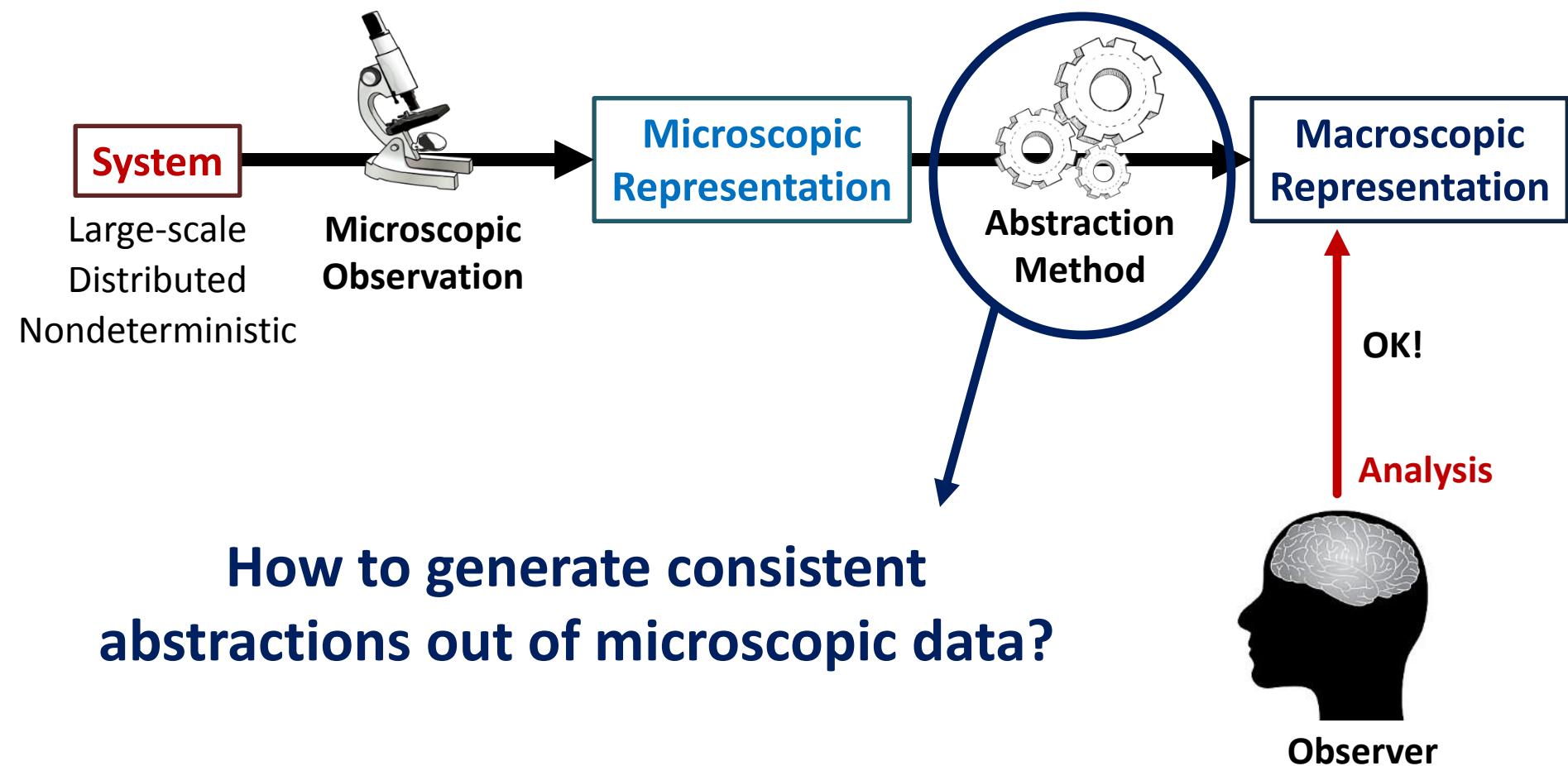
The Analysis of Large-scale Systems



The Analysis of Large-scale Systems



The Analysis of Large-scale Systems

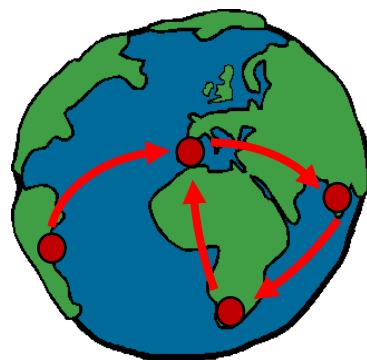


Analysis of International Relations



Geographer

International
System



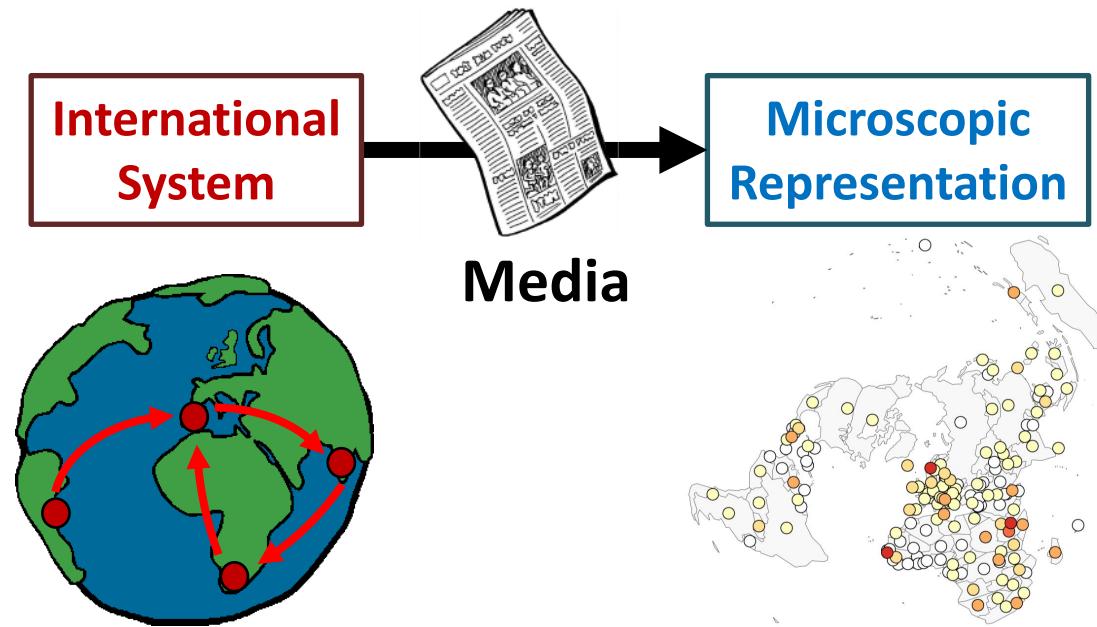
Analysis of International Relations

Hypothesis: media constitute an adequate instrument to observe the national level

[Grasland *et al.*, 2011]



Geographer



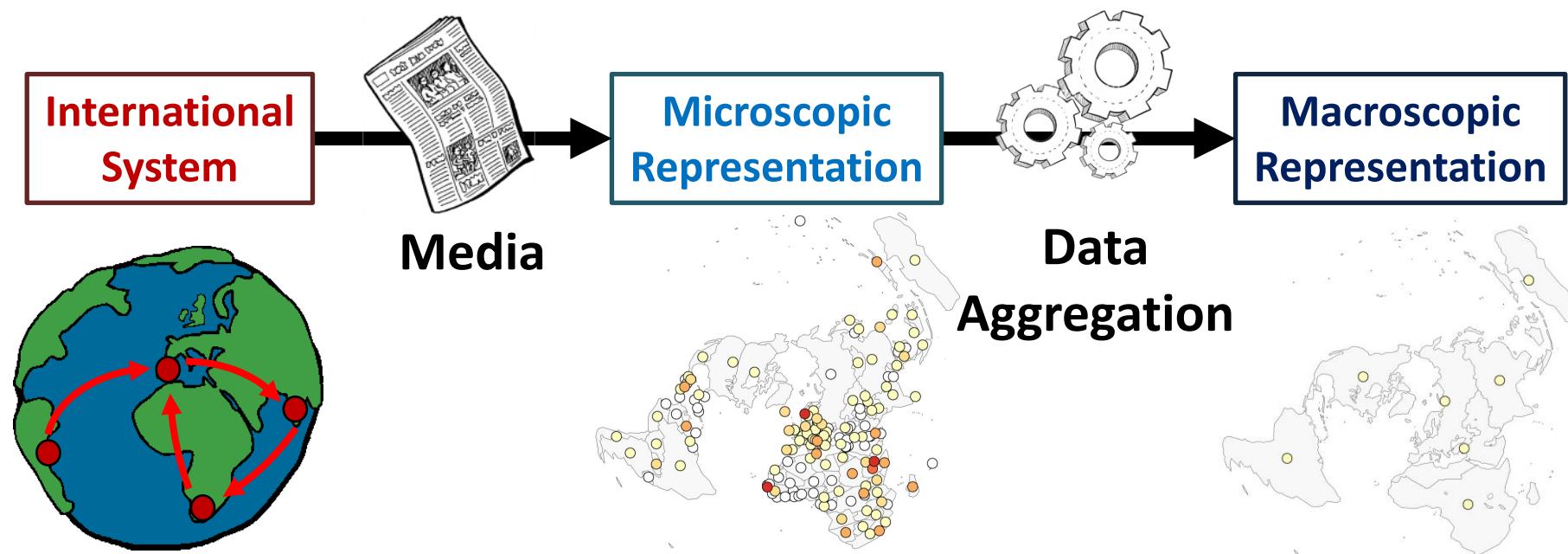
Analysis of International Relations

Hypothesis: media constitute an adequate instrument to observe the national level

[Grasland *et al.*, 2011]



Geographer



Data from Print Media

THE GUARDIAN

Paper 1



“Japan”

THE TIMES OF INDIA

Paper 2



“Madrid”

Paper 3



“French”

“Spain”

The GEOMEDIA Database
(ANR CORPUS GUI-AAP-04)

150 newspapers

1,944,000 papers

GEOGRAPHIC INFORMATION
193 countries (UN members)

Data from Print Media

THE GUARDIAN



“Japan”

THE TIMES OF INDIA



Paper 2

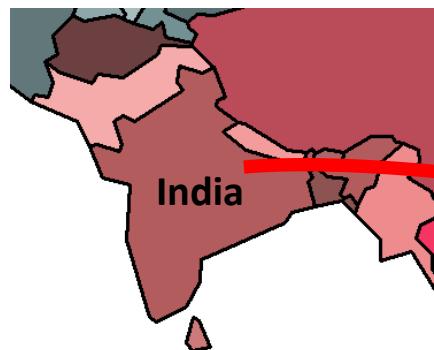
“Madrid”



Paper 3

“French”

“Spain”



The GEOMEDIA Database
(ANR CORPUS GUI-AAP-04)

150 newspapers

1,944,000 papers

GEOGRAPHIC INFORMATION
193 countries (UN members)

Data from Print Media

THE GUARDIAN

Paper 1



“Japan”

30th May 2011

THE TIMES OF INDIA

Paper 2



“Madrid”

30th May 2011

Paper 3



“French”

19th July 2012

“Spain”

The GEOMEDIA Database
(ANR CORPUS GUI-AAP-04)

150 newspapers

1,944,000 papers

GEOGRAPHIC INFORMATION
193 countries (UN members)

TEMPORAL INFORMATION

889 days / 127 weeks
(from the 3rd May 2011 to today)

Microscopic Representation of the International System

Newspaper LE MONDE

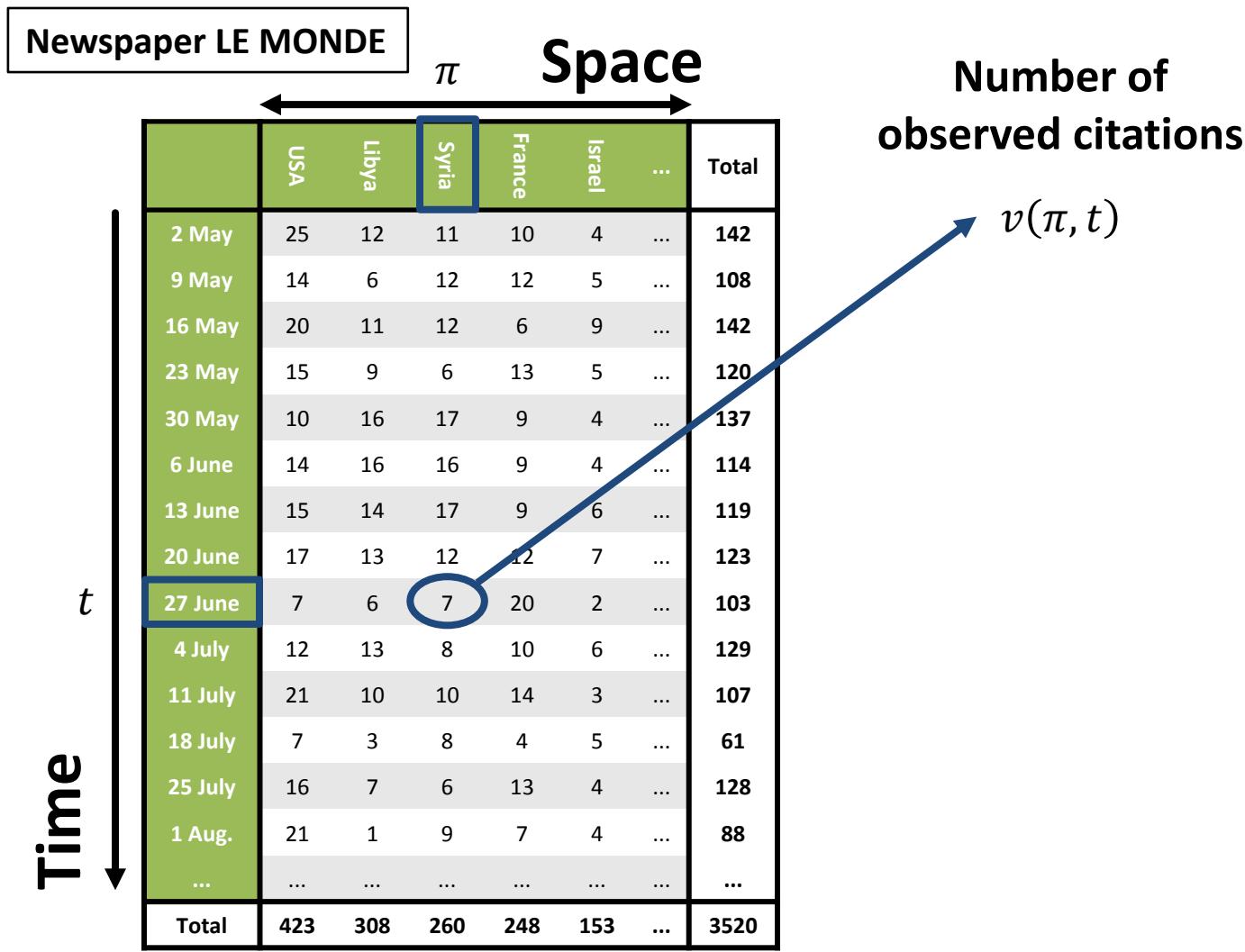
Space



Time

	USA	Libya	Syria	France	Israel	...	Total
2 May	25	12	11	10	4	...	142
9 May	14	6	12	12	5	...	108
16 May	20	11	12	6	9	...	142
23 May	15	9	6	13	5	...	120
30 May	10	16	17	9	4	...	137
6 June	14	16	16	9	4	...	114
13 June	15	14	17	9	6	...	119
20 June	17	13	12	12	7	...	123
27 June	7	6	7	20	2	...	103
4 July	12	13	8	10	6	...	129
11 July	21	10	10	14	3	...	107
18 July	7	3	8	4	5	...	61
25 July	16	7	6	13	4	...	128
1 Aug.	21	1	9	7	4	...	88
...
Total	423	308	260	248	153	...	3520

Microscopic Representation of the International System



Microscopic Representation of the International System

Newspaper LE MONDE

π

Space

	USA	Libya	Syria	France	Israel	...	Total
2 May	25	12	11	10	4	...	142
9 May	14	6	12	12	5	...	108
16 May	20	11	12	6	9	...	142
23 May	15	9	6	13	5	...	120
30 May	10	16	17	9	4	...	137
6 June	14	16	16	9	4	...	114
13 June	15	14	17	9	6	...	119
20 June	17	13	12	12	7	...	123
27 June	7	6	7	20	2	...	103
4 July	12	13	8	10	6	...	129
11 July	21	10	10	14	3	...	107
18 July	7	3	8	4	5	...	61
25 July	16	7	6	13	4	...	128
1 Aug.	21	1	9	7	4	...	88
...
Total	423	308	260	248	153	...	3520

t

Time

Number of observed citations

$$v(\pi, t)$$

..

$$v(\pi, .)$$

$$v(., .)$$

Number of expected citations

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

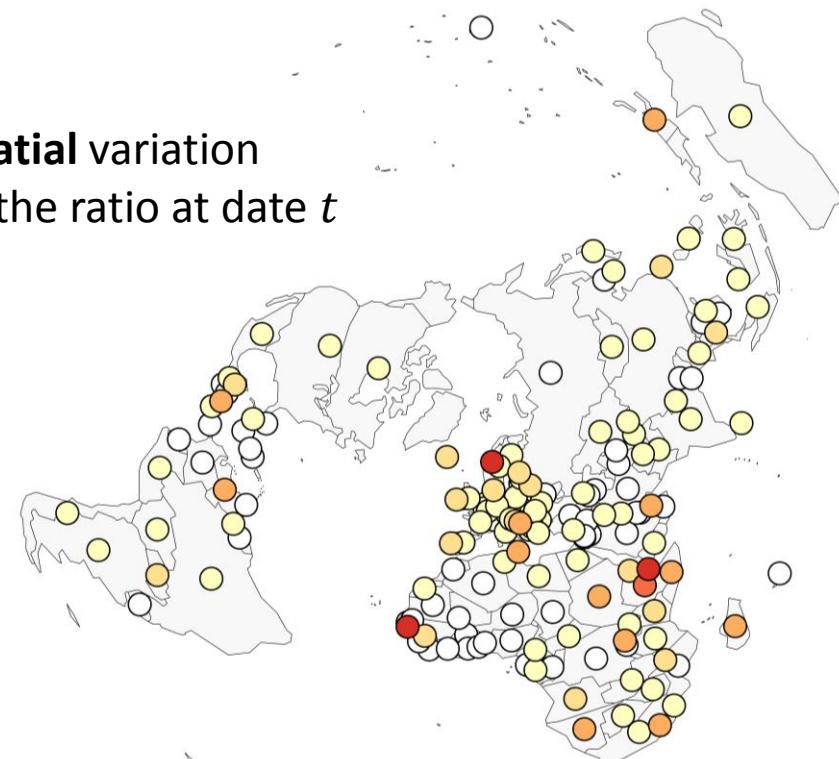
Geographical Representation

Space

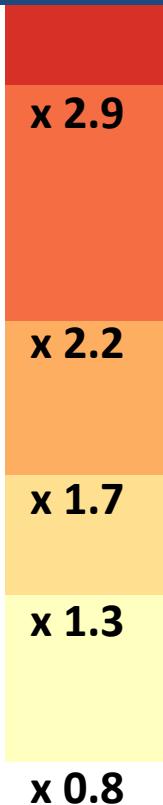
	USA	Libya	Syria	France	Israel	...	Total
2 May	25	12	11	10	4	...	142
9 May	14	6	12	12	5	...	108
16 May	20	11	12	6	9	...	142
23 May	15	9	6	13	5	...	120
30 May	10	16	17	9	4	...	137
6 June	14	16	16	9	4	...	114
13 June	15	14	17	9	6	...	119
20 June	17	13	12	12	7	...	123
27 June	7	6	7	20	2	...	103
4 July	12	13	8	10	6	...	129
11 July	21	10	10	14	3	...	107
18 July	7	3	8	4	5	...	61
25 July	16	7	6	13	4	...	128
1 Aug.	21	1	9	7	4	...	88
...
Total	423	308	260	248	153	...	3520

Observed-to-expected
ratio of citation number

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

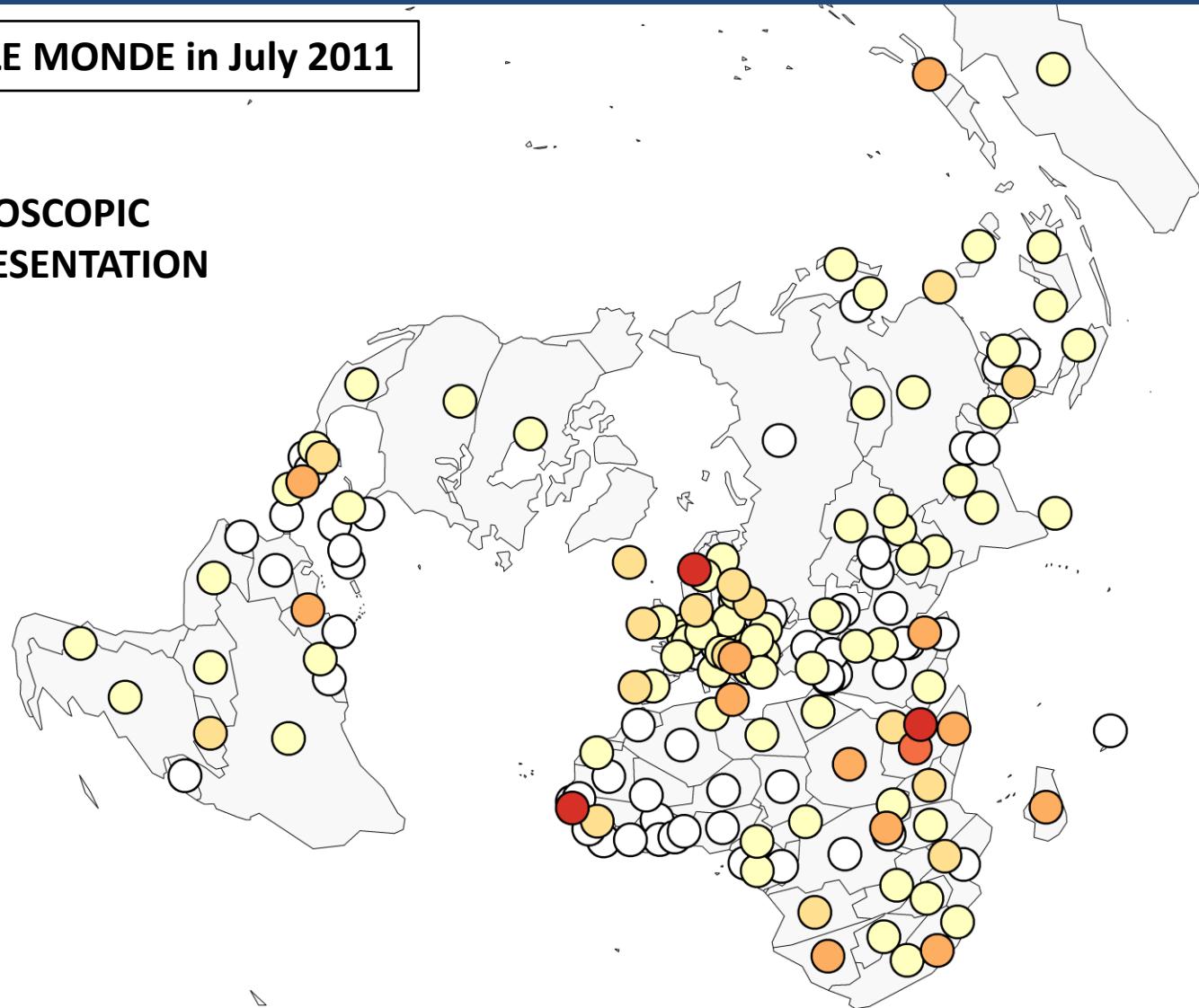


Detection of Media Events

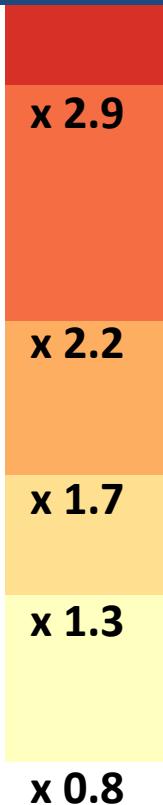


Newspaper LE MONDE in July 2011

MICROSCOPIC
REPRESENTATION

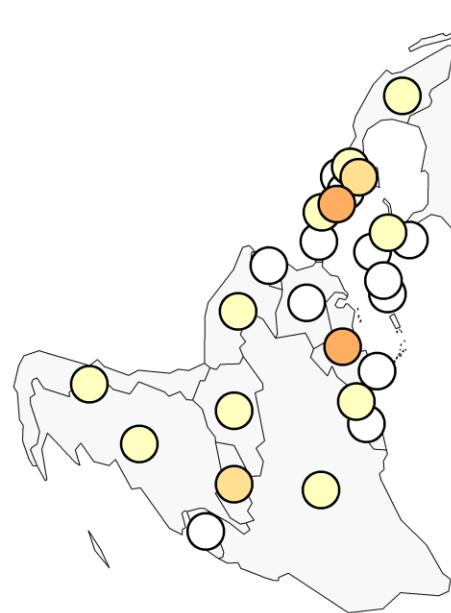


Detection of Media Events

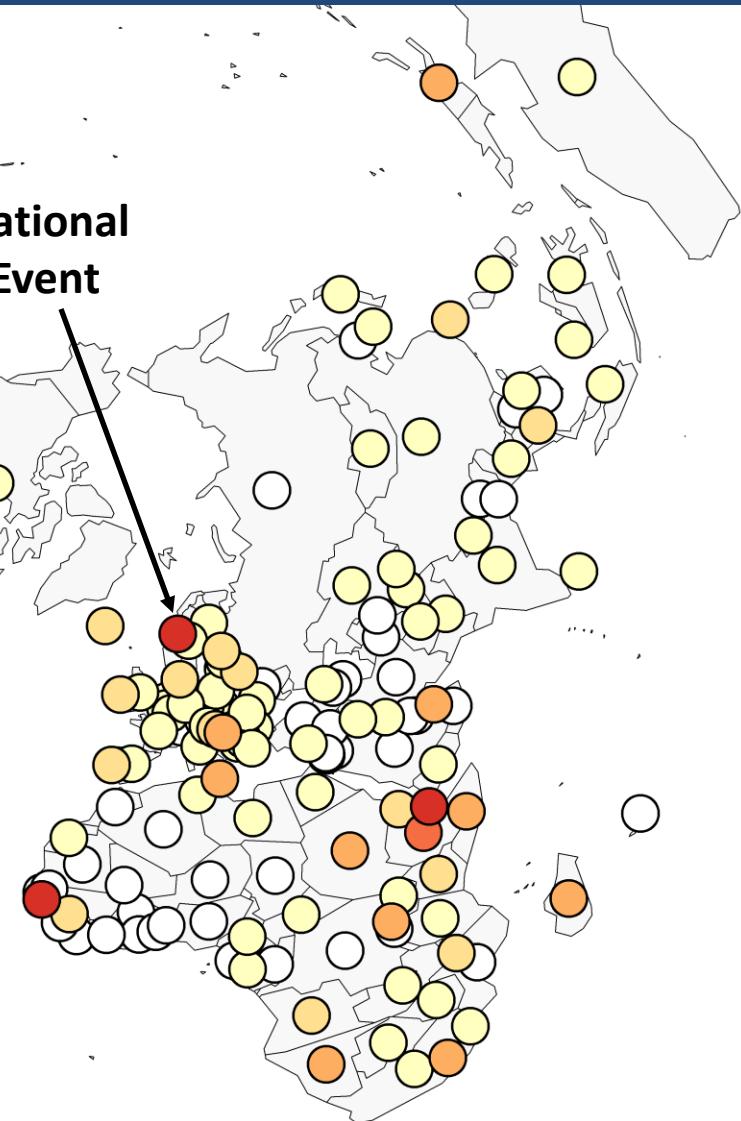


Newspaper LE MONDE in July 2011

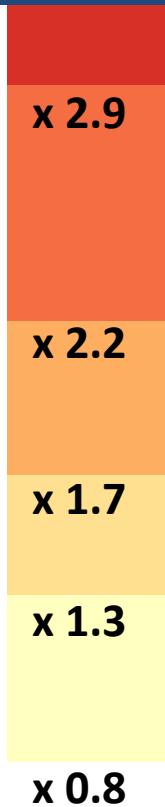
MICROSCOPIC
REPRESENTATION



National
Event

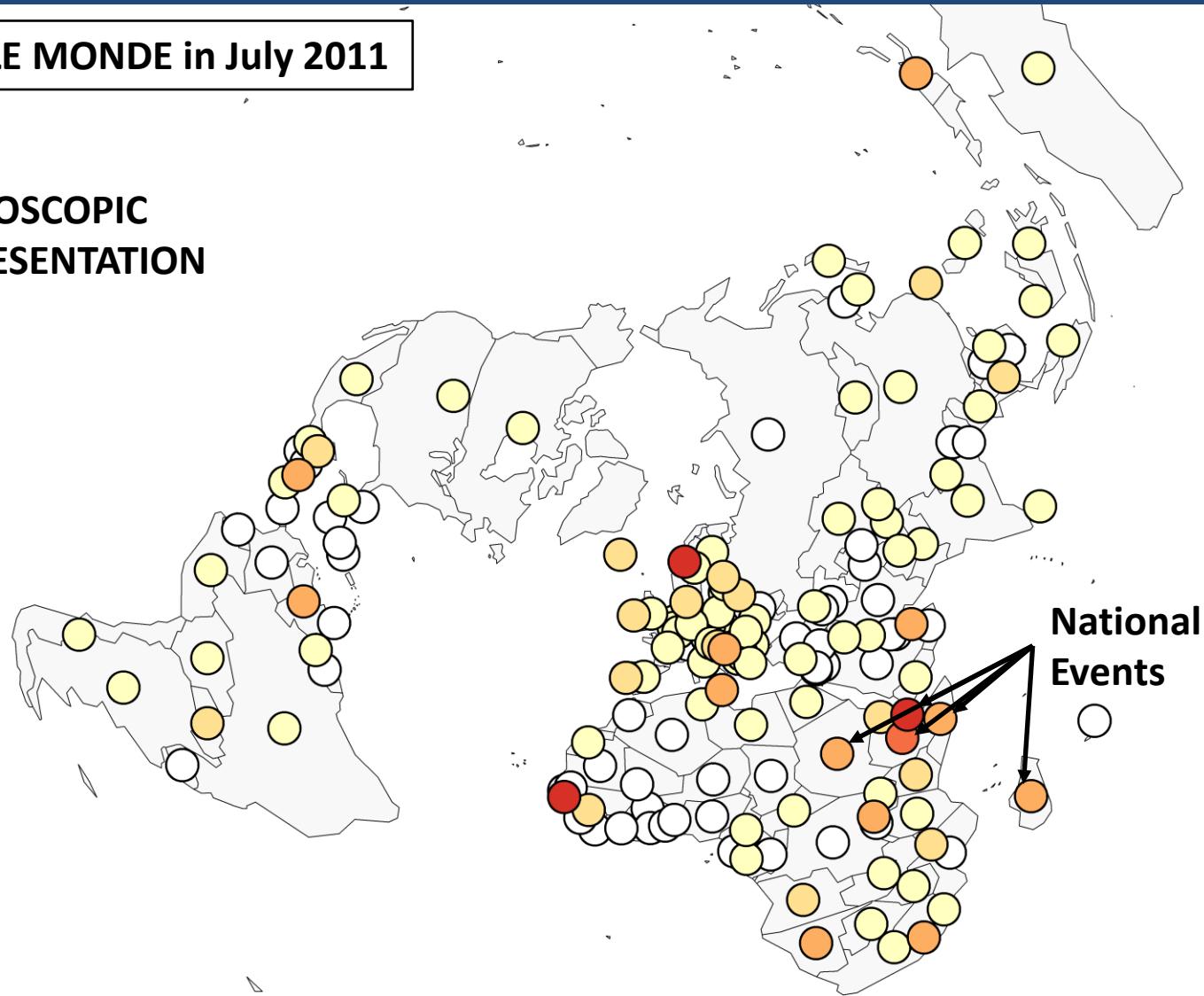


Detection of Media Events

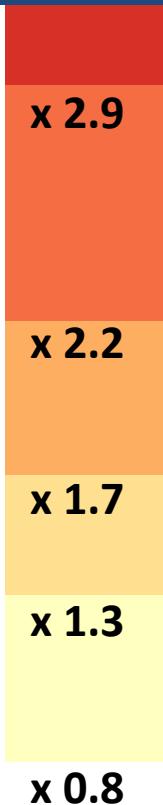


Newspaper LE MONDE in July 2011

MICROSCOPIC
REPRESENTATION

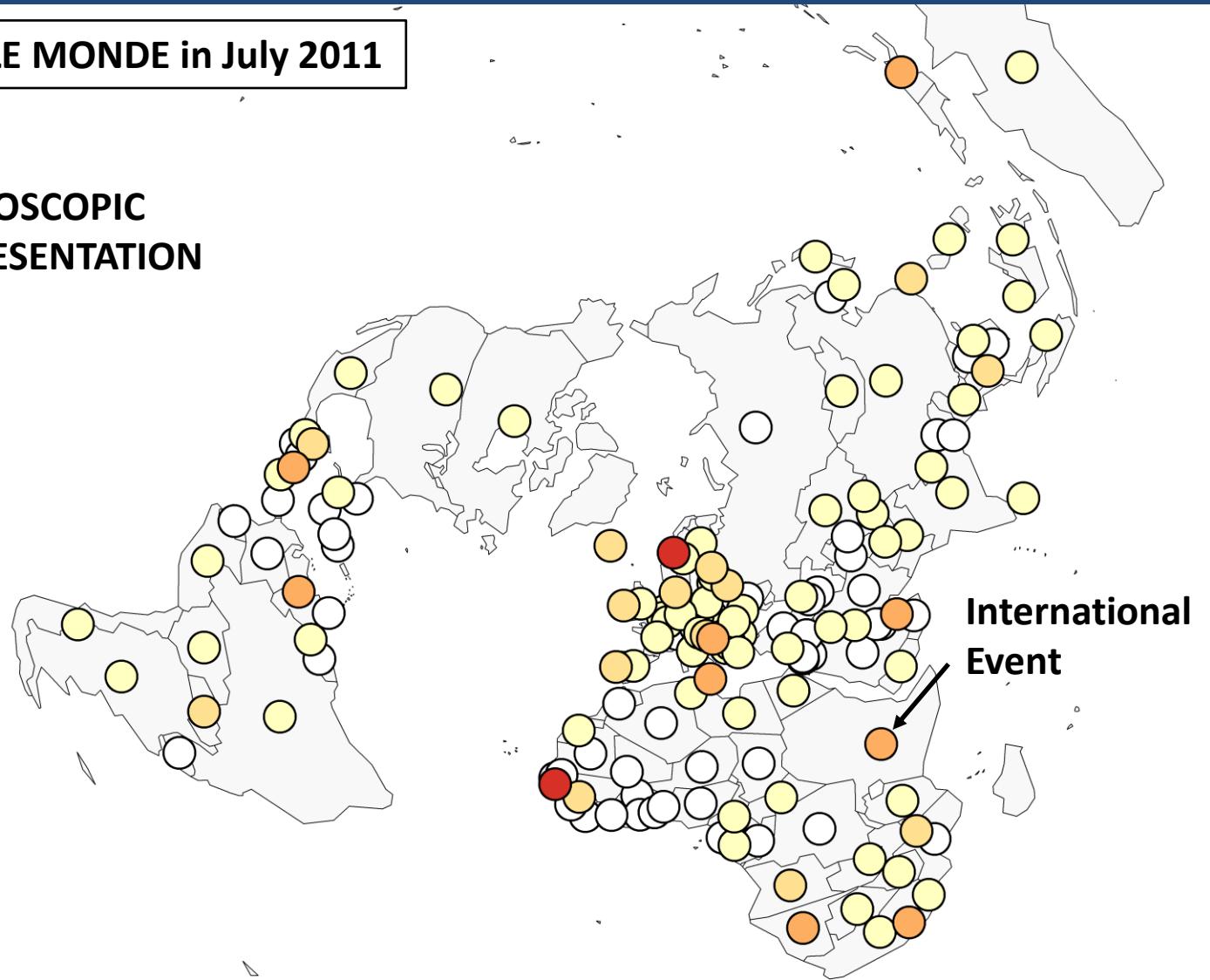


Detection of Media Events



Newspaper LE MONDE in July 2011

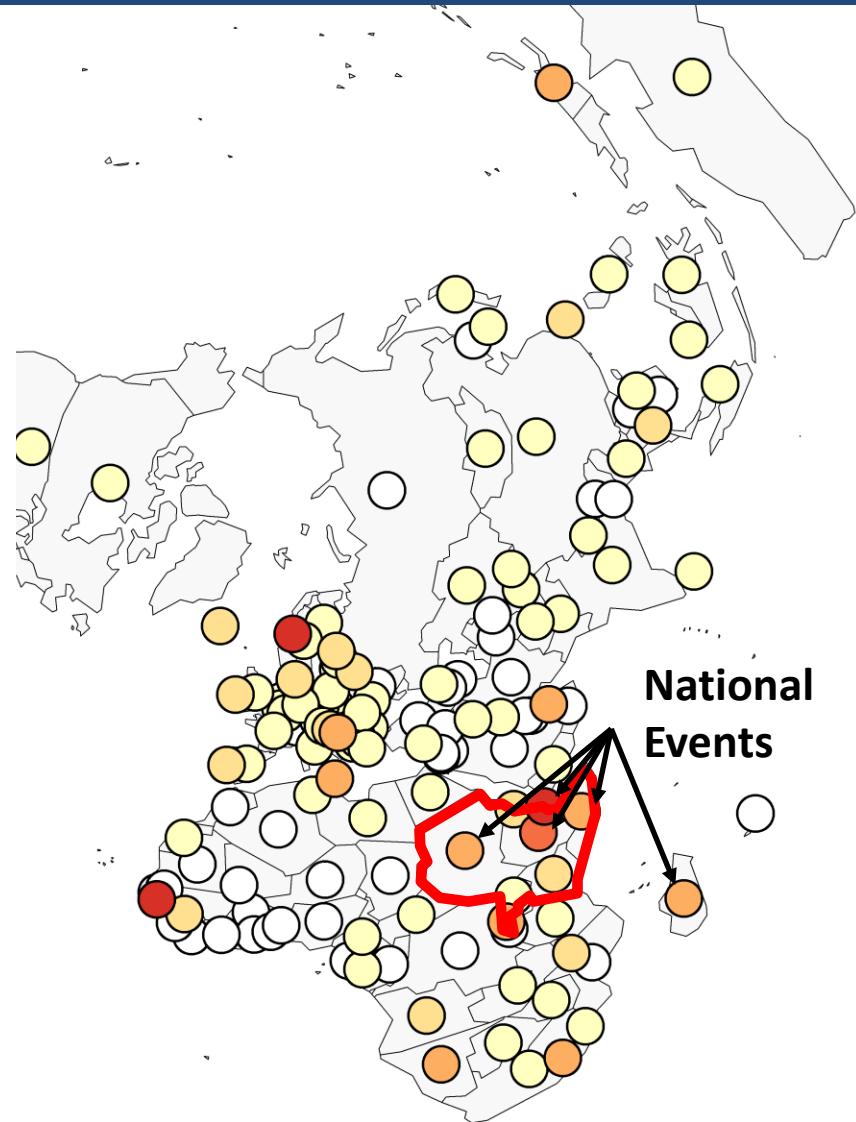
MICROSCOPIC
REPRESENTATION



Data Aggregation

Space

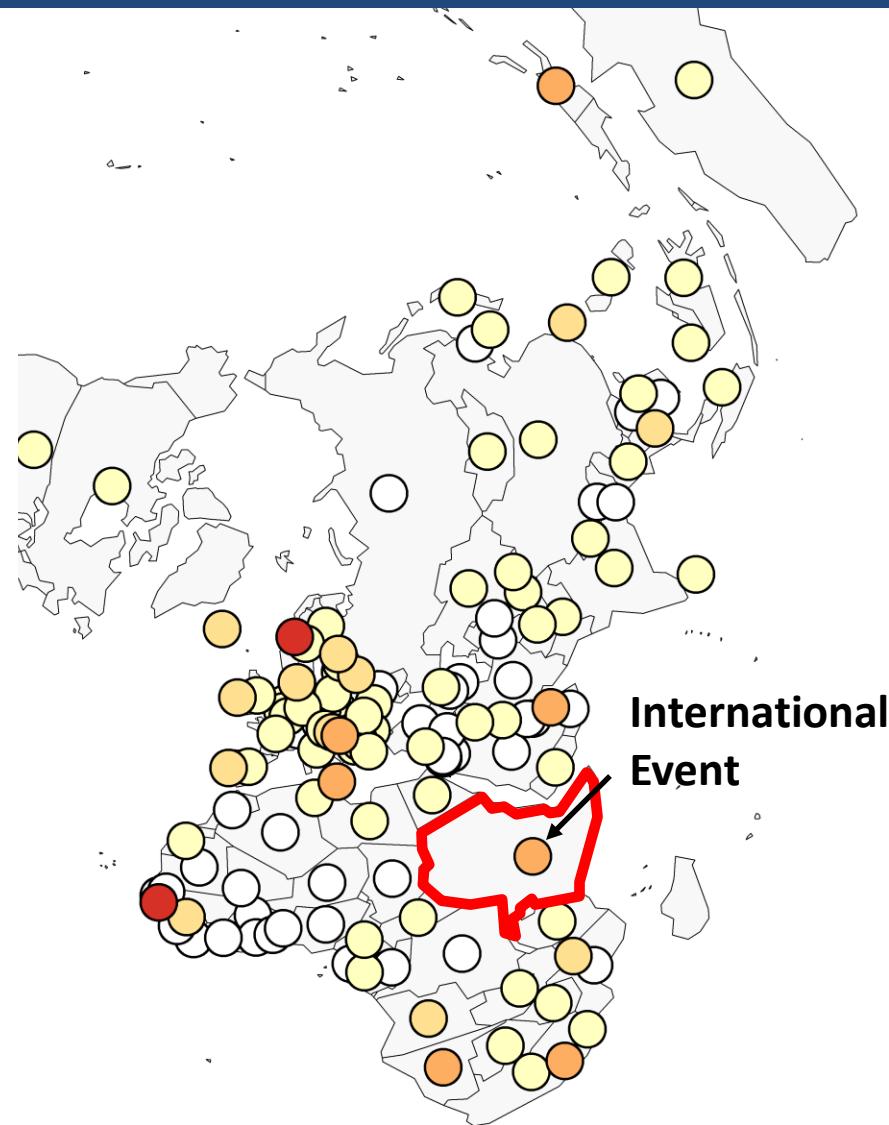
	USA	Libya	Syria	France	Israel	...	Total
2 May	25	12	11	10	4	...	142
9 May	14	6	12	12	5	...	108
16 May	20	11	12	6	9	...	142
23 May	15	9	6	13	5	...	120
30 May	10	16	17	9	4	...	137
6 June	14	16	16	9	4	...	114
13 June	15	14	17	9	6	...	119
20 June	17	13	12	12	7	...	123
27 June	7	6	7	20	2	...	103
4 July	12	13	8	10	6	...	129
11 July	21	10	10	14	3	...	107
18 July	7	3	8	4	5	...	61
25 July	16	7	6	13	4	...	128
1 Aug.	21	1	9	7	4	...	88
...
Total	423	308	260	248	153	...	3520



Data Aggregation

Space

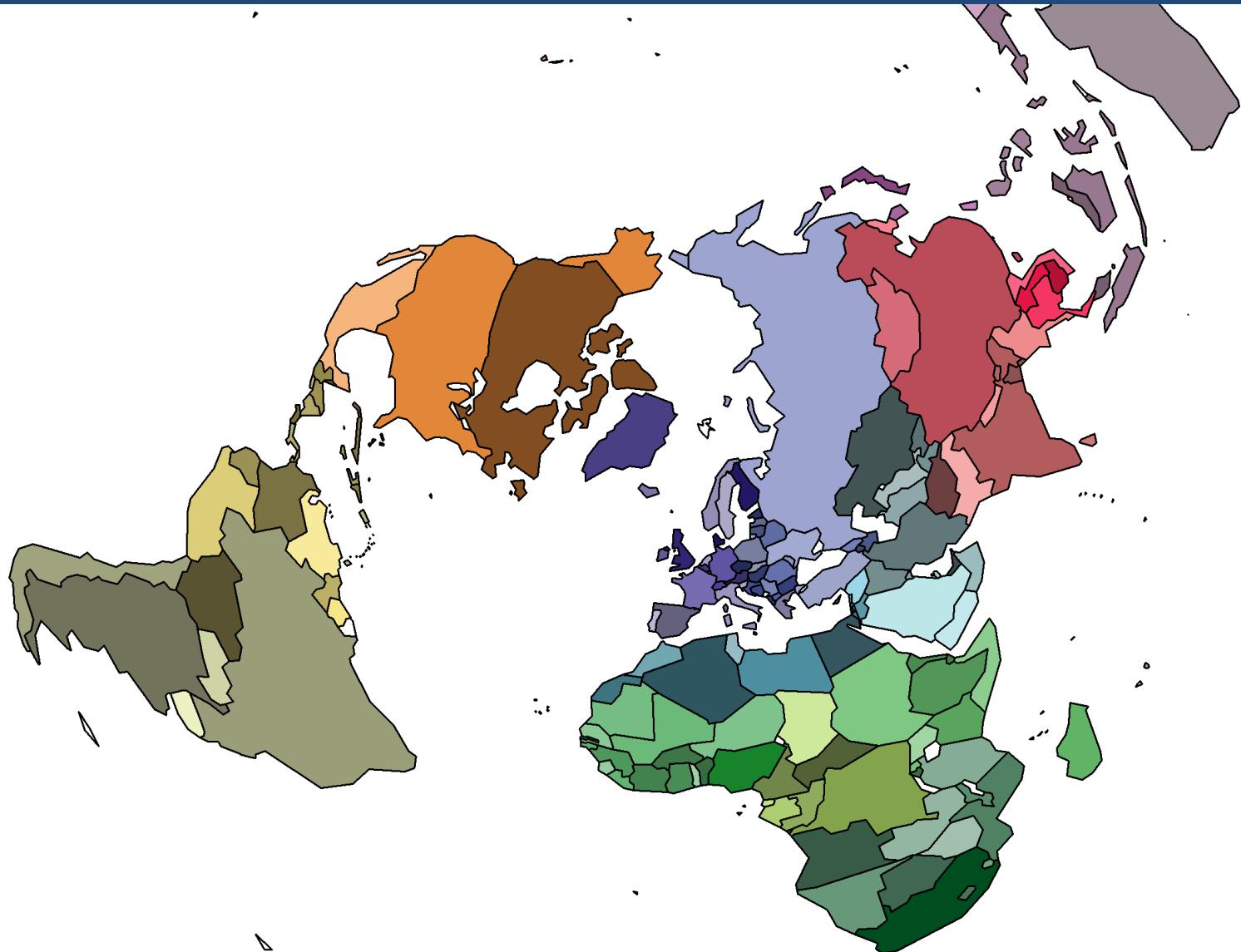
	USA	Aggregate	Israel	...	Total
2 May	25	13+11+10	4	...	142
9 May	14	6+12+12	5	...	108
16 May	20	11+12+6	9	...	142
23 May	15	9+6+13	5	...	120
30 May	10	16+17+9	4	...	137
6 June	14	16+16+9	4	...	114
13 June	15	14+17+9	6	...	119
20 June	17	13+12+12	7	...	123
27 June	7	6+7+20	2	...	103
4 July	12	13+8+10	6	...	129
11 July	21	10+10+14	3	...	107
18 July	7	3+8+4	5	...	61
25 July	16	7+6+13	4	...	128
1 Aug.	21	1+9+7	4	...	88
...
Total	423	308+260+248	153	...	3520



P1: The Semantics of Geographical Aggregates



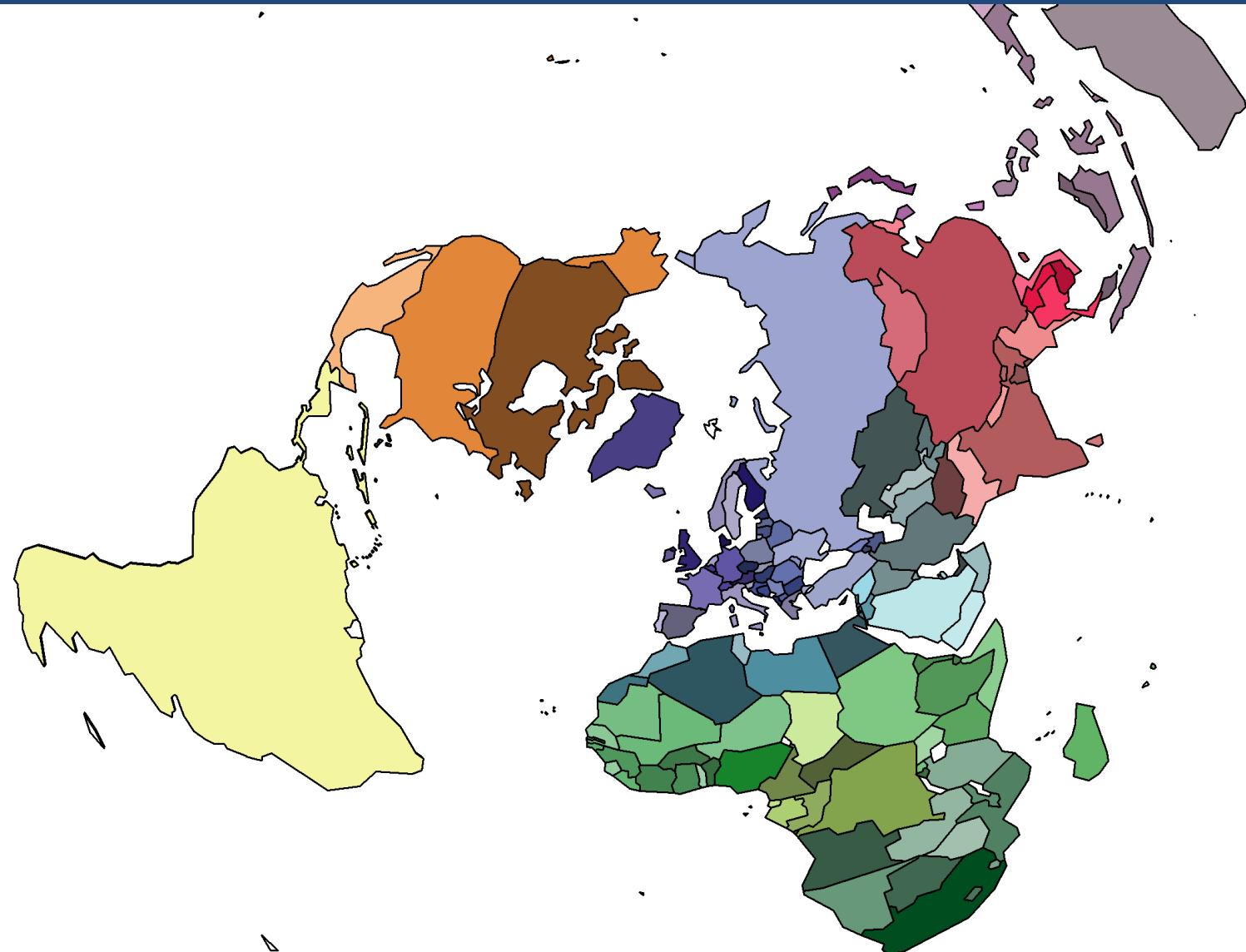
Geographer



P1: The Semantics of Geographical Aggregates



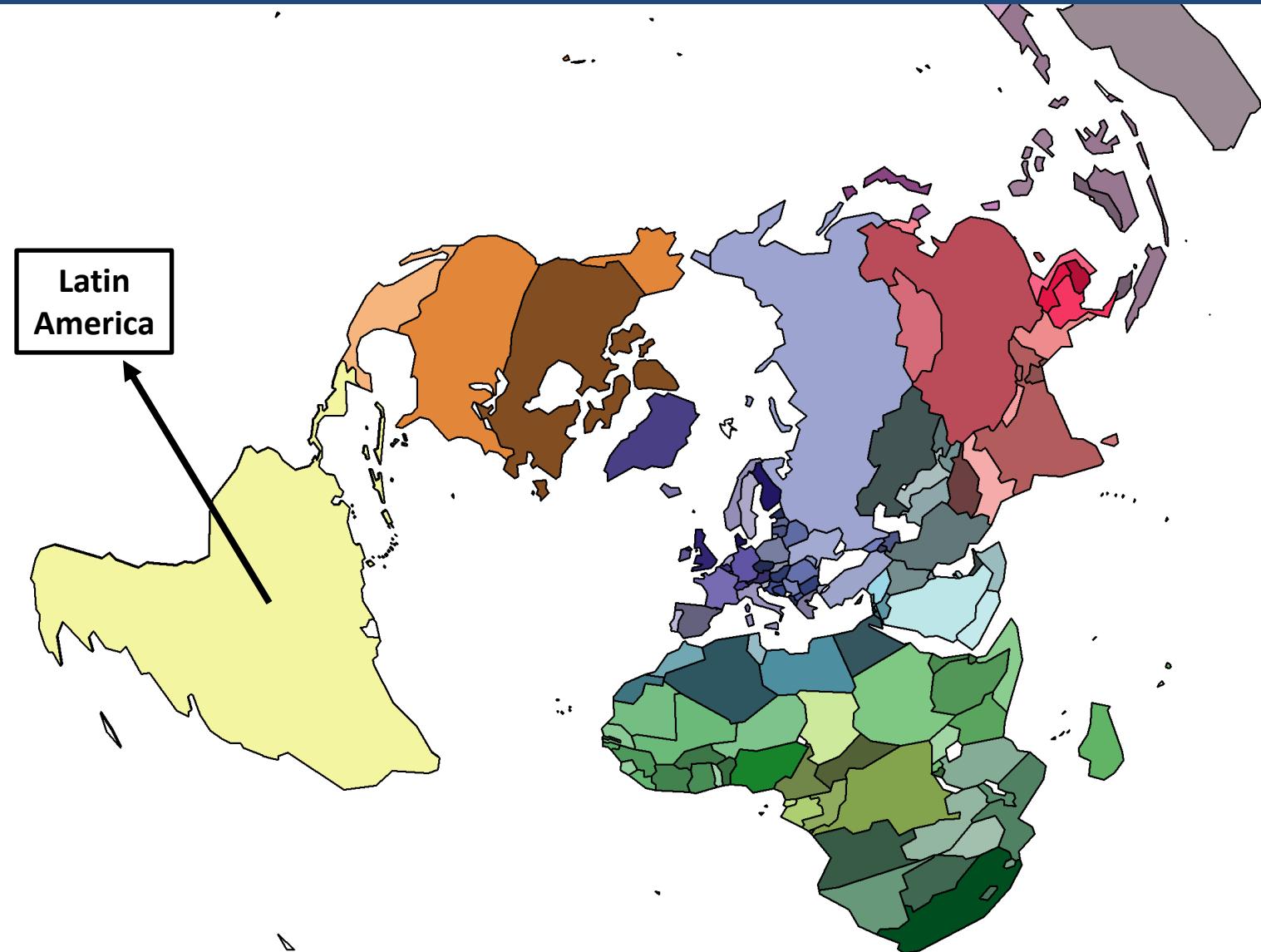
Geographer



P1: The Semantics of Geographical Aggregates



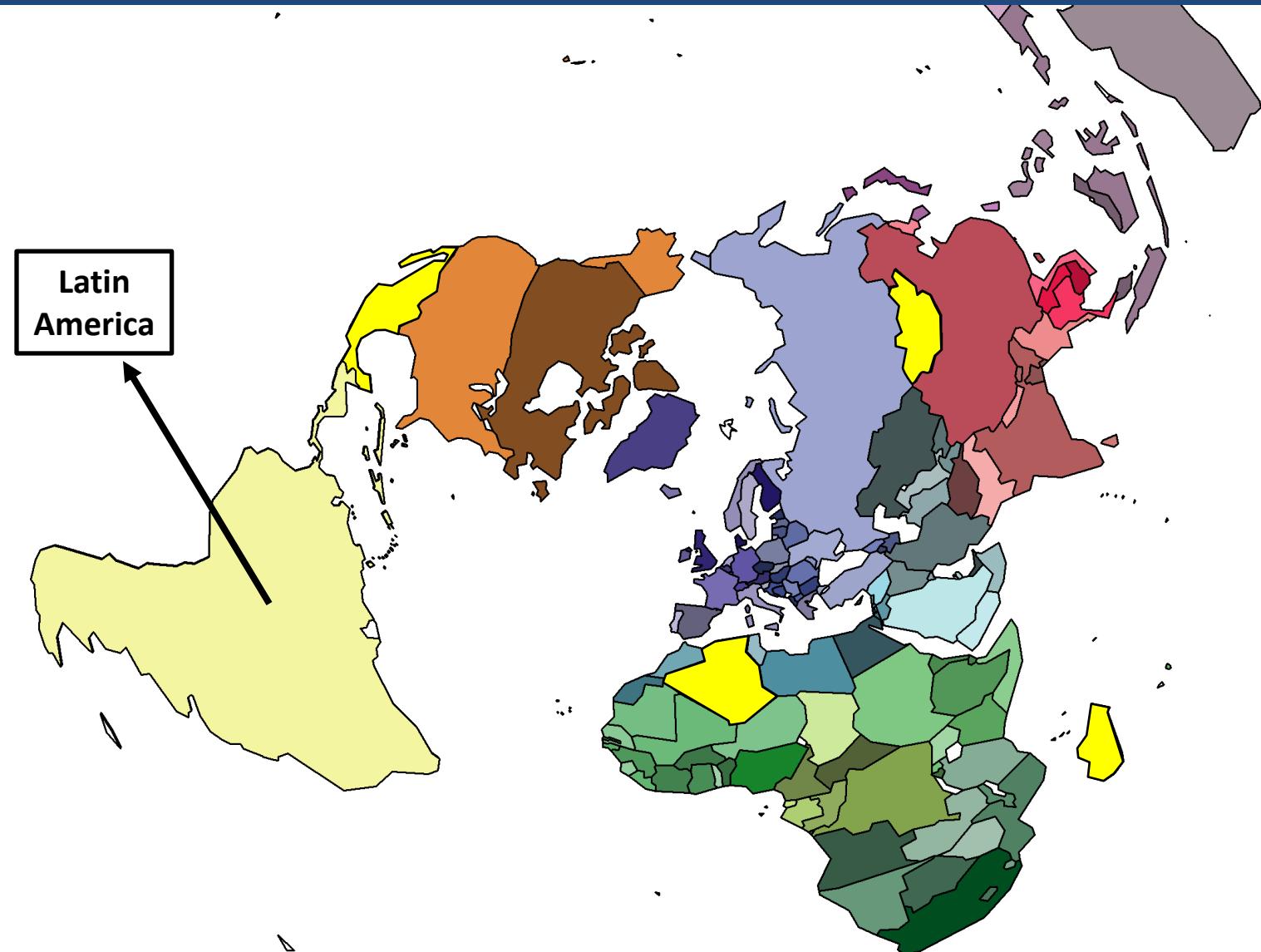
Geographer



P1: The Semantics of Geographical Aggregates



Geographer



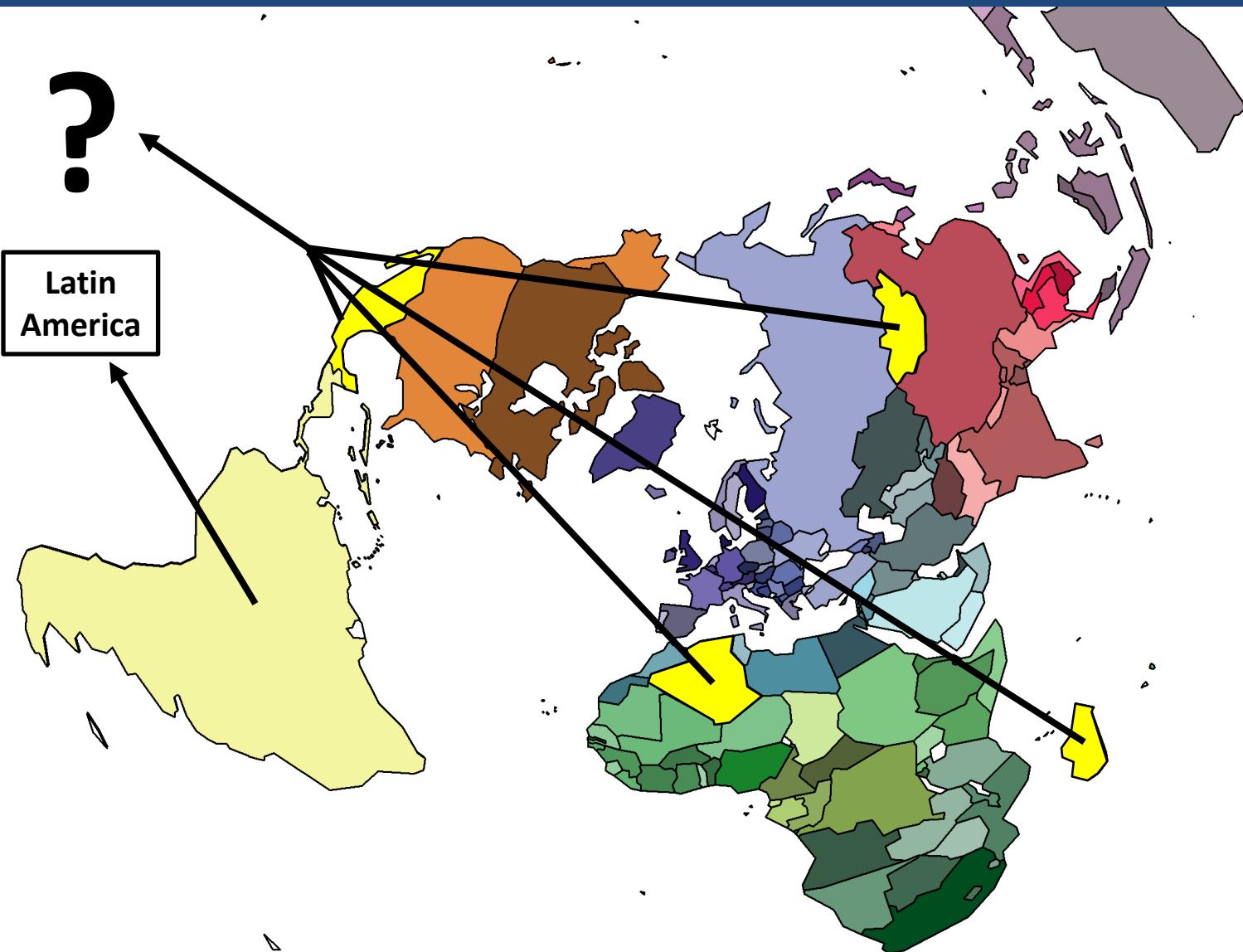
P1: The Semantics of Geographical Aggregates



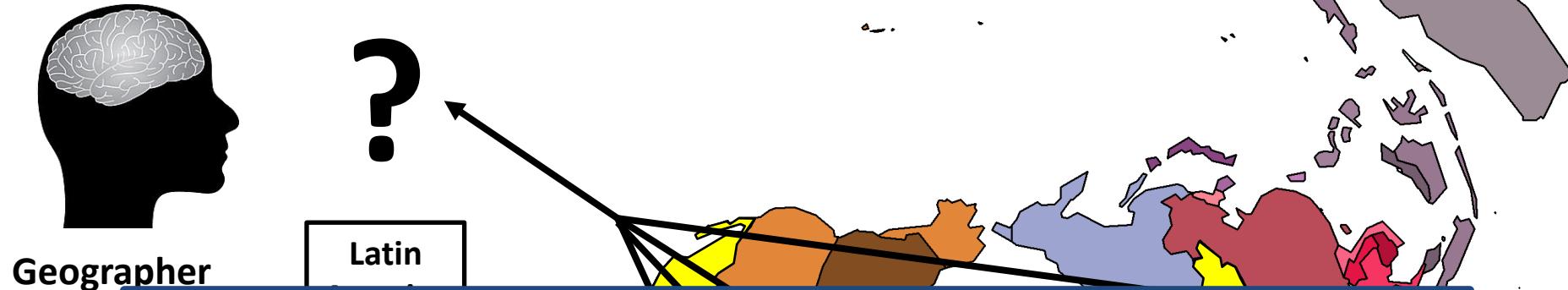
Geographer

?

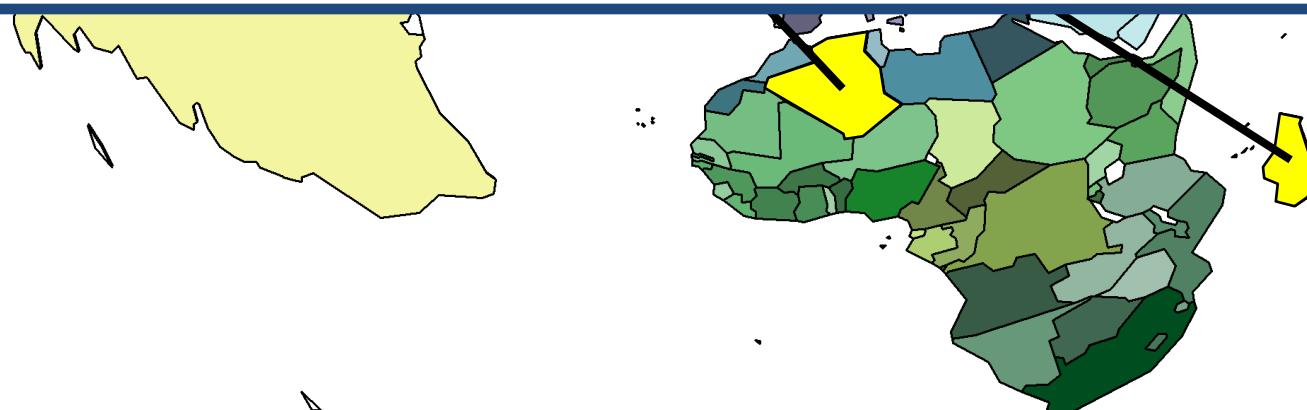
Latin
America



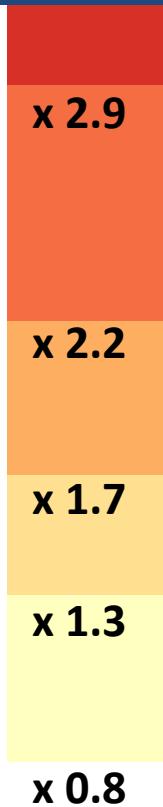
P1: The Semantics of Geographical Aggregates



Problem 1: How to generate abstractions
that are **meaningful** for the observer?

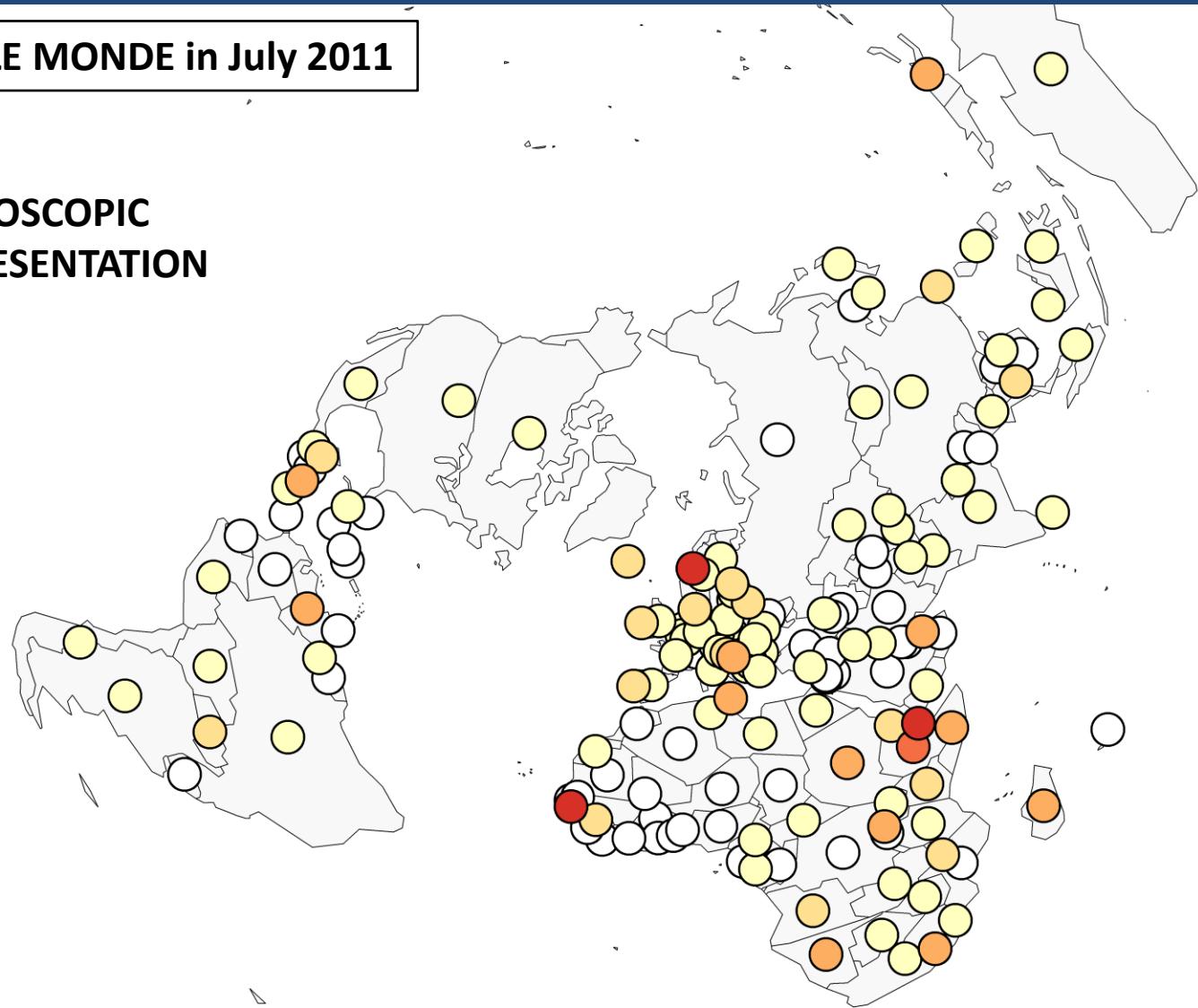


P2: The Levels of Representation

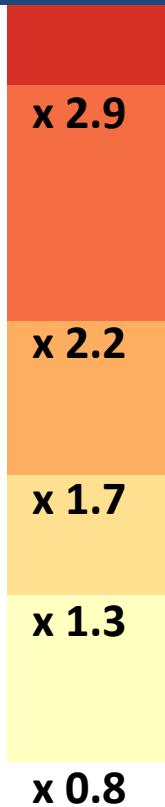


Newspaper LE MONDE in July 2011

MICROSCOPIC
REPRESENTATION

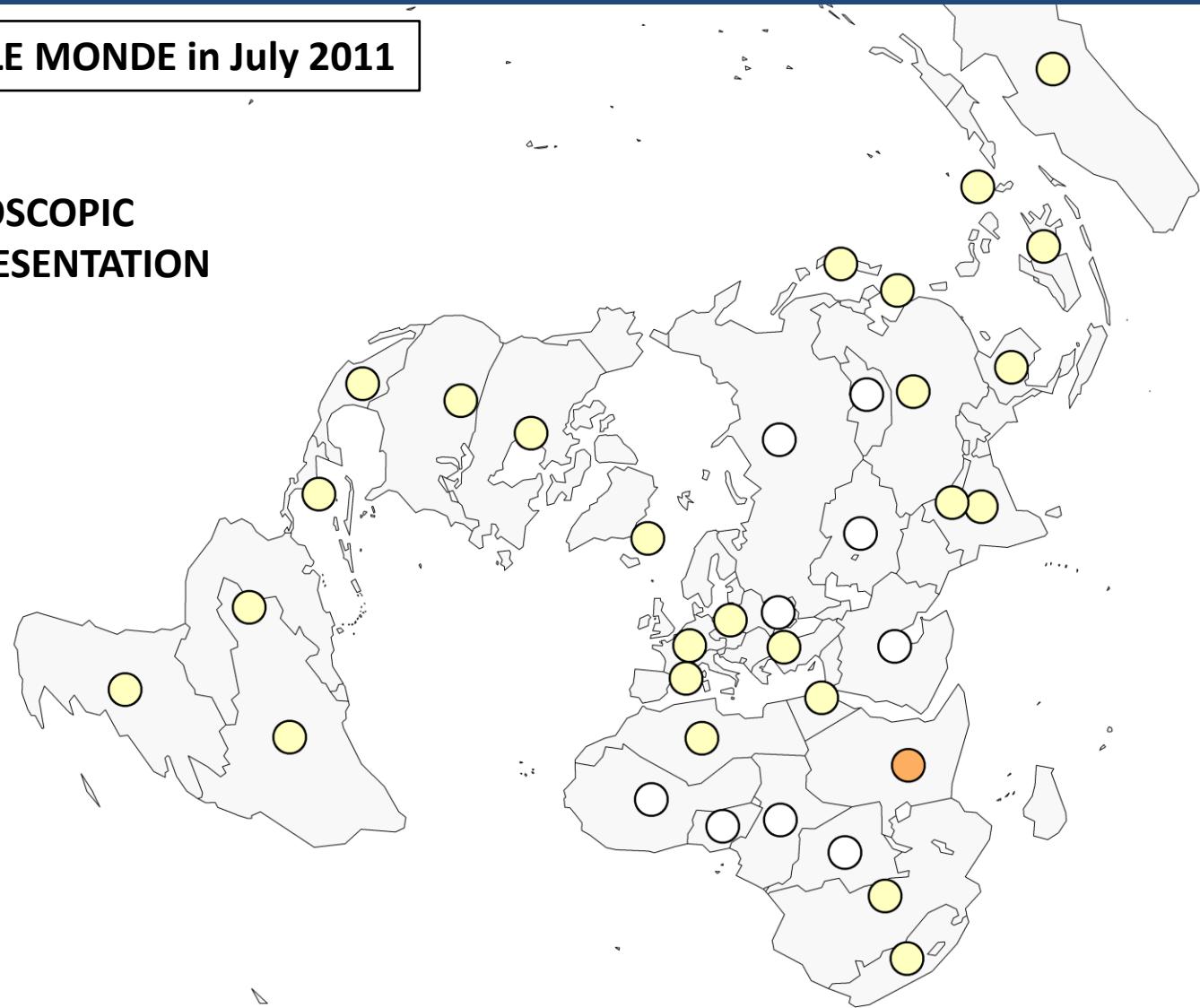


P2: The Levels of Representation

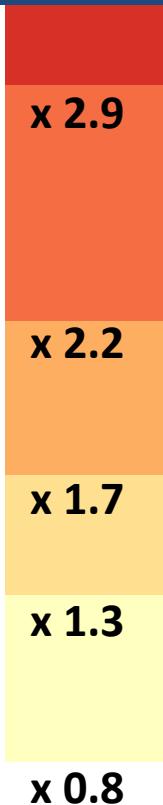


Newspaper LE MONDE in July 2011

MESOSCOPIC
REPRESENTATION

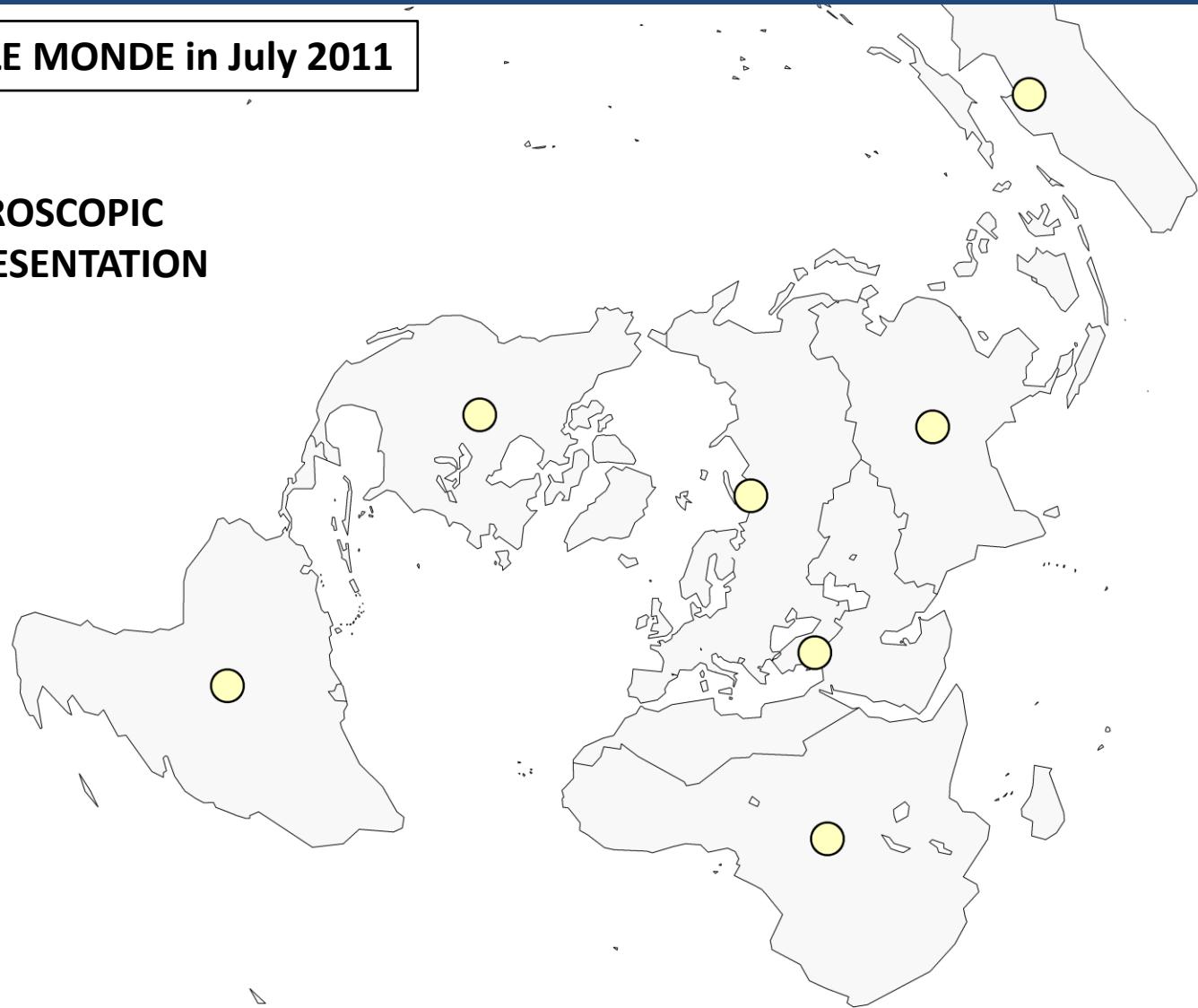


P2: The Levels of Representation

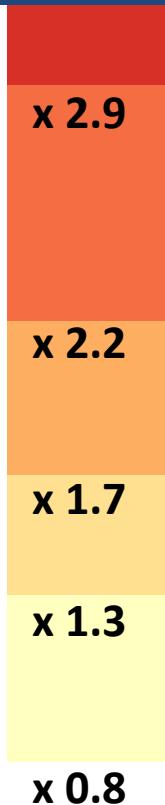


Newspaper LE MONDE in July 2011

MACROSCOPIC
REPRESENTATION

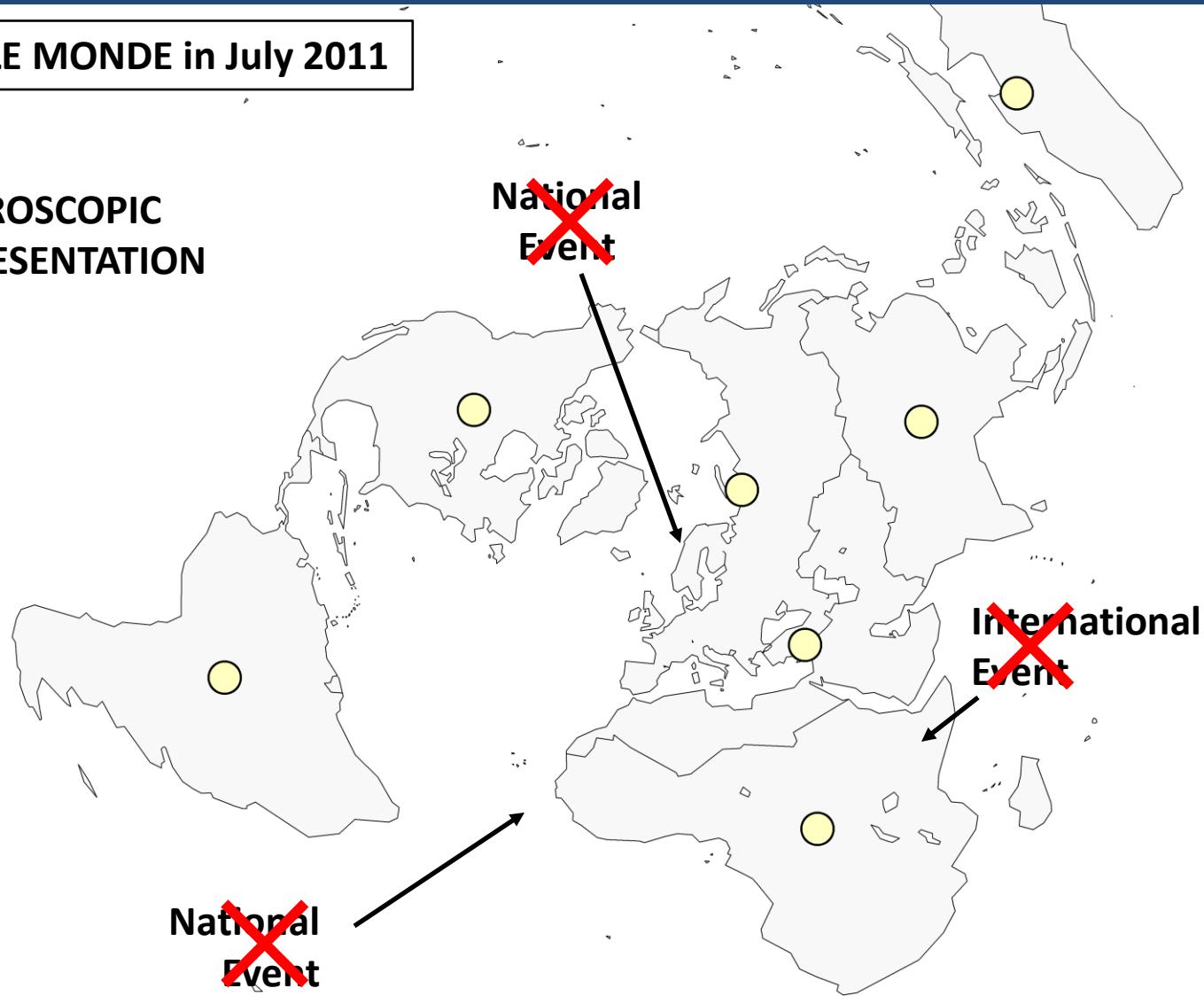


P2: The Levels of Representation

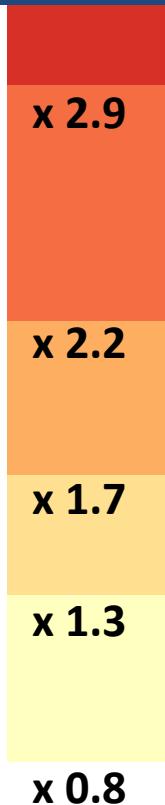


Newspaper LE MONDE in July 2011

MACROSCOPIC
REPRESENTATION

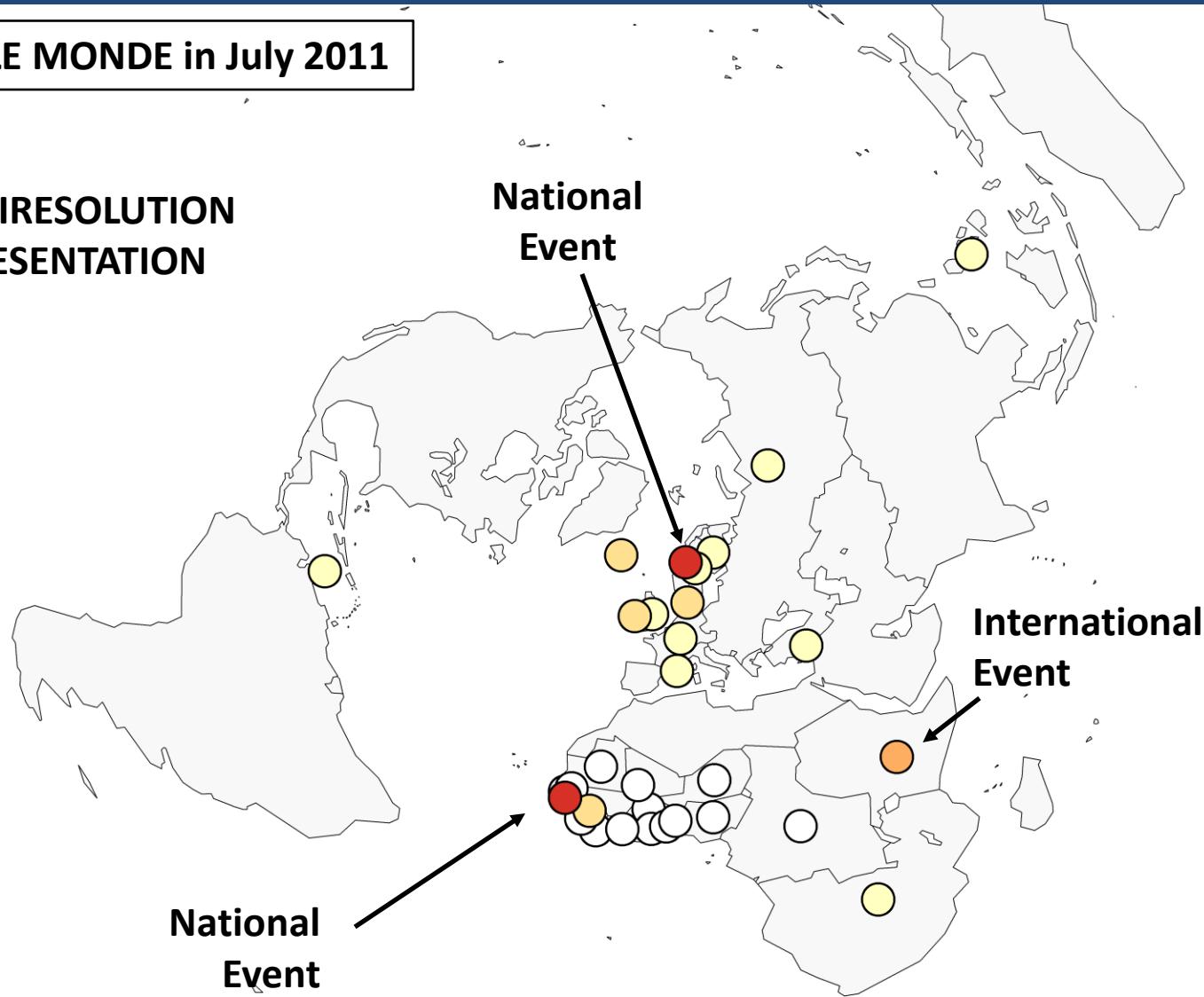


P2: The Levels of Representation



Newspaper LE MONDE in July 2011

MULTIRESOLUTION
REPRESENTATION



P2: The Levels of Representation



Newspaper LE MONDE in July 2011

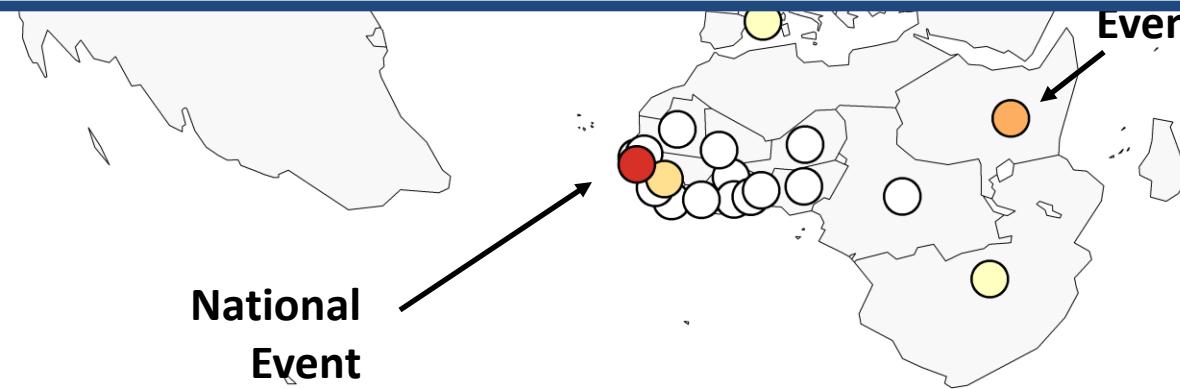
MULTIRESOLUTION
REPRESENTATION

National
Event

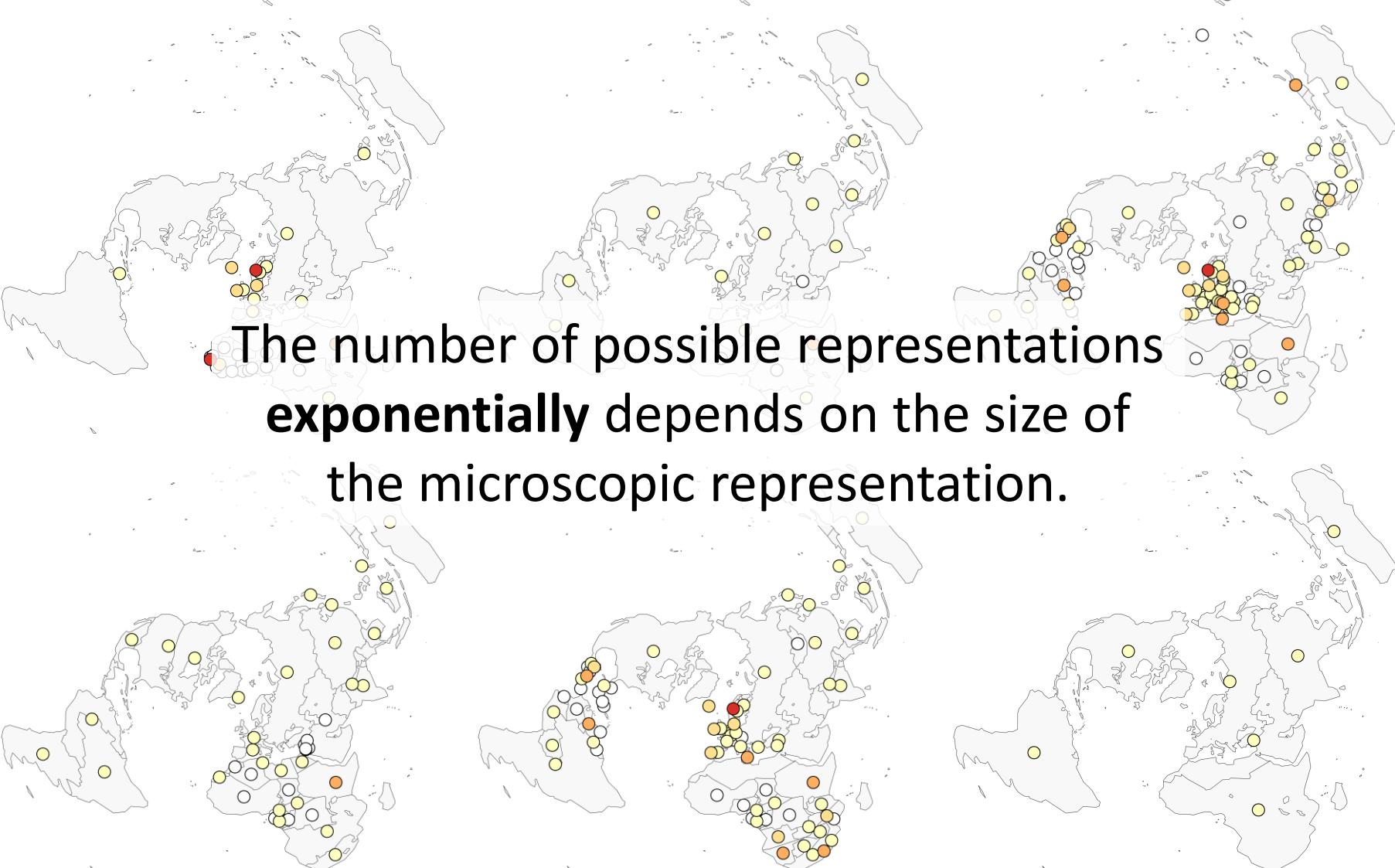
Problem 2: How to find the representation levels that are **relevant** for the analysis?

National
Event

Event



P3: The Computation of the Best Representations

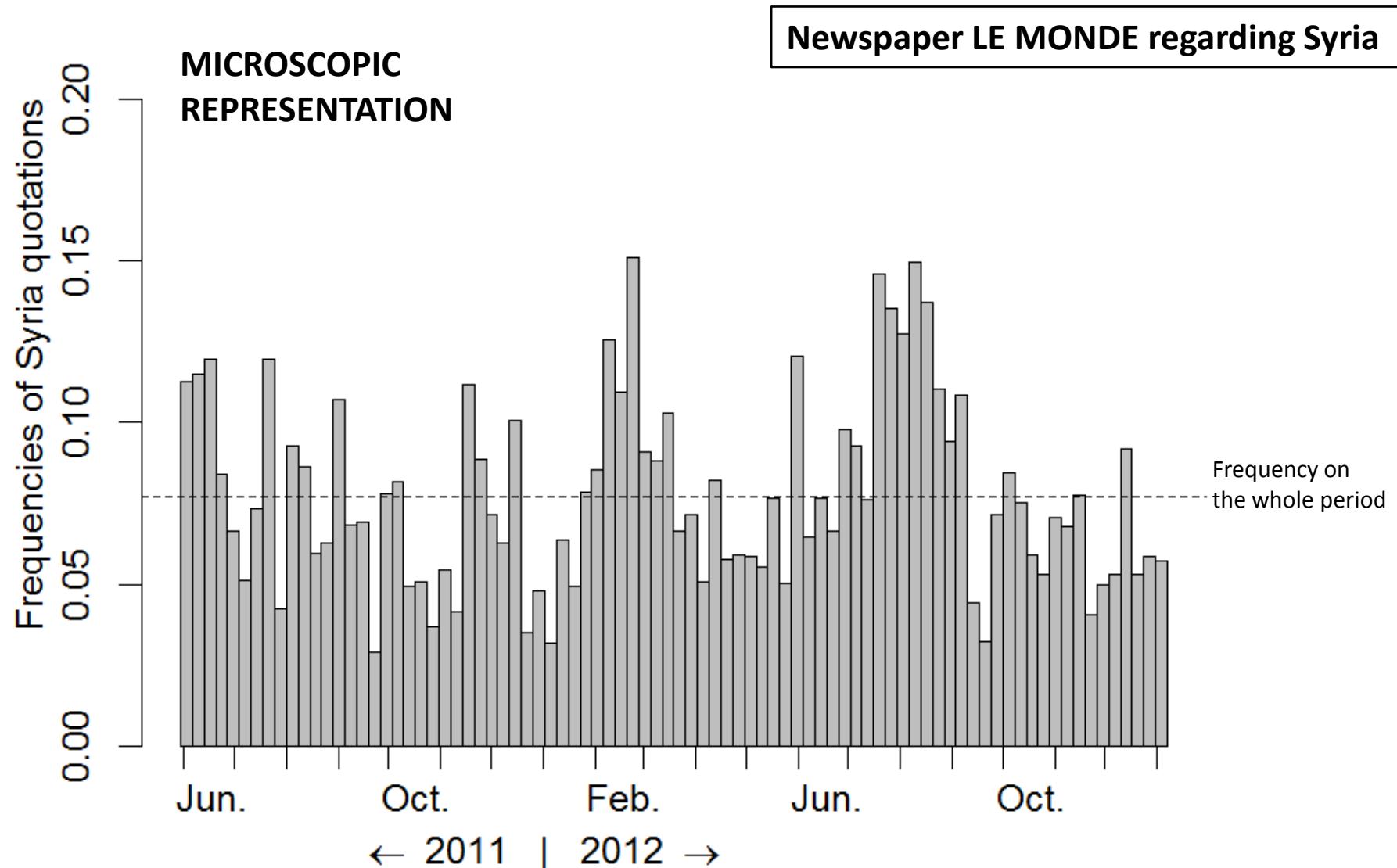


The number of possible representations **exponentially** depends on the size of the microscopic representation.

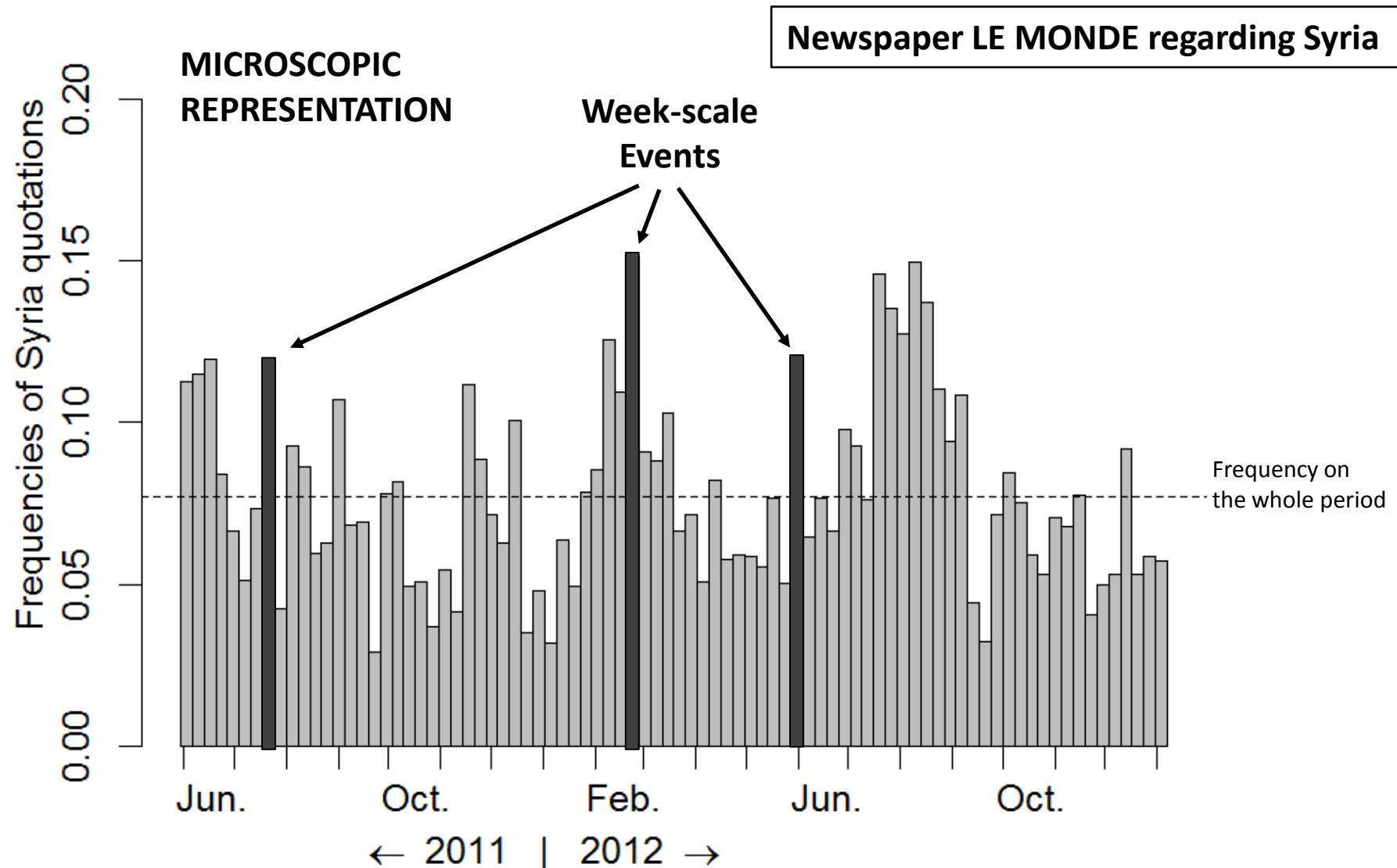
P3: The Computation of the Best Representations

Problem 3 : How to compute the optimal representation in an efficient way?

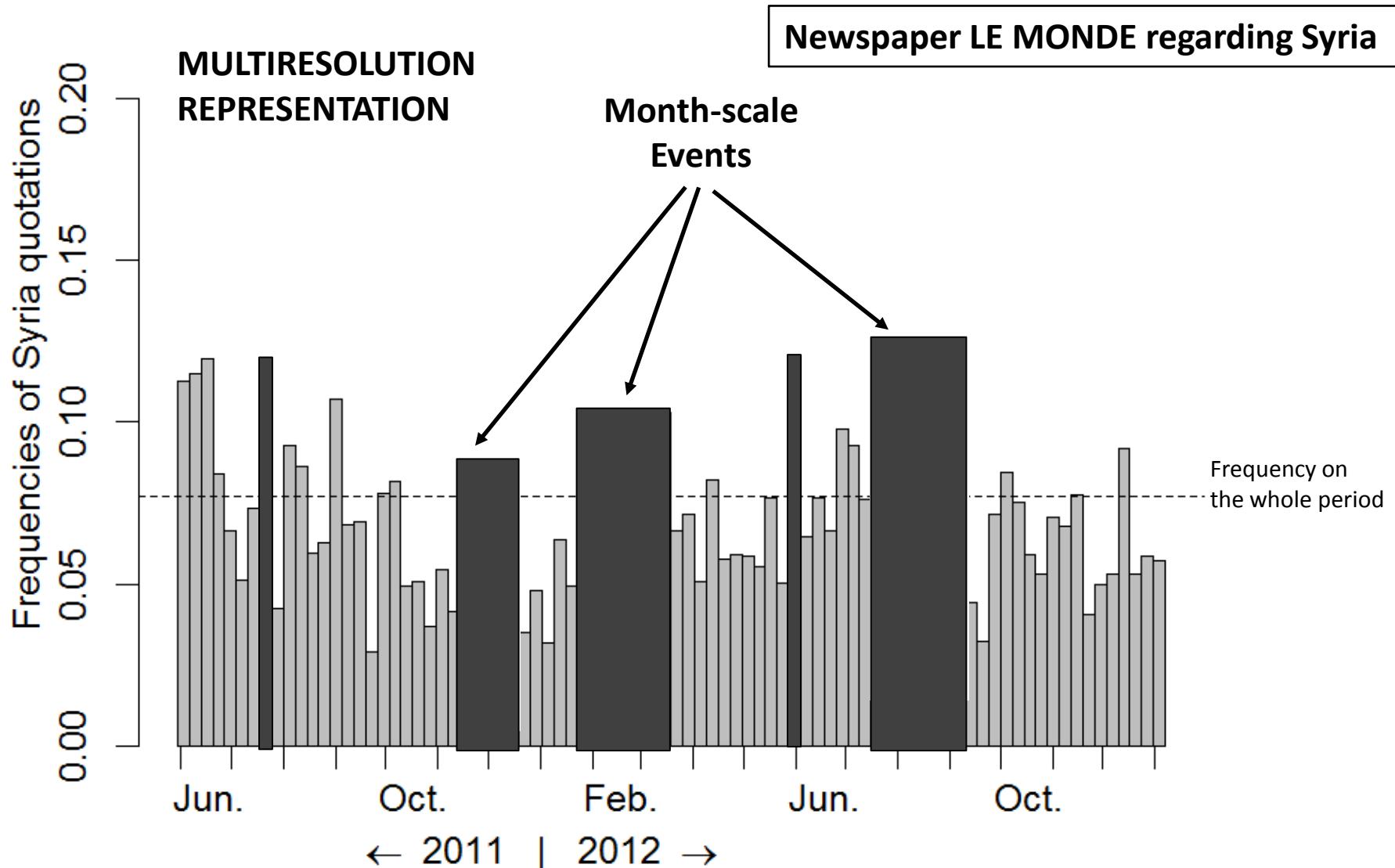
And for Other Dimensions?



And for Other Dimensions?



And for Other Dimensions?



Outline

P0 To characterize the aggregation process

P1 To preserve the system's semantics

To aggregate according to several dimension

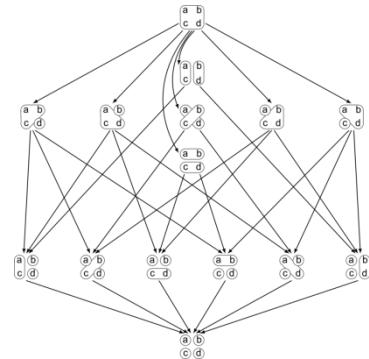
P2 To evaluate and compare the representations

To offer several granularity levels

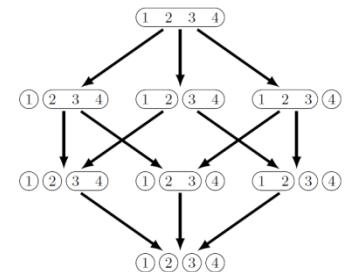
P3 To compute the best representations

My Approach

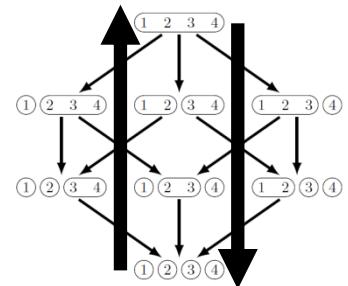
P0 To characterize the aggregation process
→ The algebra of possible partitions



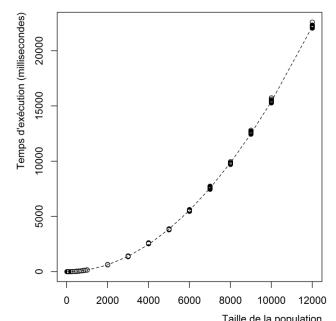
P1 To preserve the system's semantics
→ A constrained partitioning method
To aggregate according to several dimension
→ Some constraints expressing the system's topology



P2 To evaluate and compare the representations
→ Some measures of complexity and information
To offer several granularity levels
→ The optimization of a compromise



P3 To compute the best representations
→ A generic algorithm of constrained optimization



My Approach

P0

To characterize the aggregation process
→ The algebra of possible partitions

P1

To preserve the system's semantics
→ A constrained partitioning method
To aggregate according to several dimension
→ Some constraints expressing the system's topology

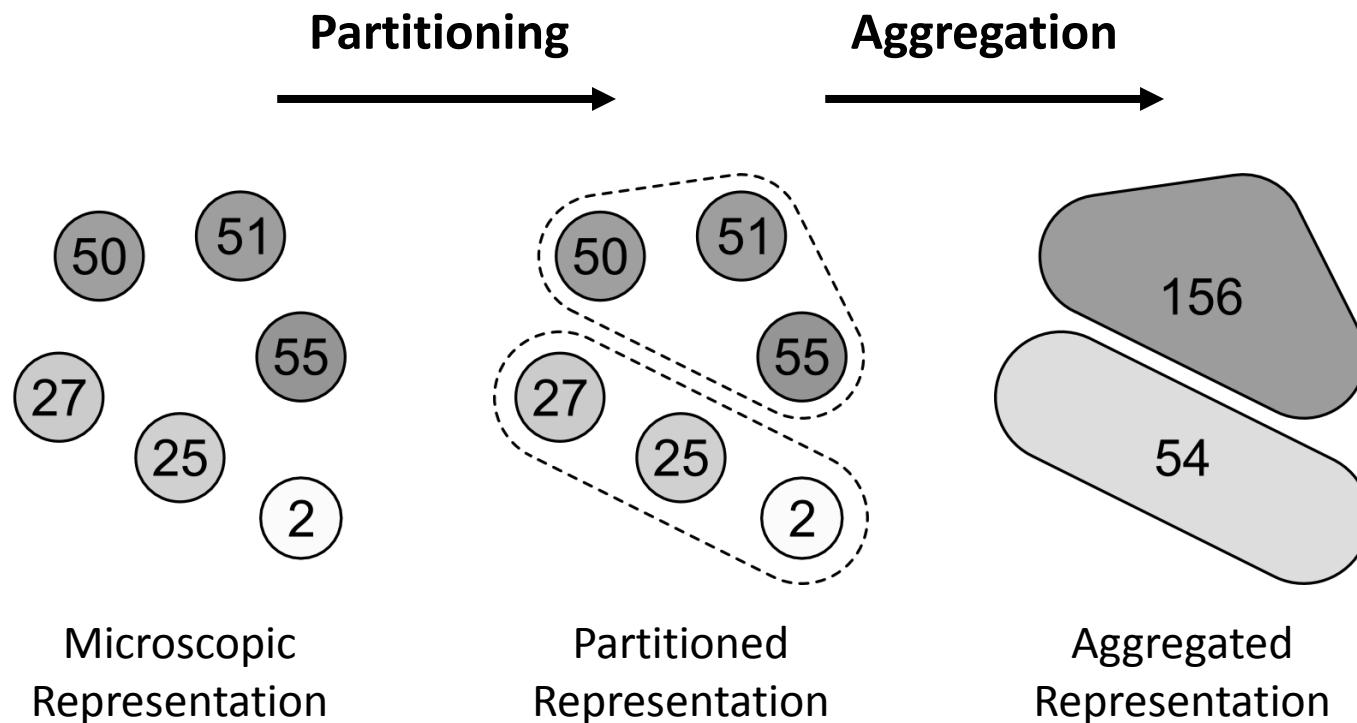
P2

To evaluate and compare the representations
→ Some measures of complexity and information
To offer several granularity levels
→ The optimization of a compromise

P3

To compute the best representations
→ A generic algorithm of constrained optimization

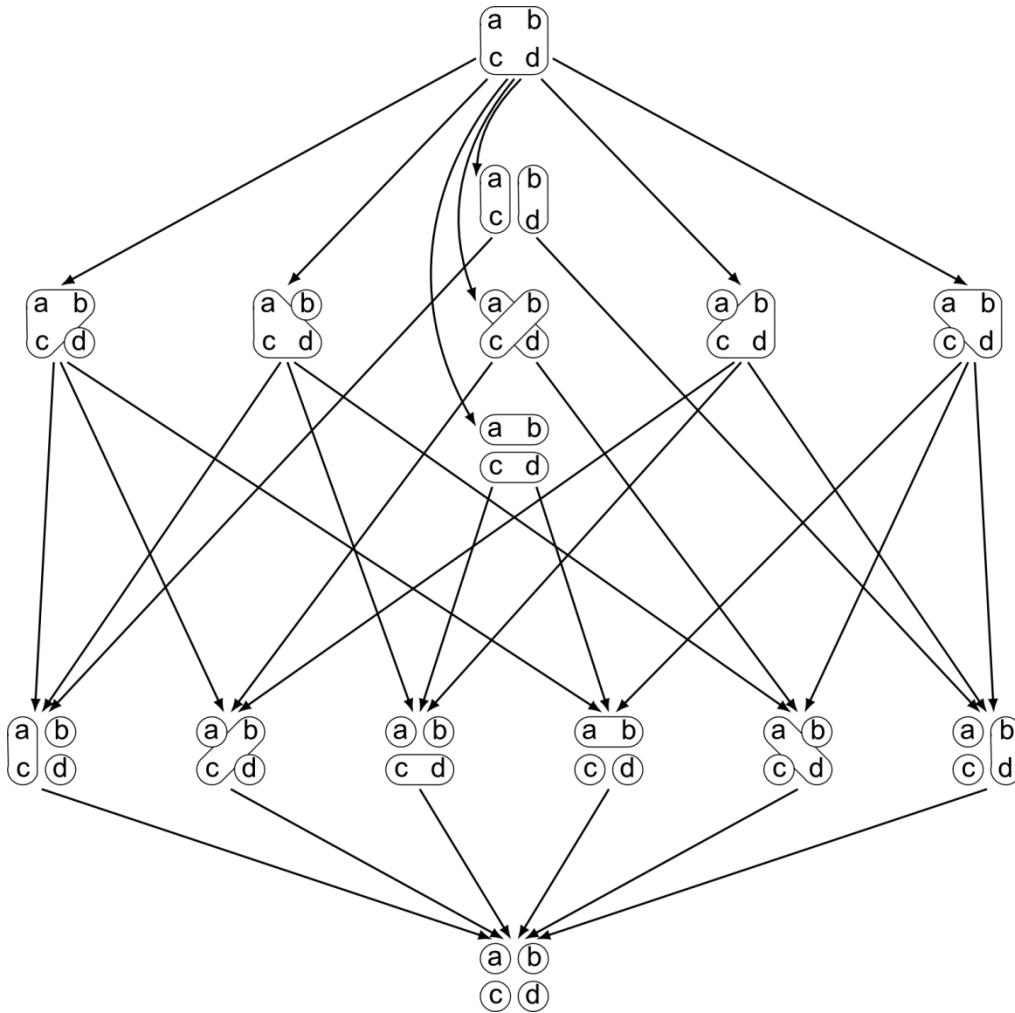
The Aggregation Process



Set of Possible Partitions



Set of Possible Partitions

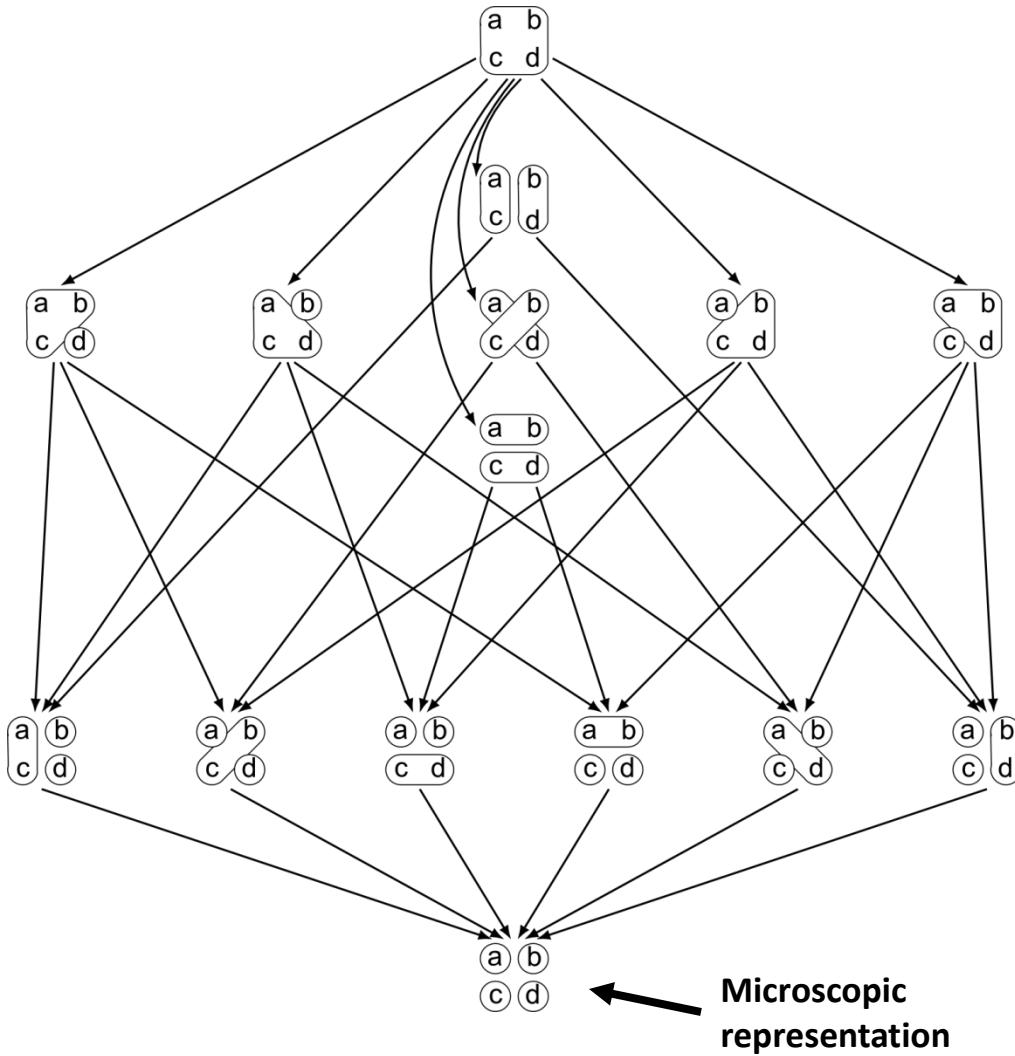


Algebraic Structure

A partial order on the set of possible partitions

→ the refinement relation

Set of Possible Partitions

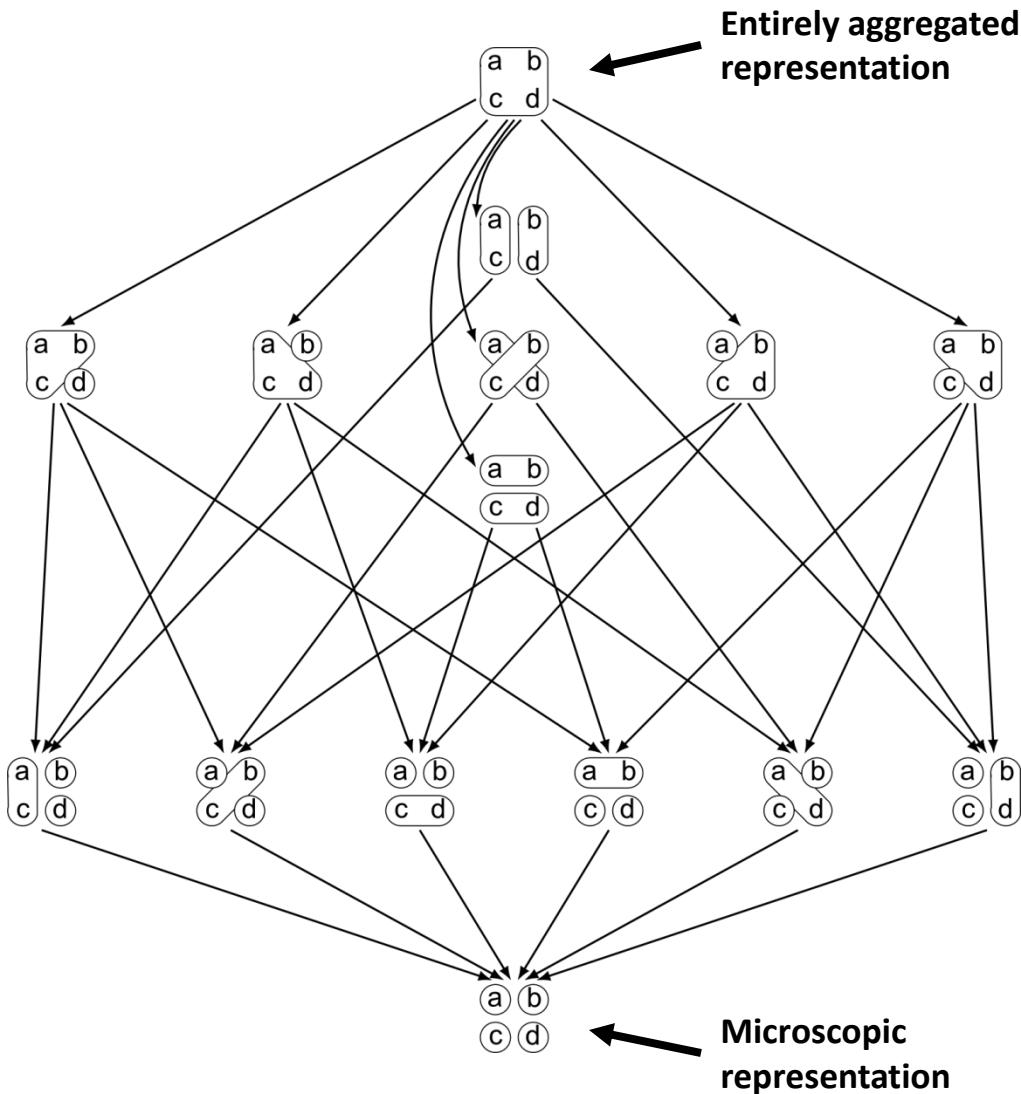


Algebraic Structure

A partial order on the set of possible partitions

→ the refinement relation

Set of Possible Partitions

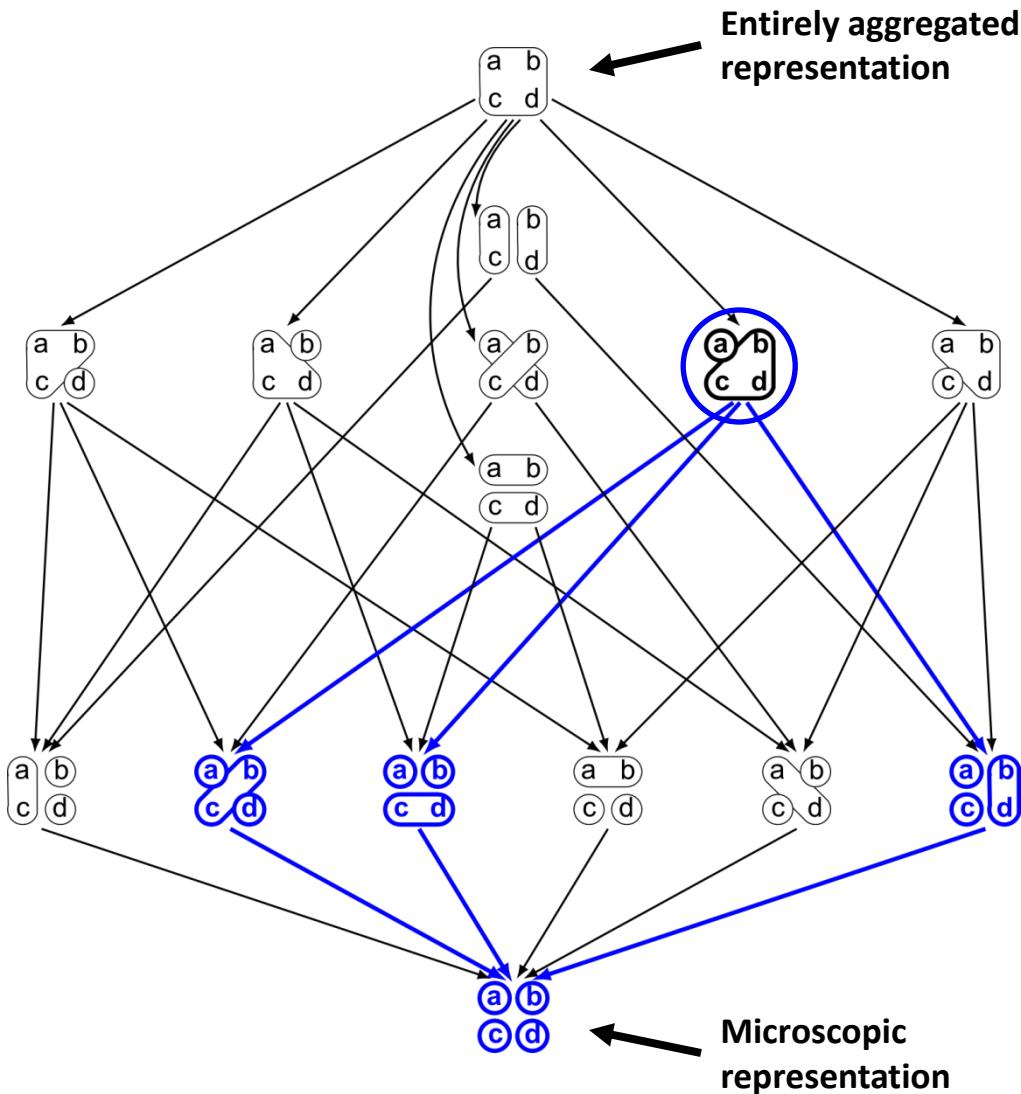


Algebraic Structure

A partial order on the set of possible partitions

→ the refinement relation

Set of Possible Partitions

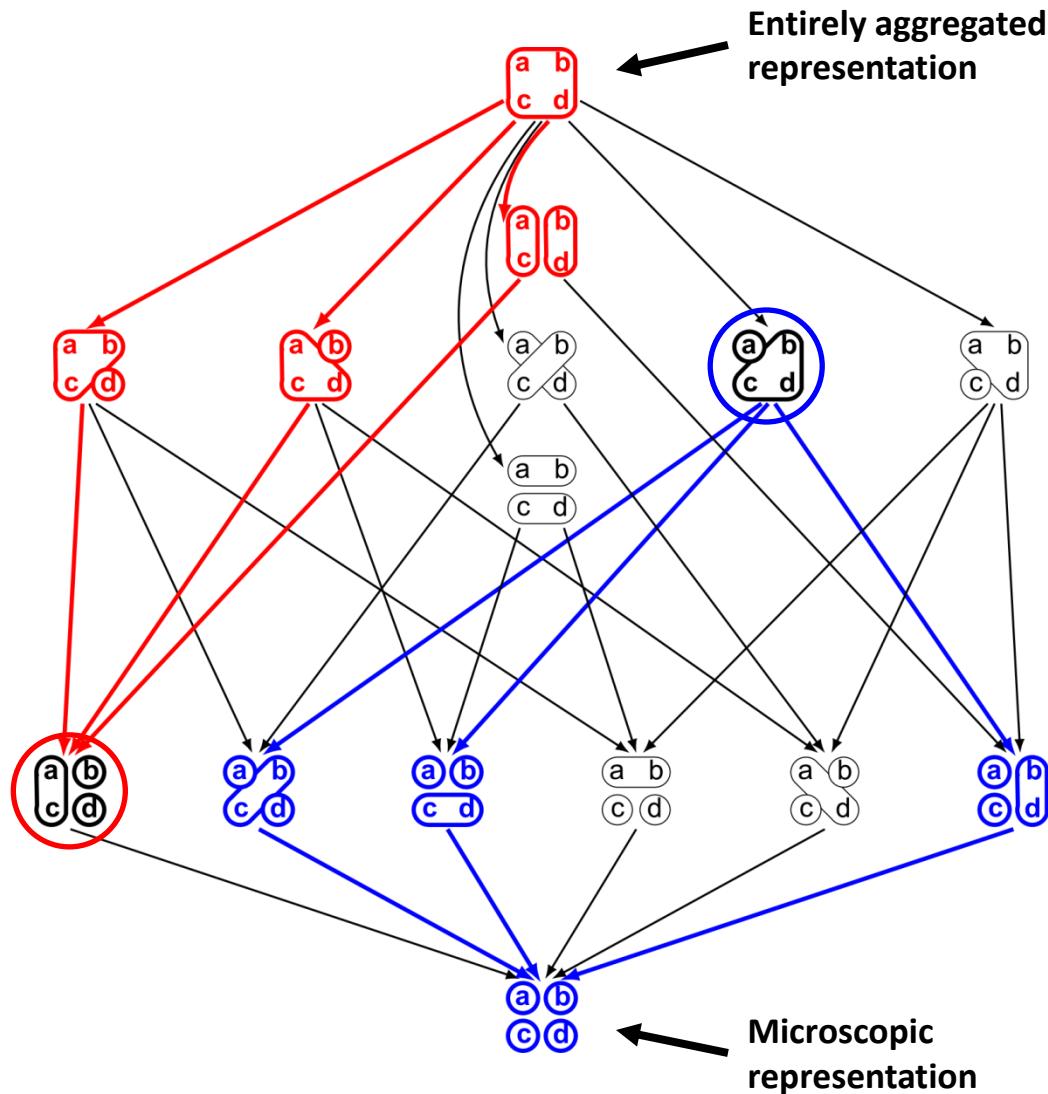


Algebraic Structure

A partial order on the set of possible partitions

→ the refinement relation

Set of Possible Partitions

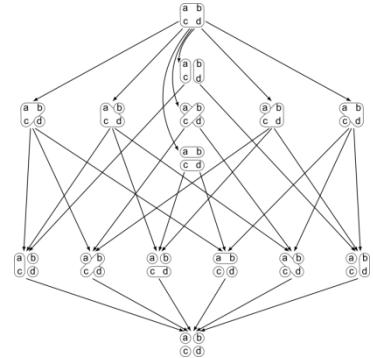


Algebraic Structure

A partial order on the set of possible partitions

→ the refinement relation

My Approach



P0 To characterize the aggregation process
→ The algebra of possible partitions

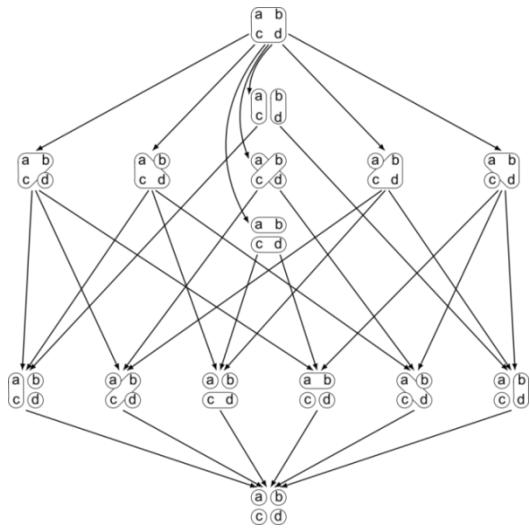
P1 To preserve the system's semantics
→ A constrained partitioning method
To aggregate according to several dimension
→ Some constraints expressing the system's topology

P2 To evaluate and compare the representations
→ Some measures of complexity and information
To offer several granularity levels
→ The optimization of a compromise

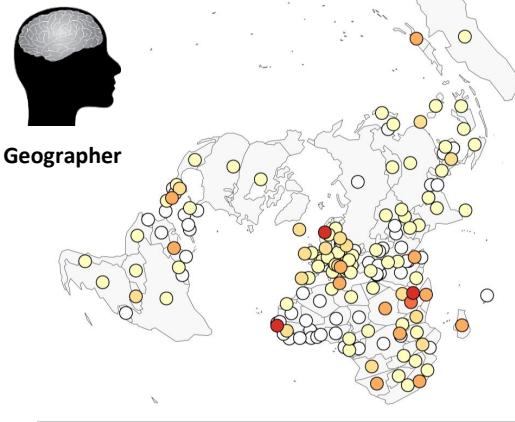
P3 To compute the best representations
→ A generic algorithm of constrained optimization

Problems and Objectives

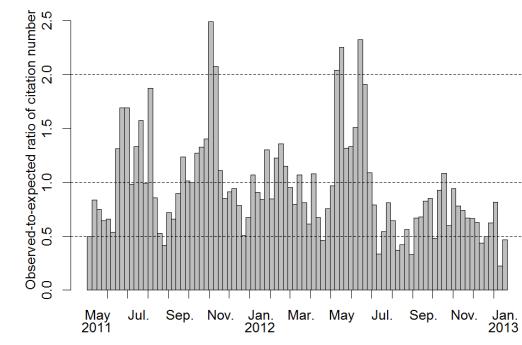
Set of possible partitions



Geographical Semantics

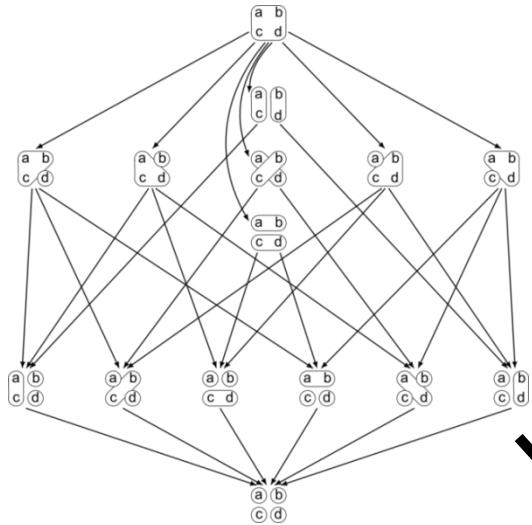


Temporal Semantics

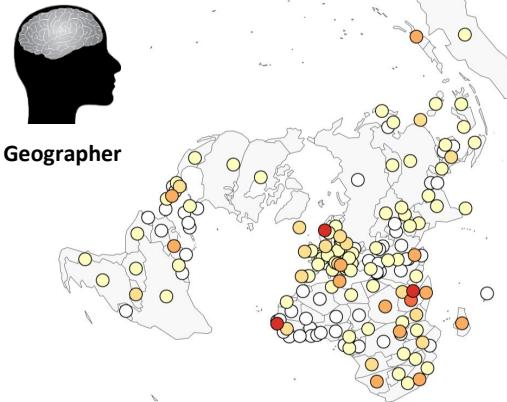


Problems and Objectives

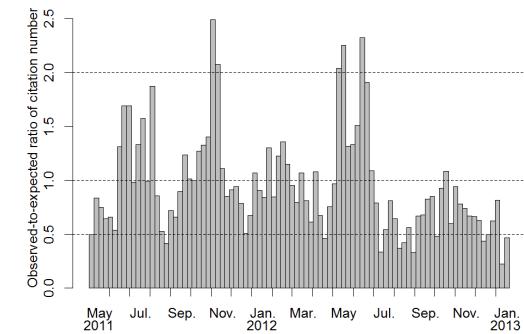
Set of possible partitions



Geographical Semantics

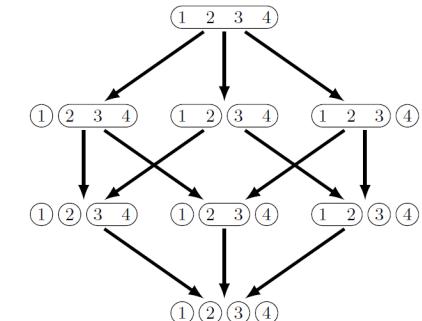
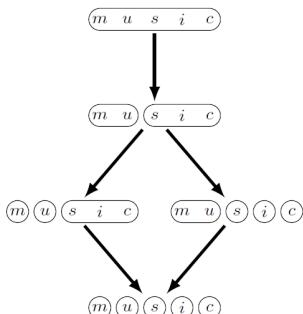


Temporal Semantics



Constraints

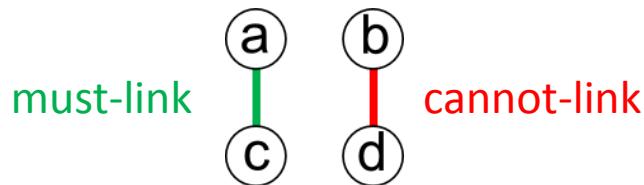
Set of admissible partitions



Constrained Partitioning

At the instances level

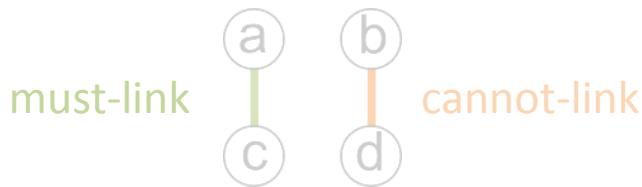
[Davidson and Basu, 2007]



Constrained Partitioning

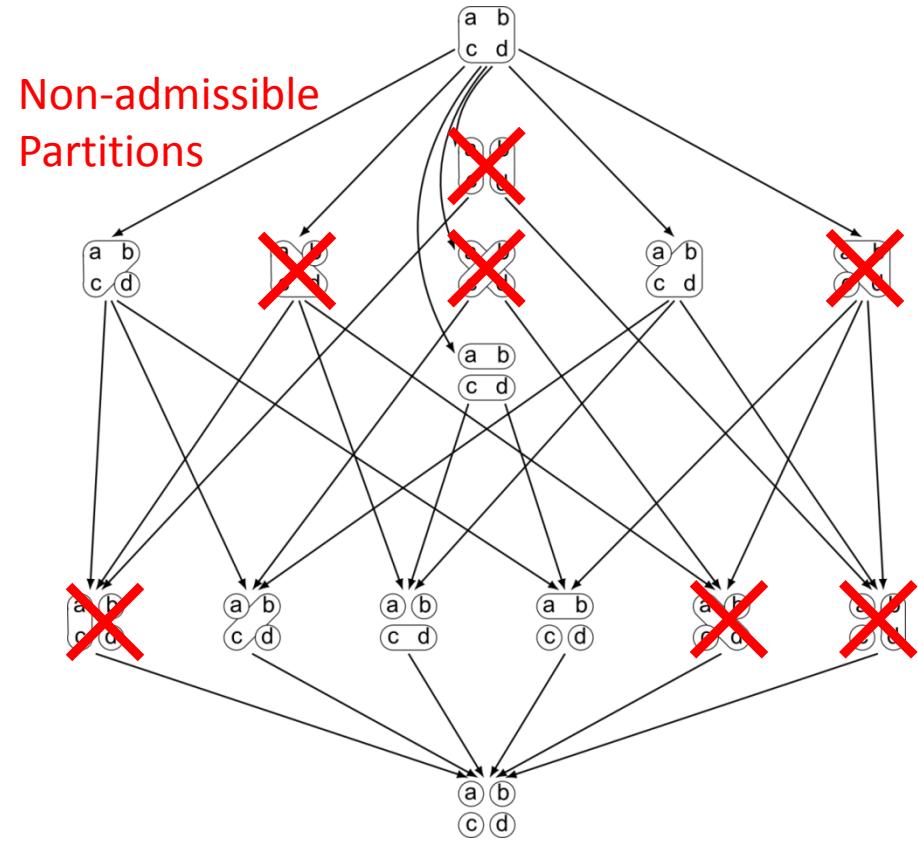
At the instances level

[Davidson and Basu, 2007]



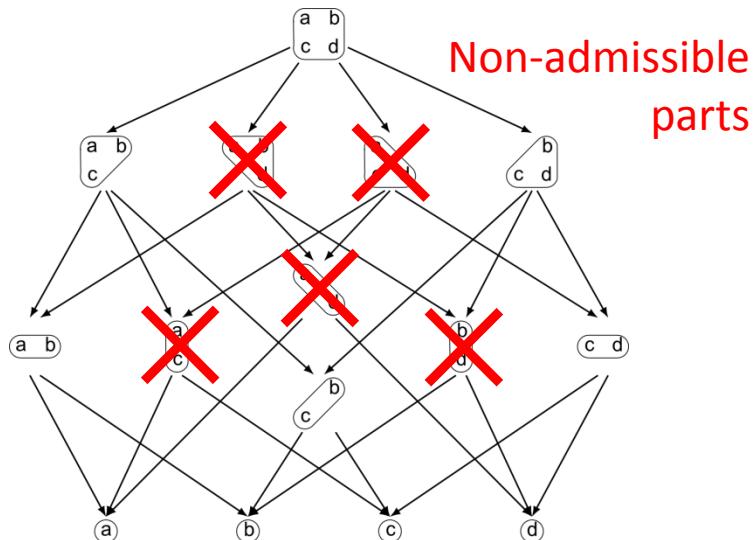
At the partitions level

[Lamarche-Perrin *et al.*, IAT 2013]



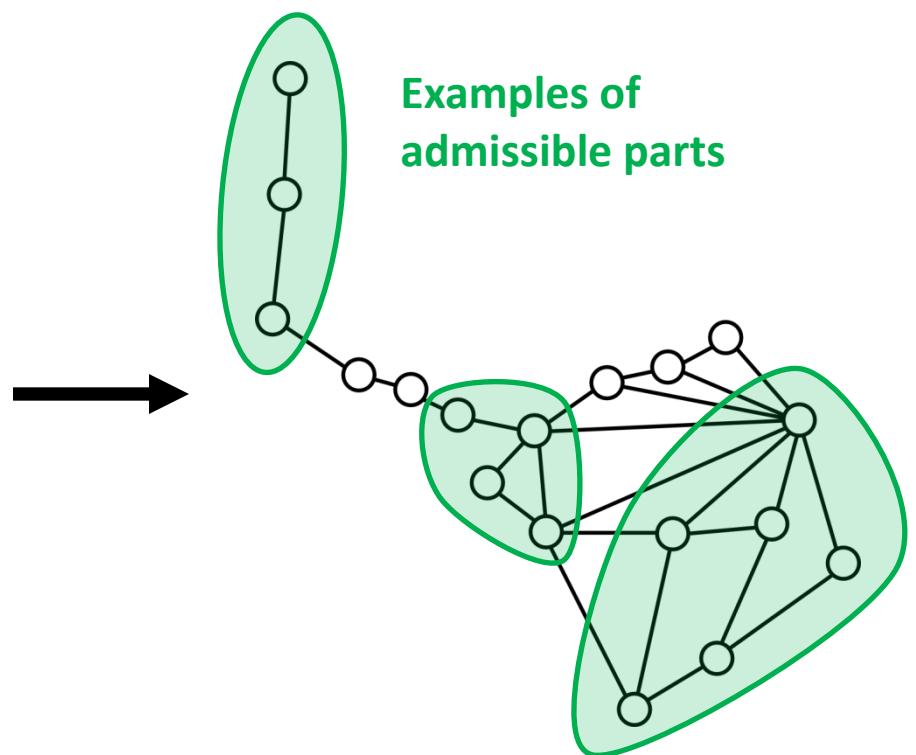
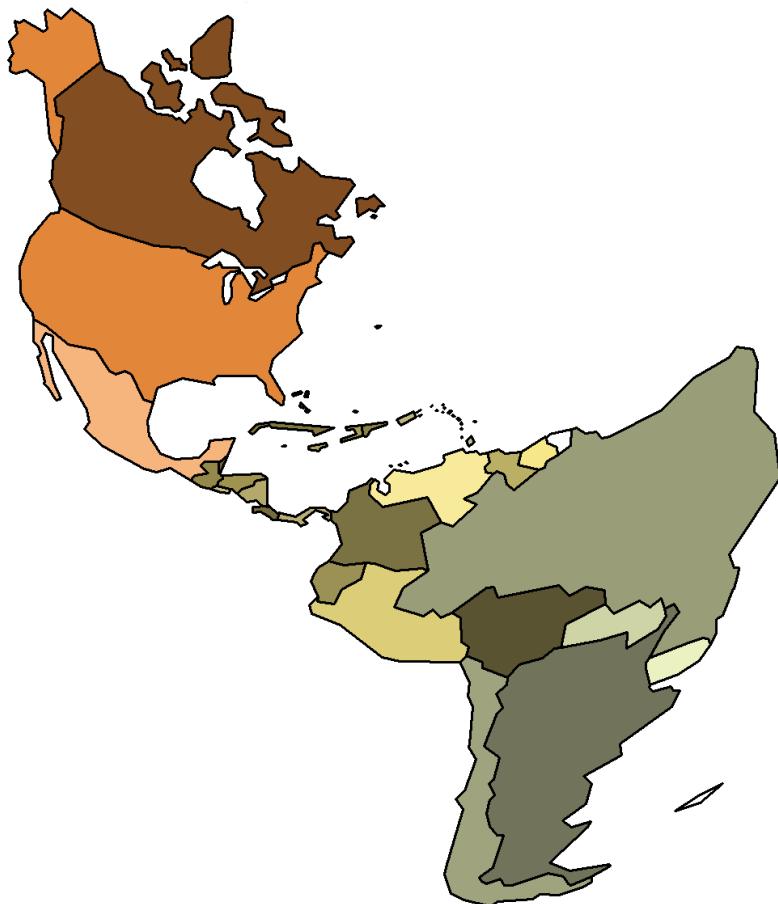
At the parts level

[Lamarche-Perrin *et al.*, IAT 2013]



Preserving the Neighborhood Relation

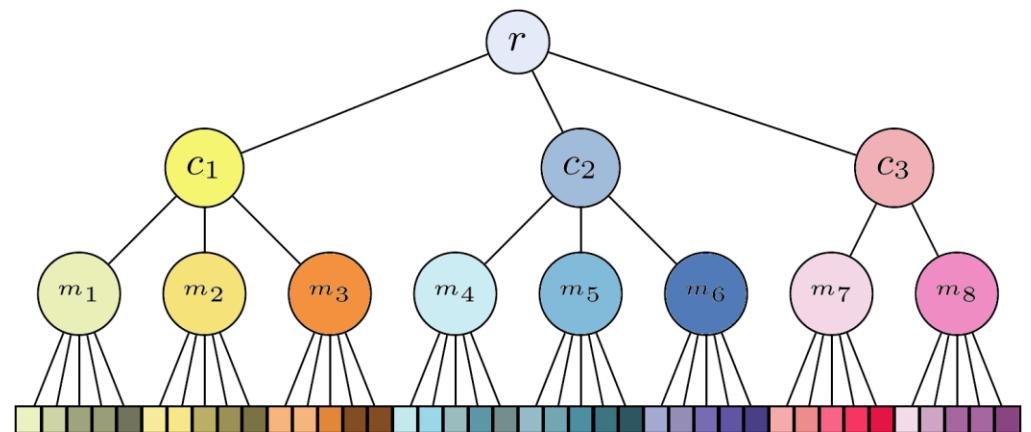
Admissible Parts: set of connected countries regarding the adjacency graph



The WUTS Hierarchy

[Grasland and Didelon, 2007]

Admissible Parts:
set of countries that are
politically, culturally and
economically consistent

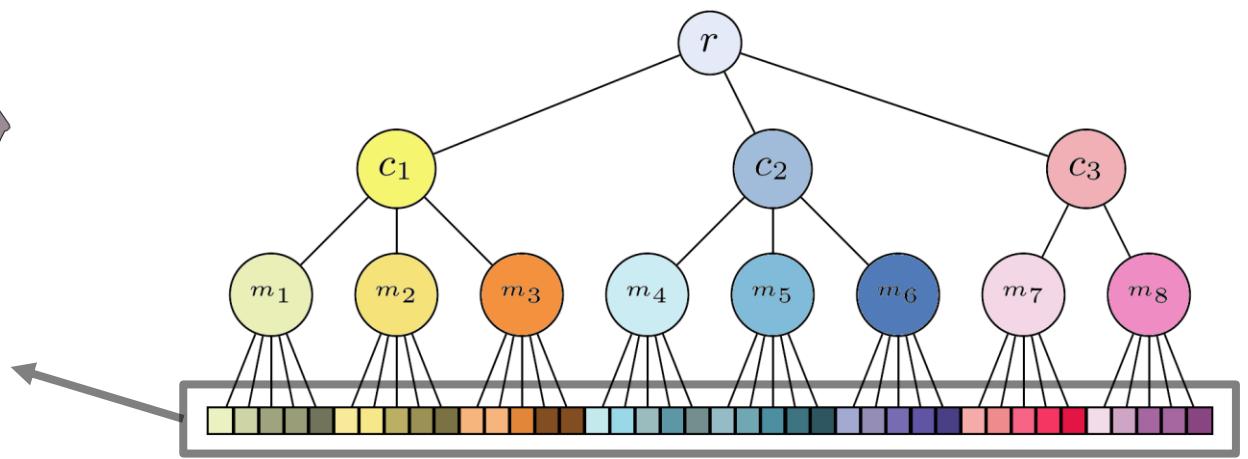
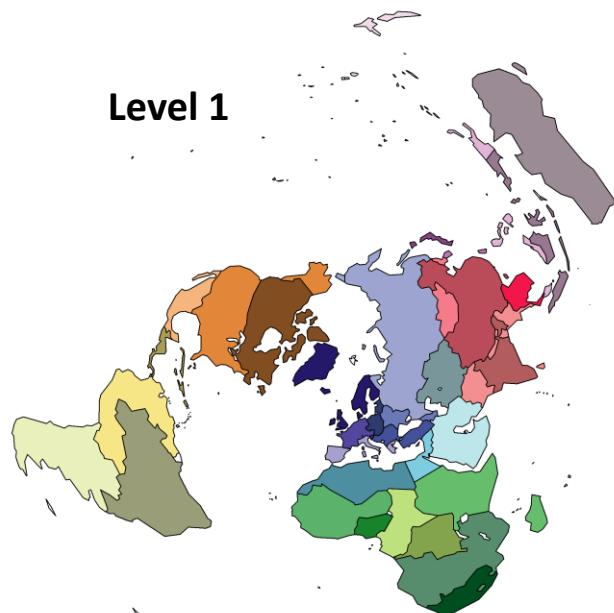


The WUTS Hierarchy

[Grasland and Didelon, 2007]

Admissible Parts:
set of countries that are
politically, culturally and
economically consistent

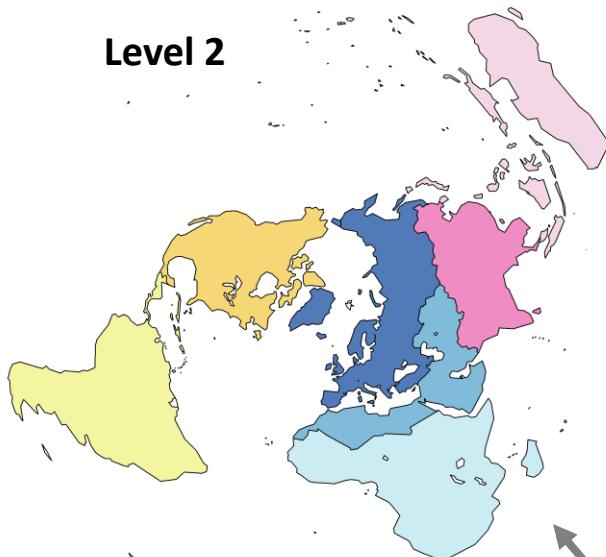
Level 1



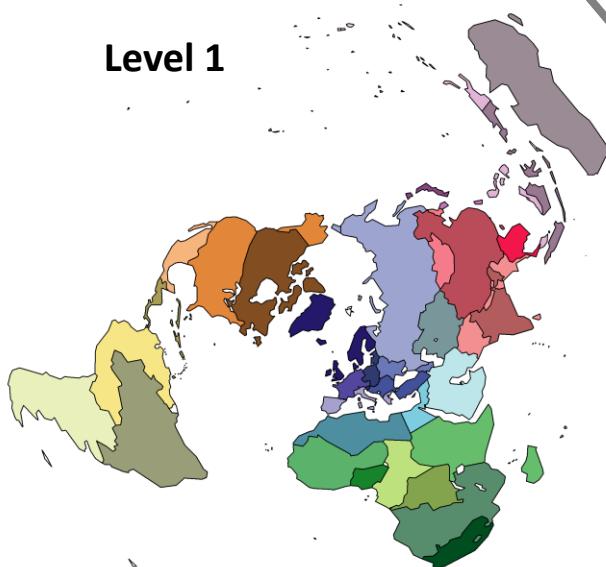
The WUTS Hierarchy

[Grasland and Didelon, 2007]

Level 2

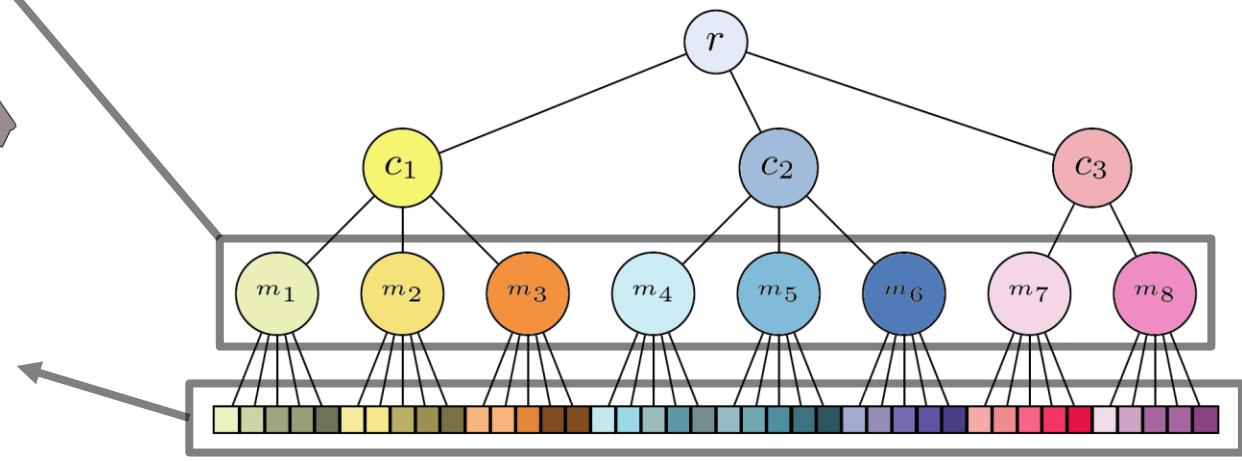


Level 1



Admissible Parts:

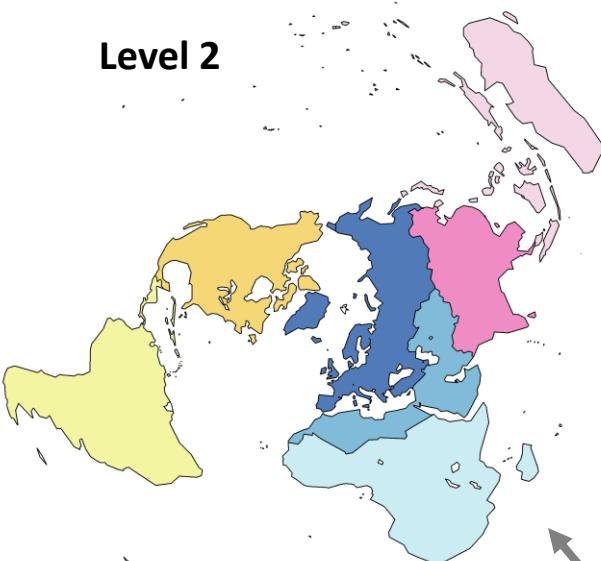
set of countries that are politically, culturally and economically consistent



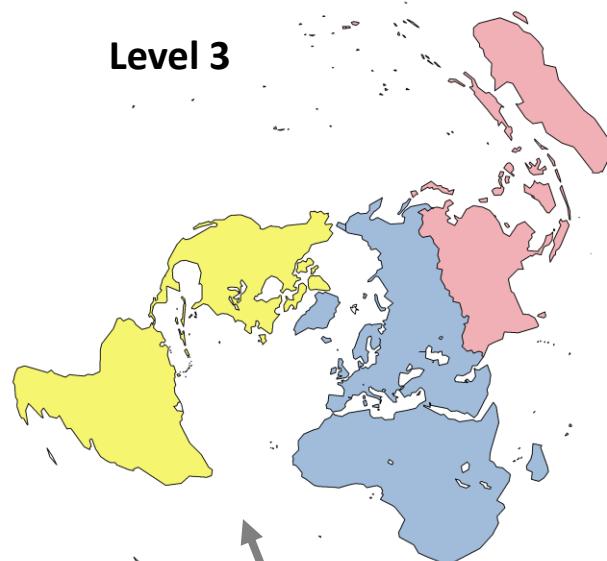
The WUTS Hierarchy

[Grasland and Didelon, 2007]

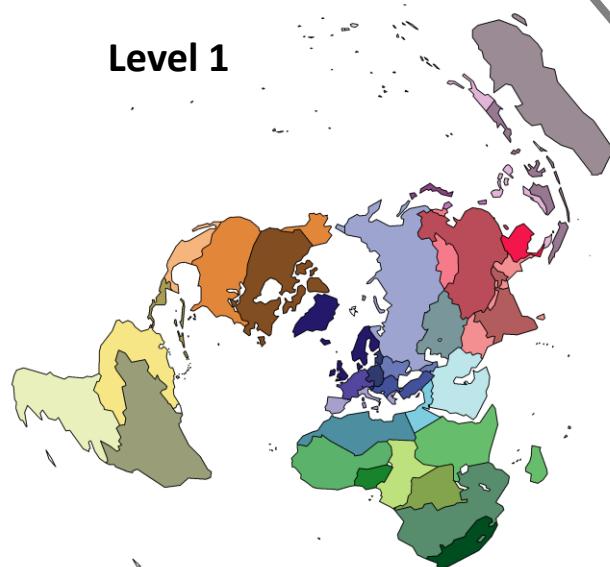
Level 2



Level 3

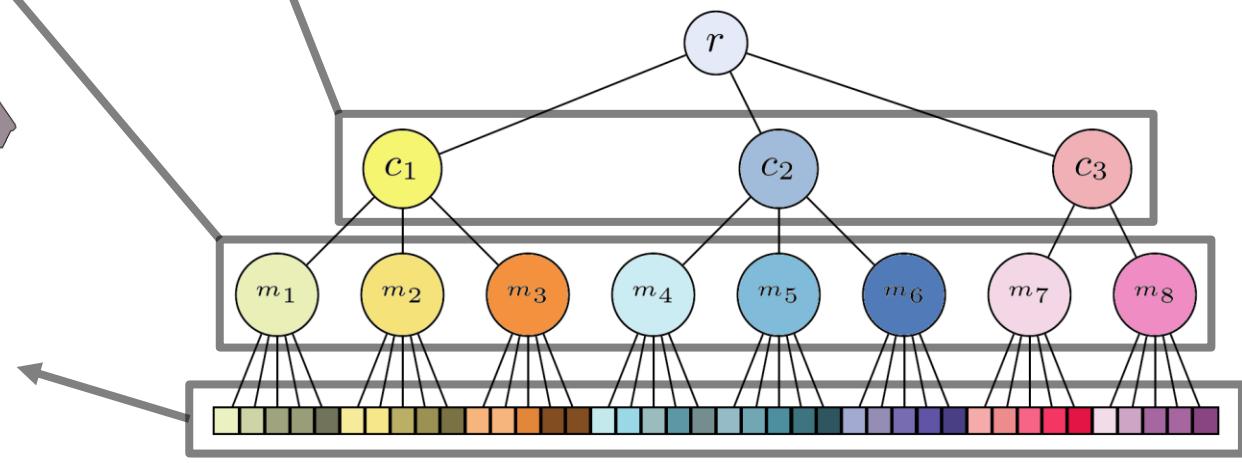


Level 1



Admissible Parts:

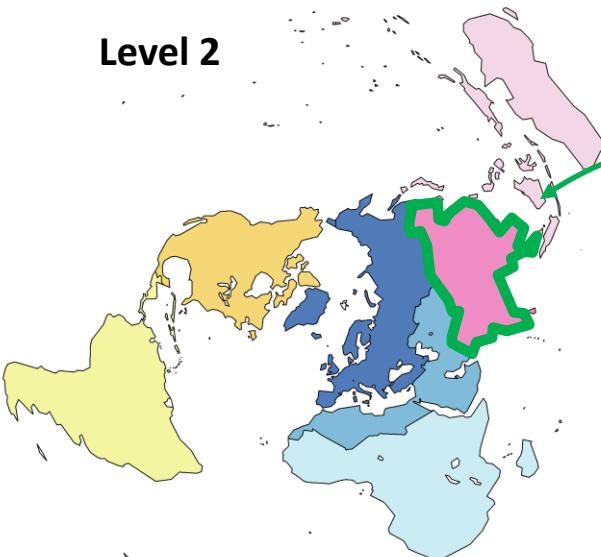
set of countries that are politically, culturally and economically consistent



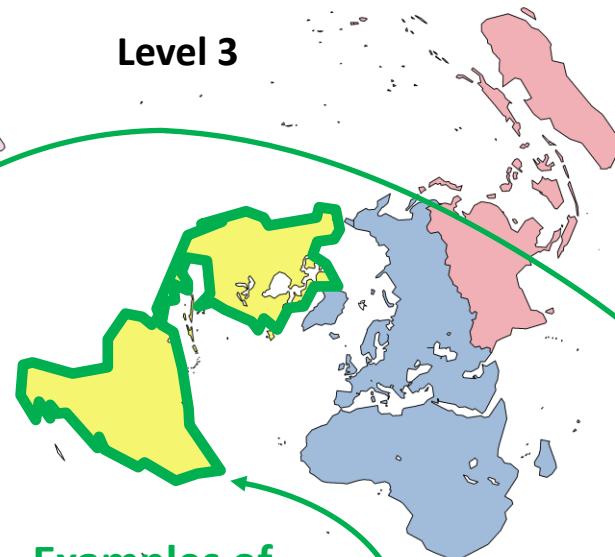
The WUTS Hierarchy

[Grasland and Didelon, 2007]

Level 2



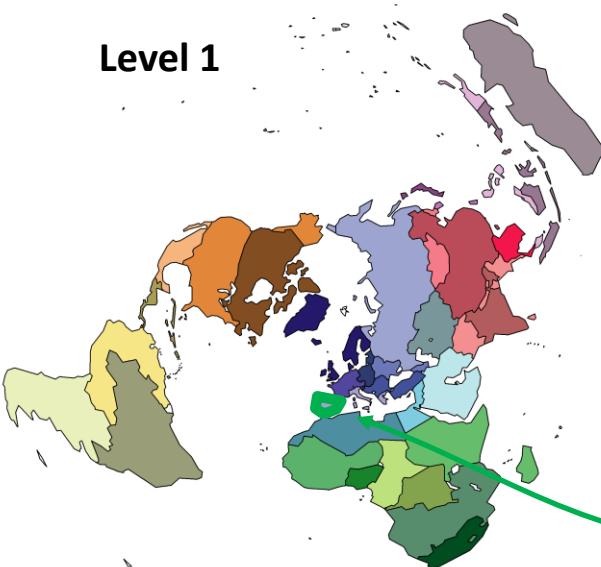
Level 3



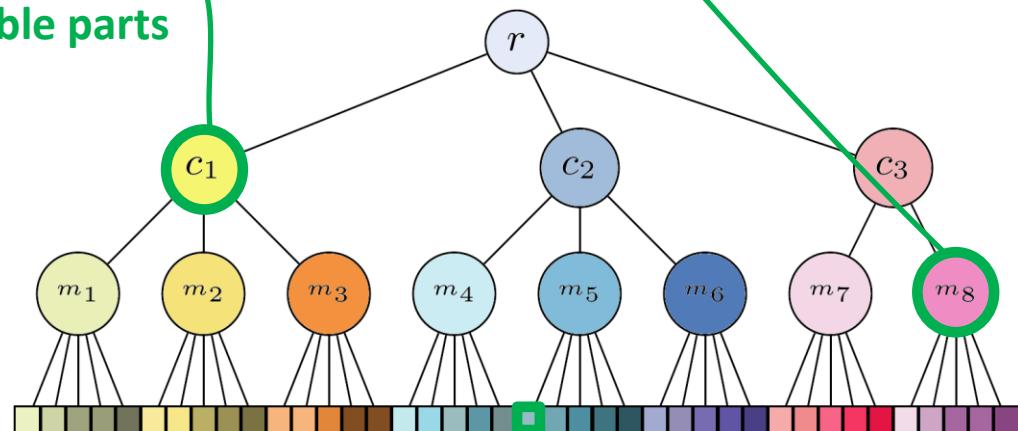
Admissible Parts:

set of countries that are politically, culturally and economically consistent

Level 1

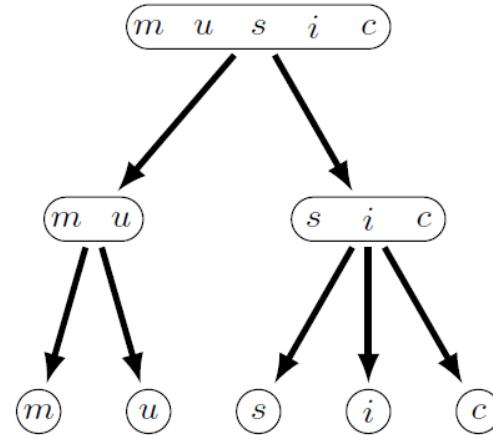


Examples of
admissible parts

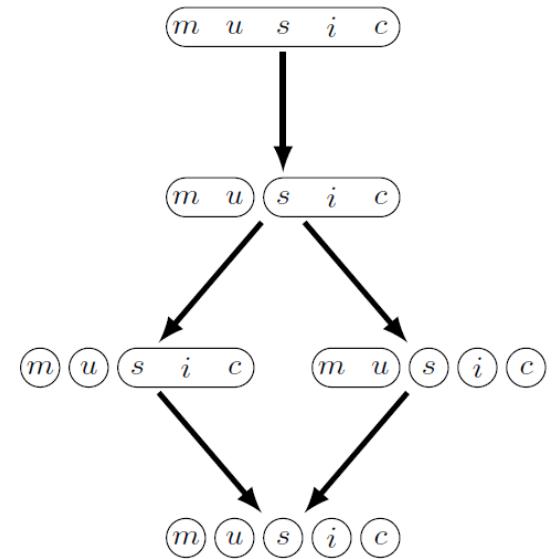


Aggregating according to a Hierarchy

Admissible Parts
(nodes of the hierarchy)

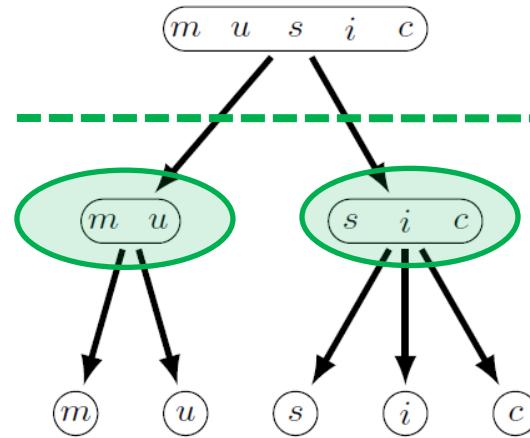


Admissible Partitions
(cuts of the hierarchy)

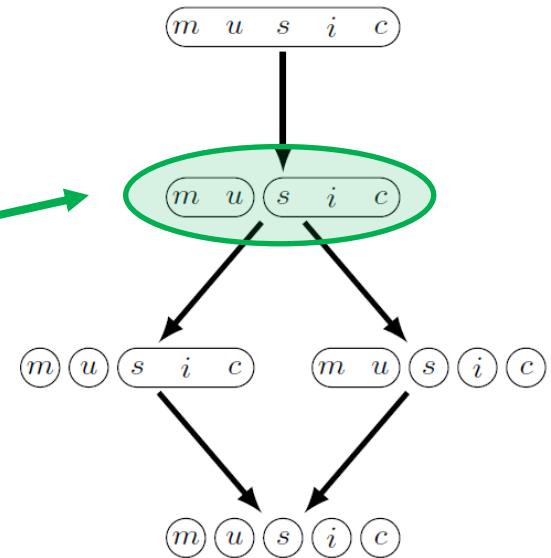


Aggregating according to a Hierarchy

Admissible Parts
(nodes of the hierarchy)

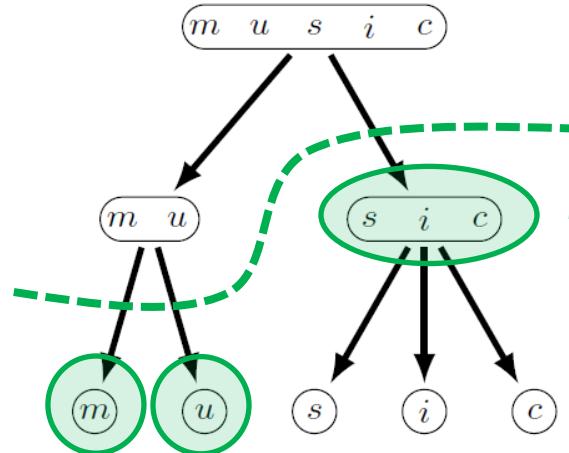


Admissible Partitions
(cuts of the hierarchy)

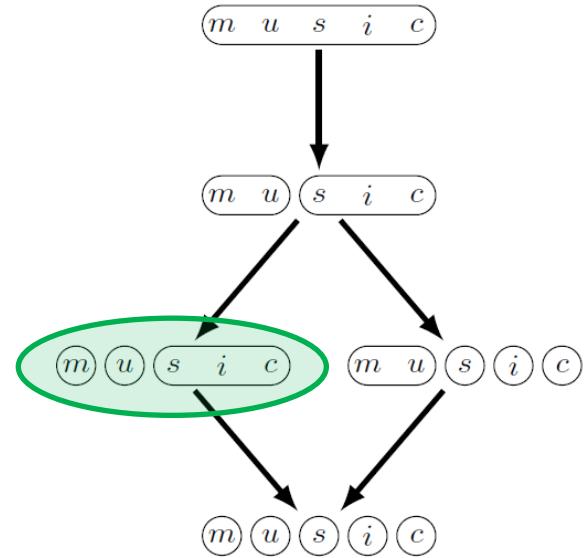


Aggregating according to a Hierarchy

Admissible Parts
(nodes of the hierarchy)



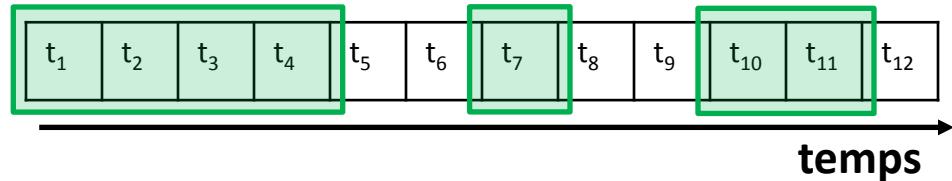
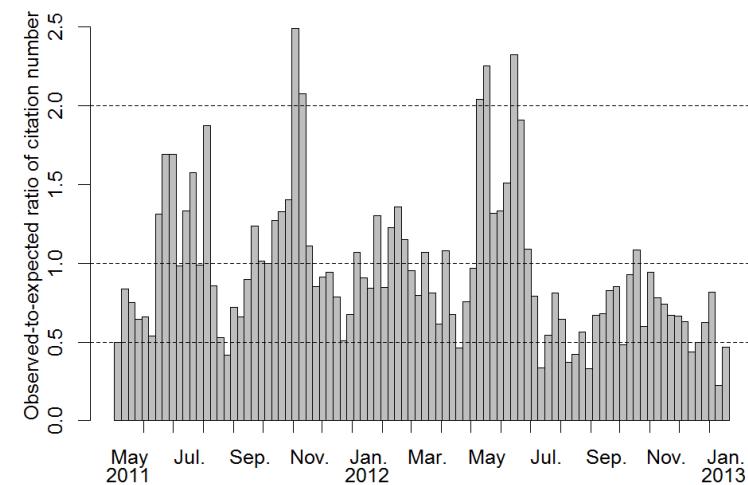
Admissible Partitions
(cuts of the hierarchy)



Preserving the Order of Time

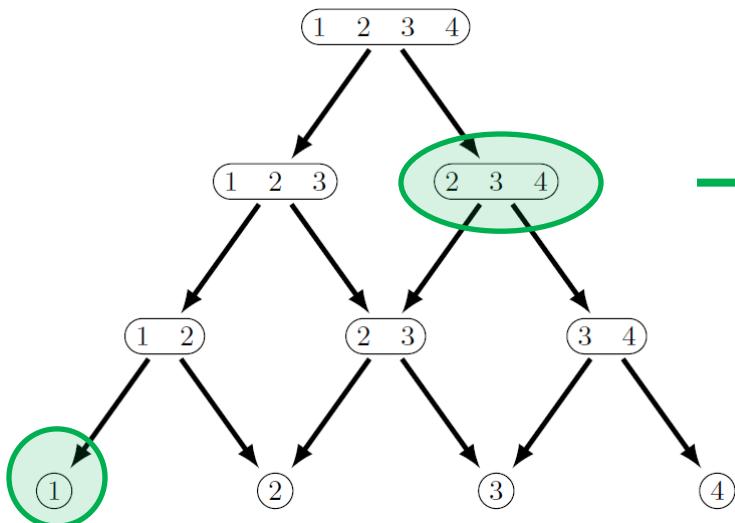
Admissible Parts:
time intervals

Examples of admissible parts

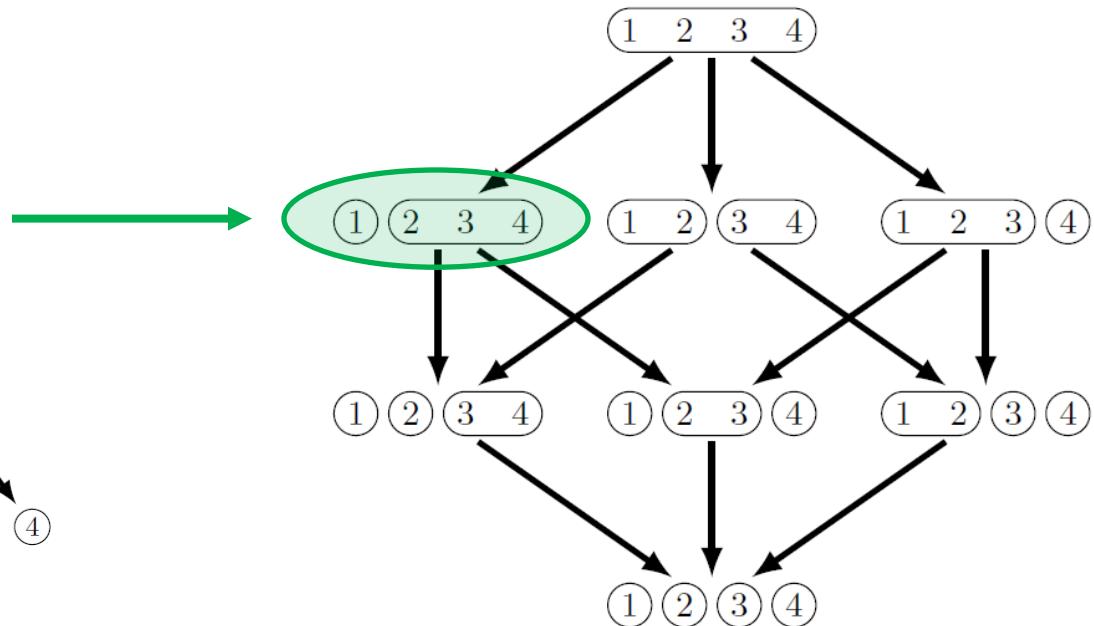


Aggregating according to a Total Order

Admissible Parts
(time intervals)

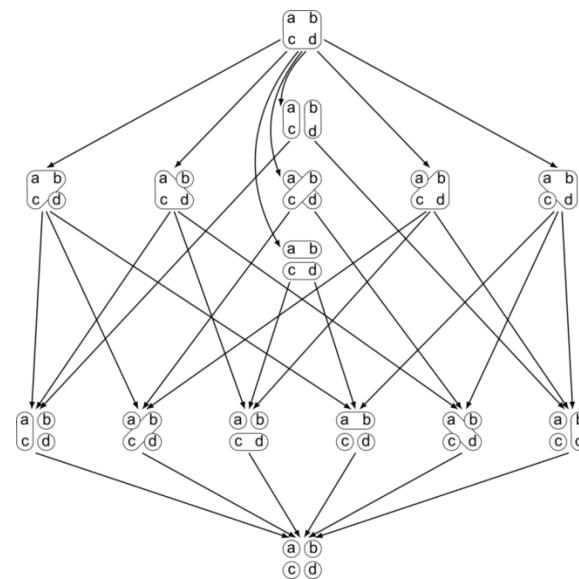


Admissible Partitions
(interval sequences)

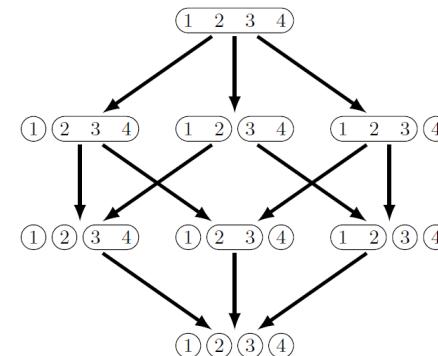


Complexity of Algebraic Structures

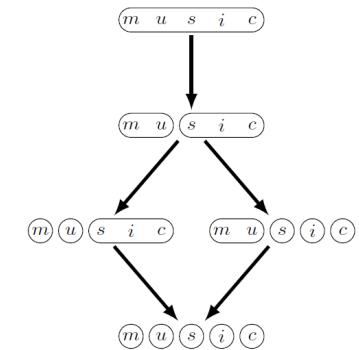
Non-constrained
partitions



Admissible partitions
according to a **total order**



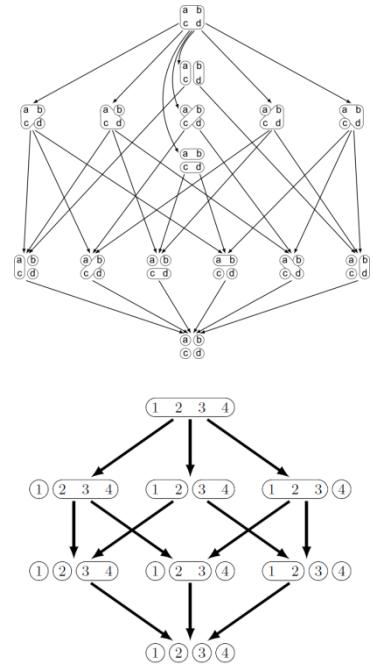
Admissible partitions
according to a **hierarchy**



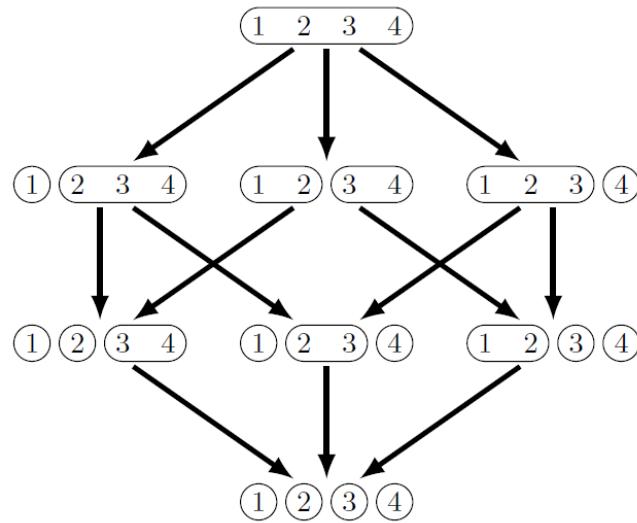
← →
Less constrained More constrained
More complex Less complex

My Approach

- P0** To characterize the aggregation process
→ The algebra of possible partitions
- P1** To preserve the system's semantics
→ A constrained partitioning method
To aggregate according to several dimension
→ Some constraints expressing the system's topology
- P2** To evaluate and compare the representations
→ Some measures of complexity and information
To offer several granularity levels
→ The optimization of a compromise
- P3** To compute the best representations
→ A generic algorithm of constrained optimization

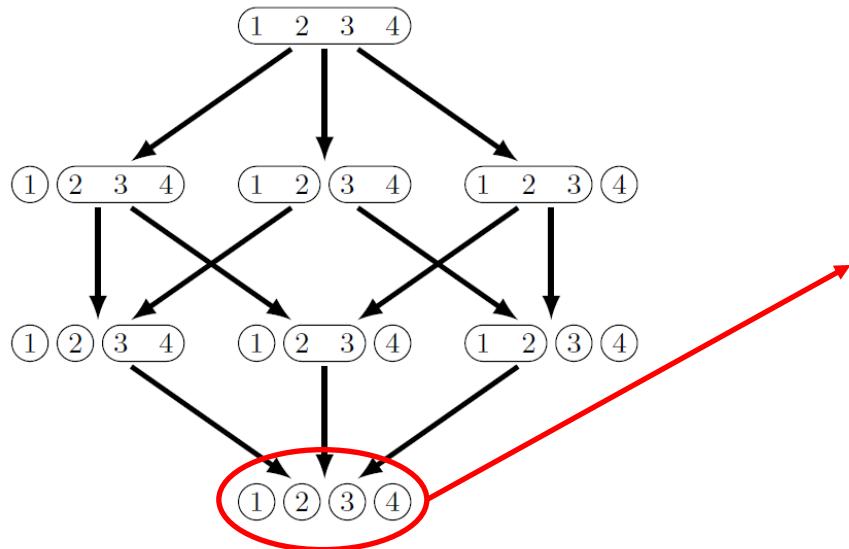


Objectives and Difficulties



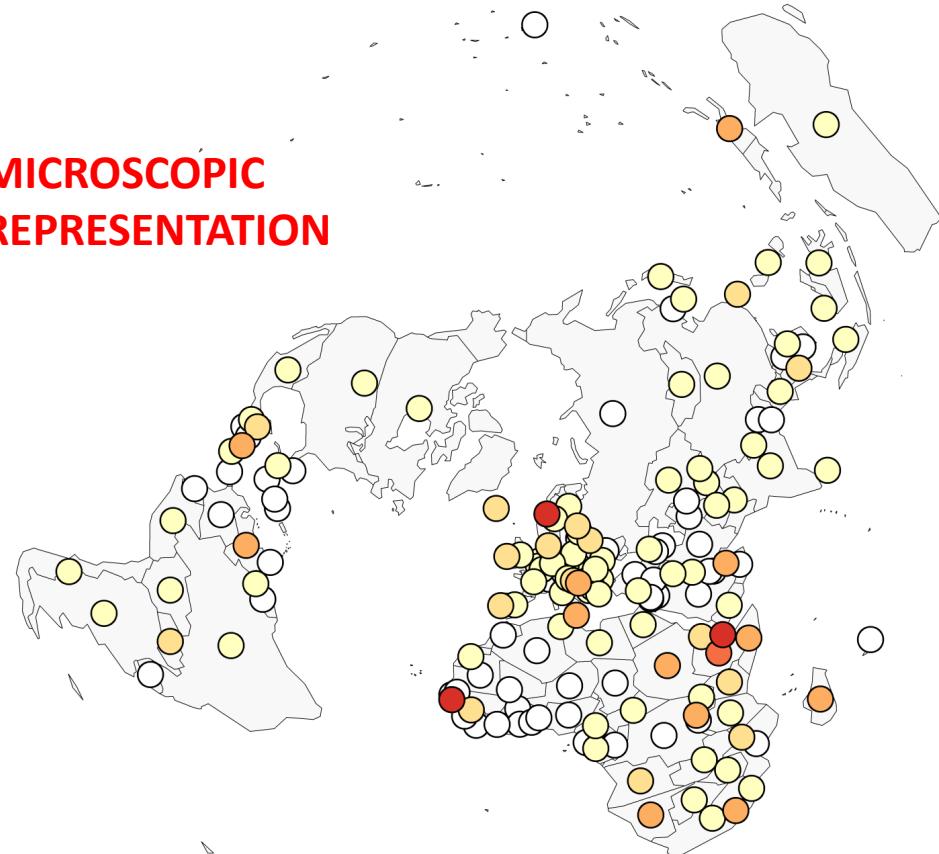
Which **admissible partition**
is the **best partition** for a
particular dataset?

Objectives and Difficulties



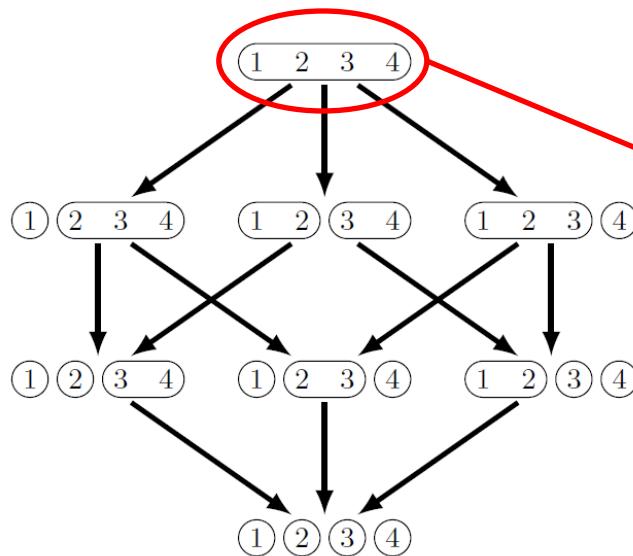
**MICROSCOPIC
REPRESENTATION**

Which **admissible partition**
is the **best partition** for a
particular dataset?



→ TOO COMPLEX TO SCALE-UP

Objectives and Difficulties



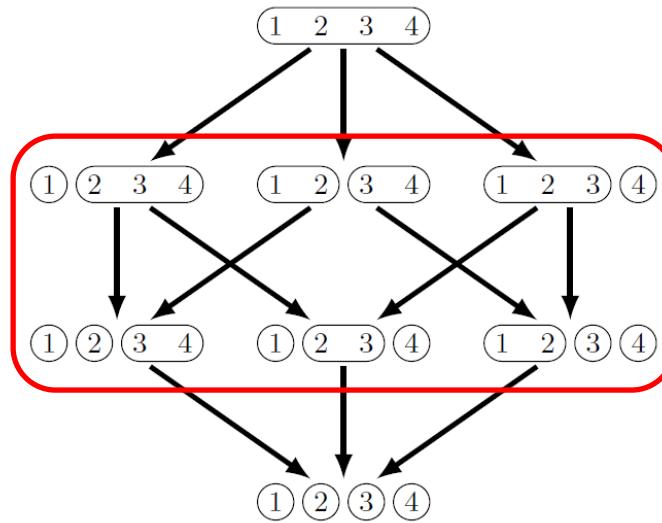
ENTIRELY AGGREGATED
REPRESENTATION



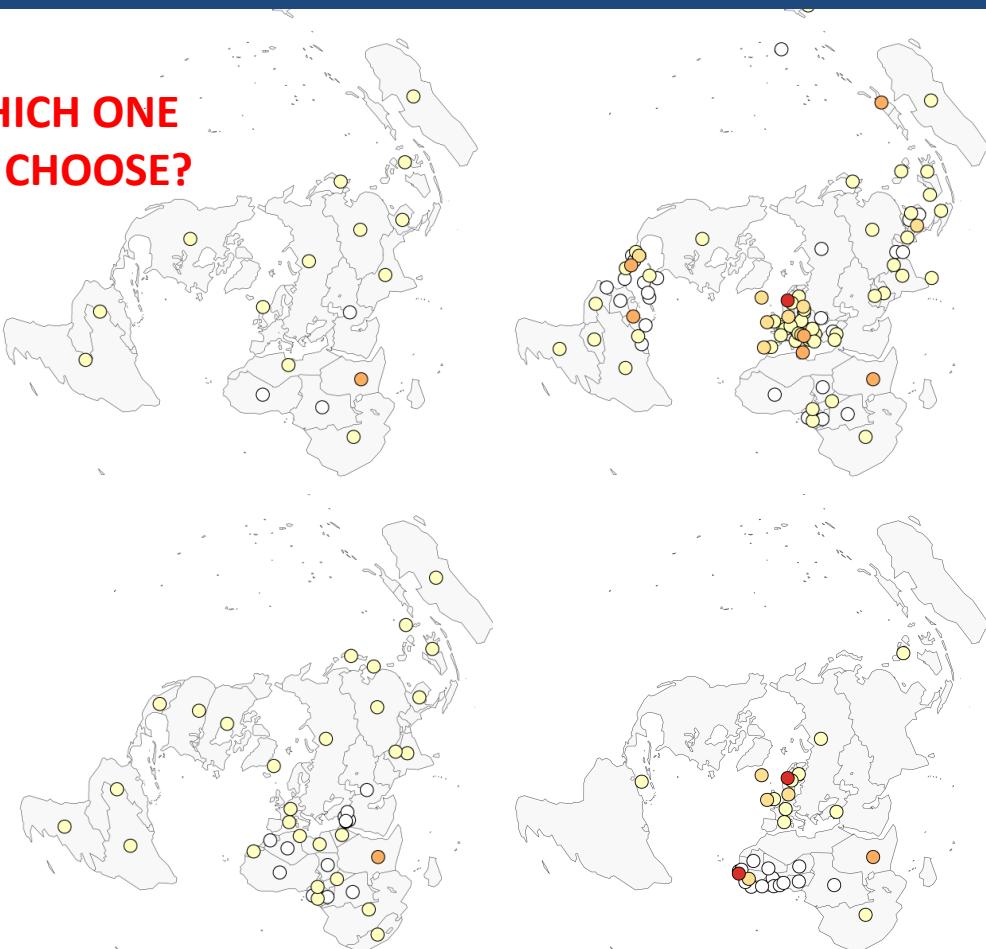
Which **admissible partition**
is the **best partition** for a
particular dataset?

→ GIVES TOO FEW INFORMATION

Objectives and Difficulties

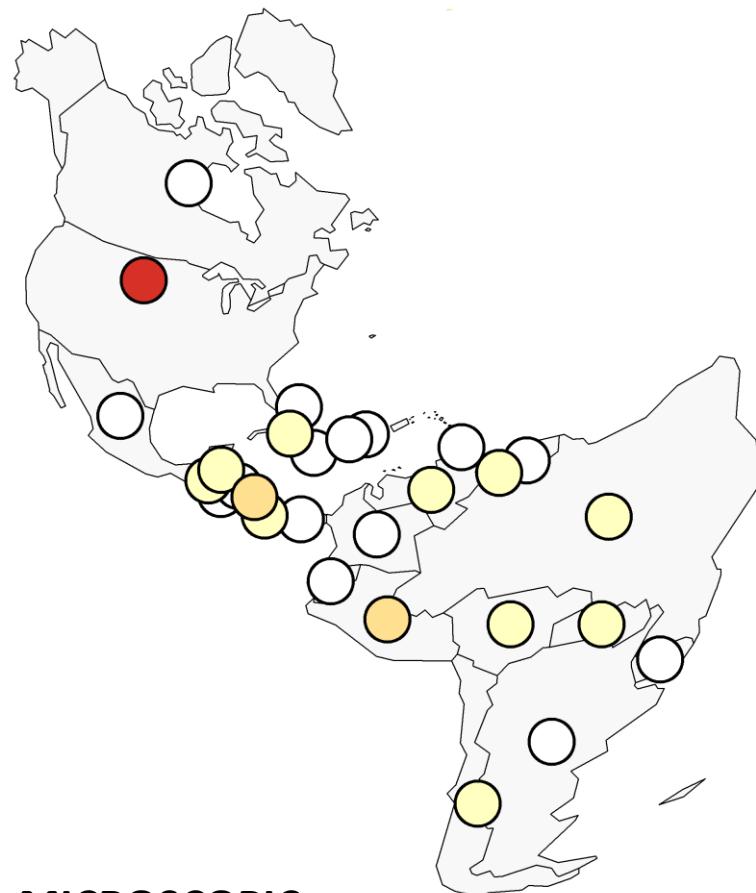


WHICH ONE
TO CHOOSE?

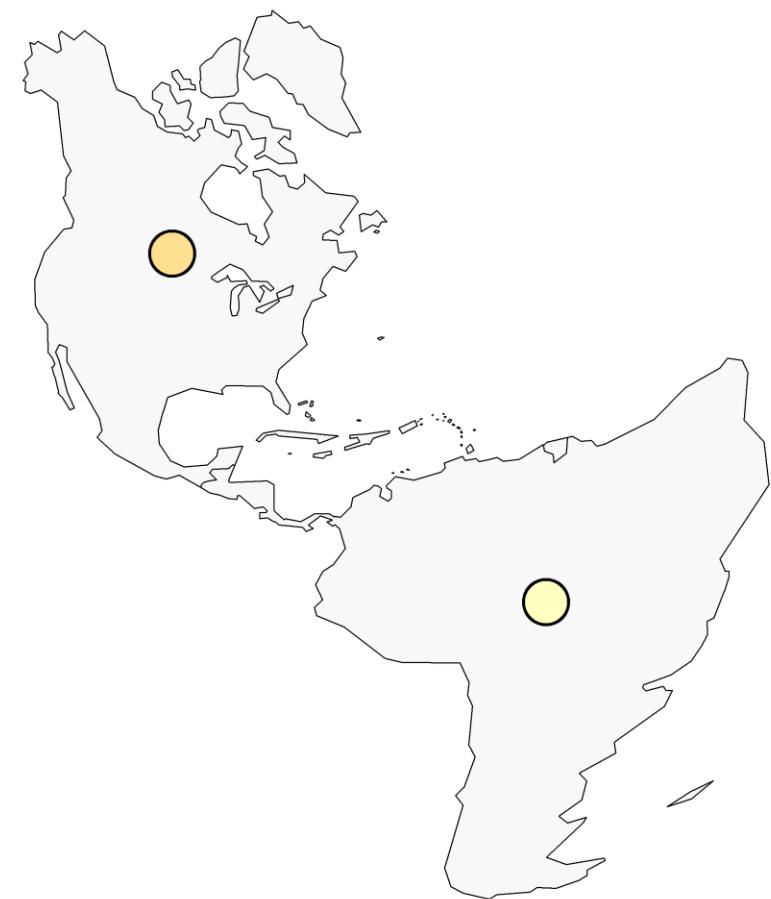


Which **admissible partition**
is the **best partition** for a
particular dataset?

Complexity and Information



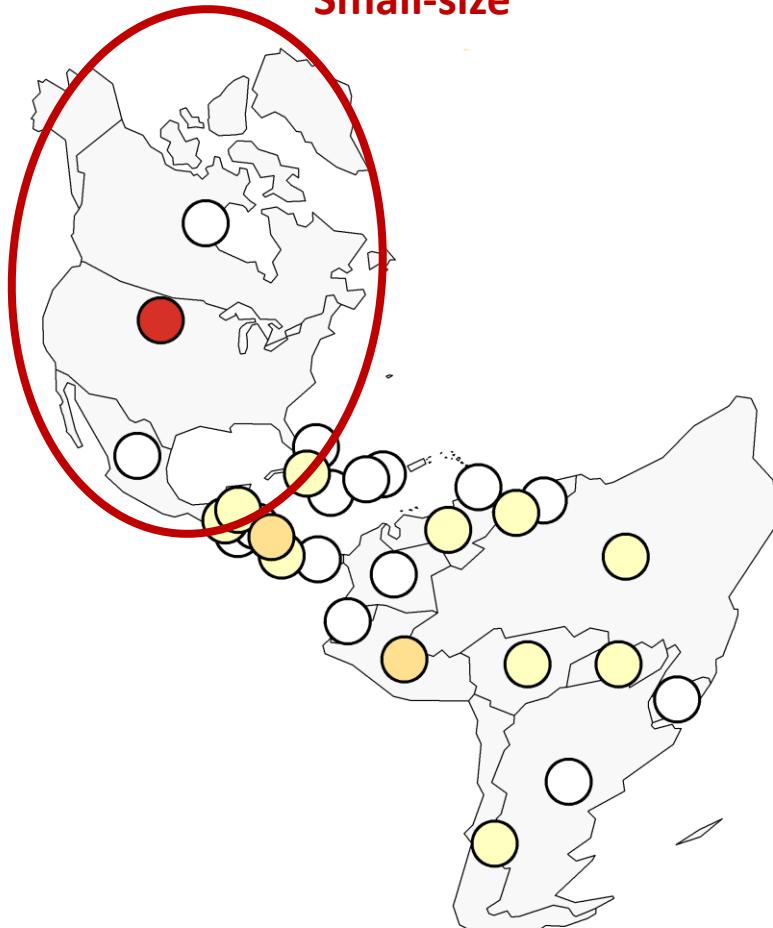
MICROSCOPIC
REPRESENTATION



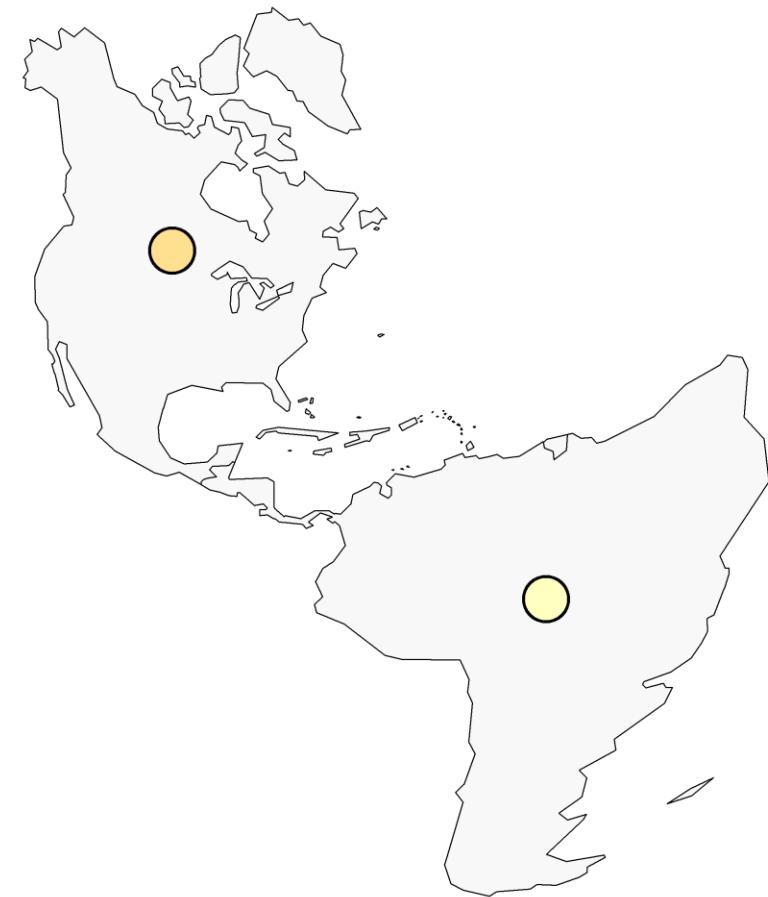
AGGREGATED
REPRESENTATION

Complexity and Information

Heterogeneous
Small-size

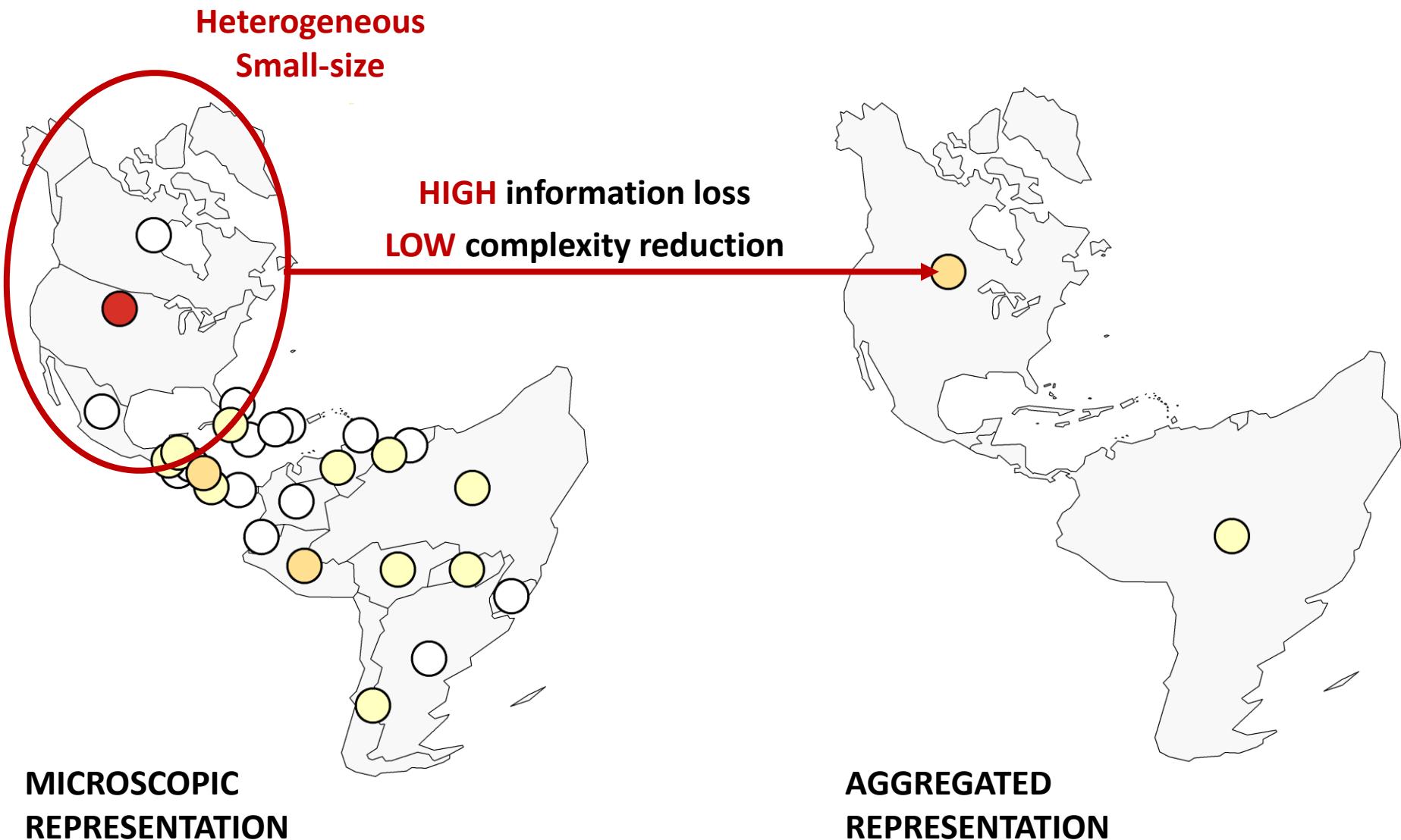


MICROSCOPIC
REPRESENTATION

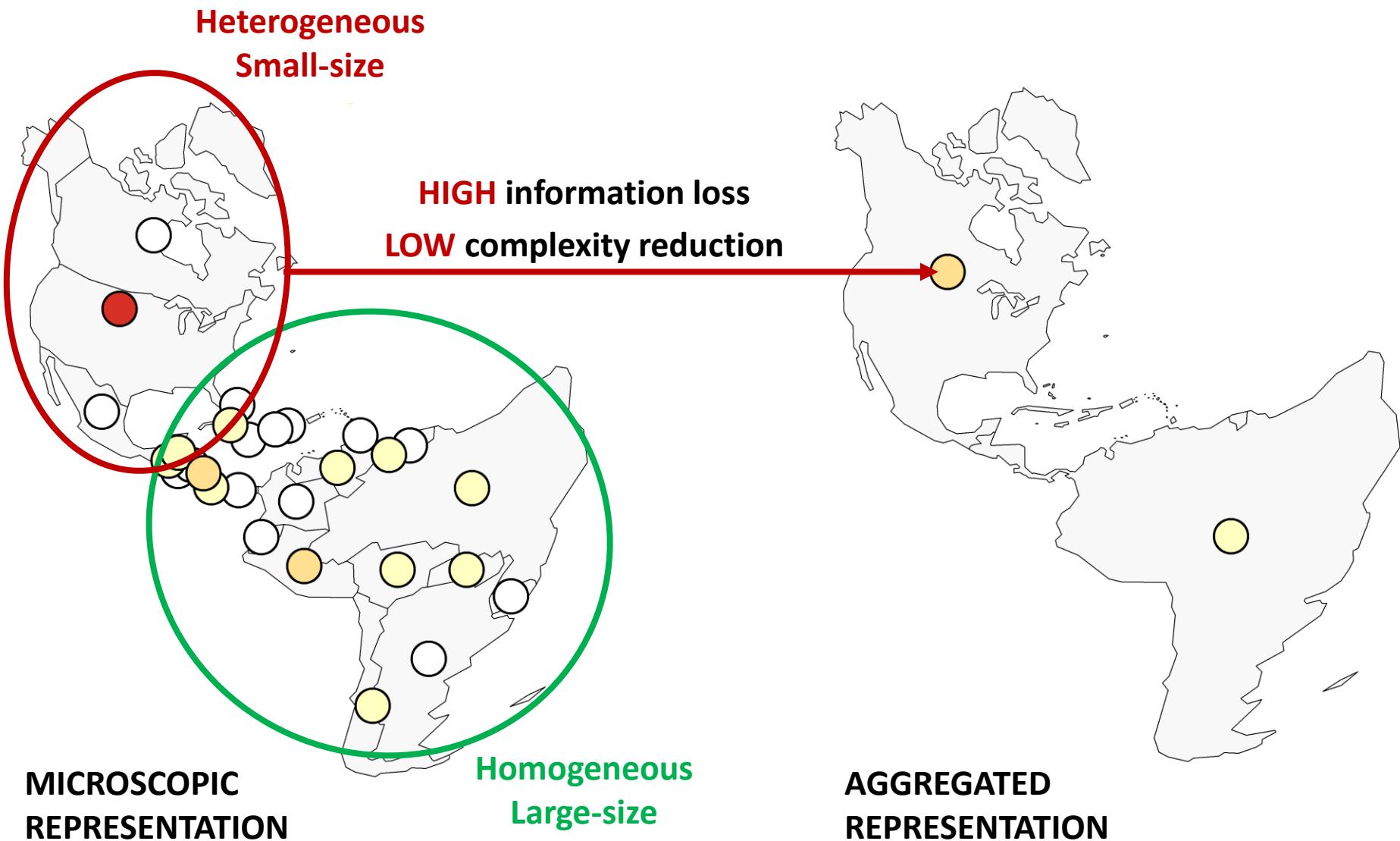


AGGREGATED
REPRESENTATION

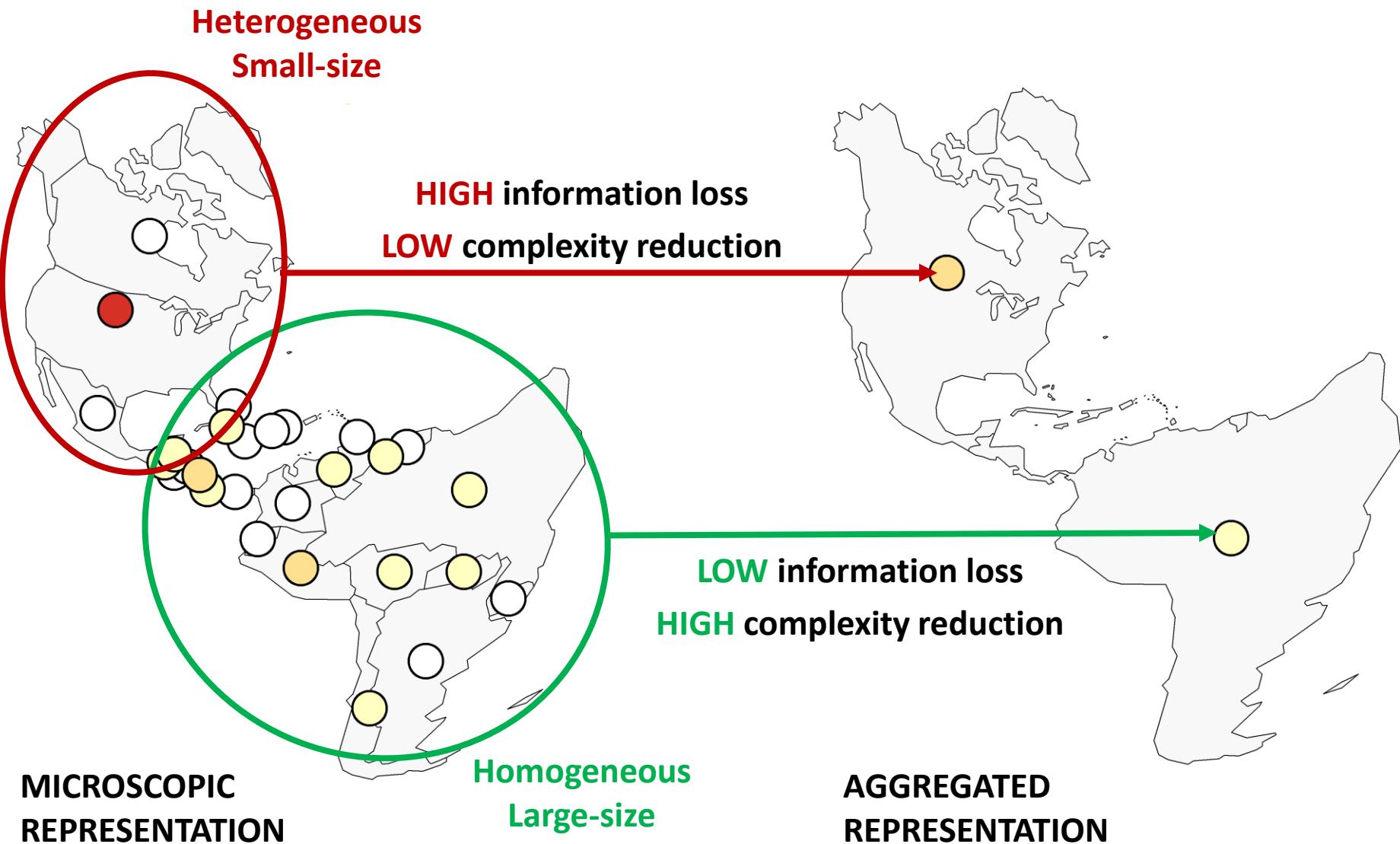
Complexity and Information



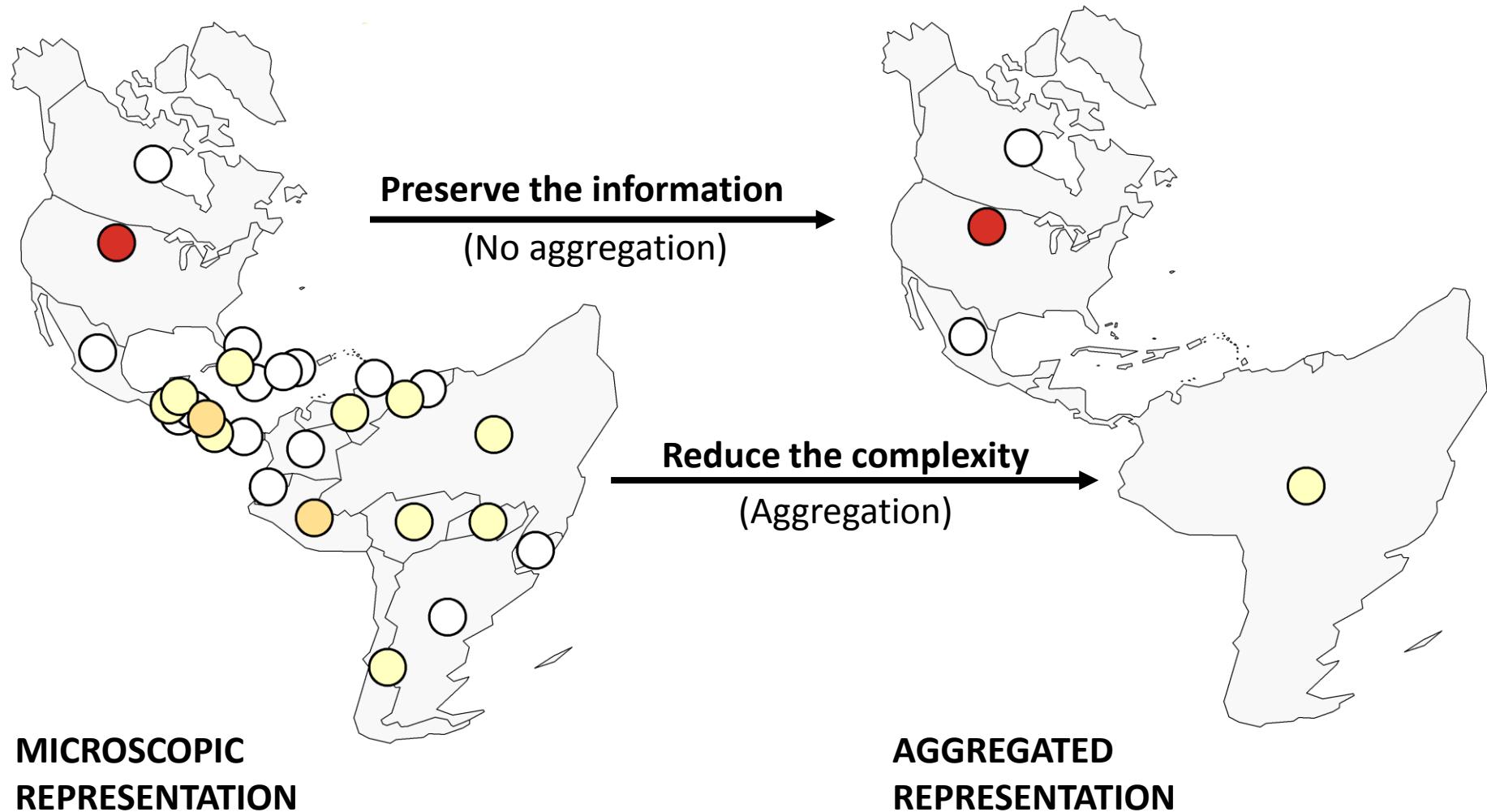
Complexity and Information



Complexity and Information



Complexity and Information



Quality Measures

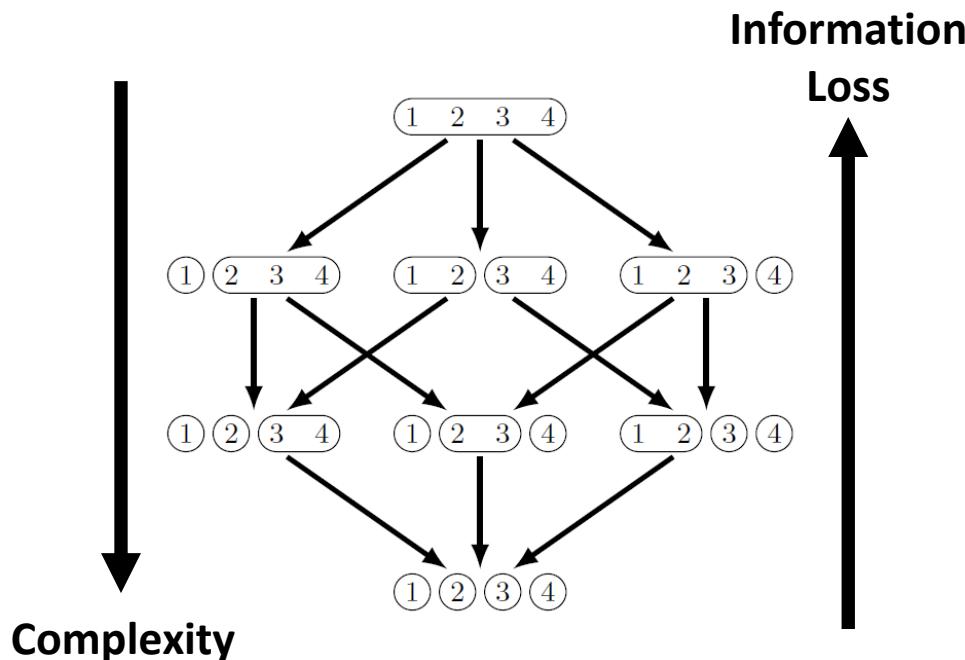
[Lamarche-Perrin *et al.*, ECCS 2012]

Complexity depends on the **tasks** we want to fulfill and the **description tools** that are available to do so

[Bonabeau and Dossal, 1997]

Information loss is measured by the **KL-divergence** between two probabilistic distributions

[Kullback et Leibler, 1951]



Quality Measures

[Lamarche-Perrin *et al.*, ECCS 2012]

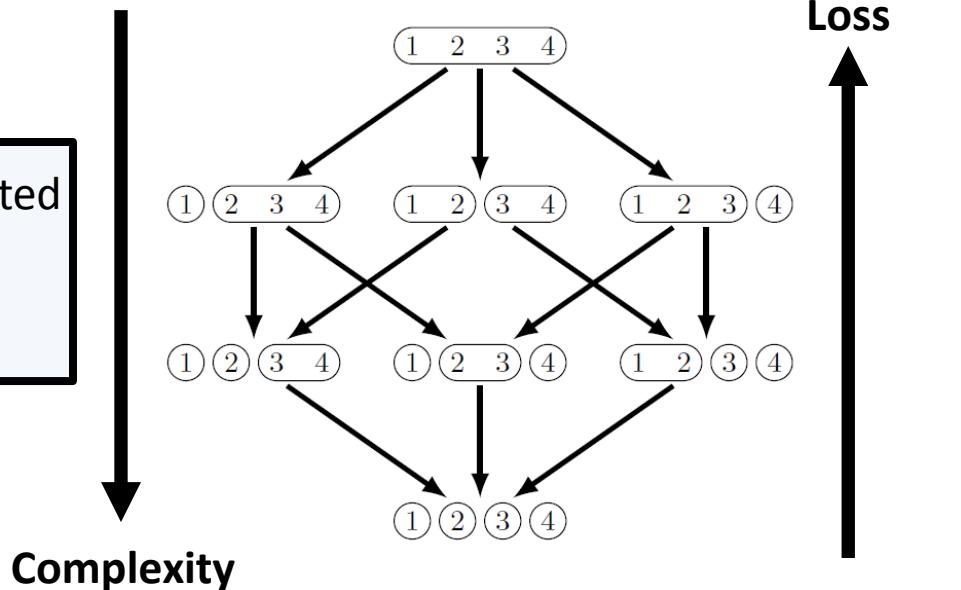
Complexity depends on the **tasks** we want to fulfill and the **description tools** that are available to do so

[Bonabeau and Dossal, 1997]

Information loss is measured by the **KL-divergence** between two probabilistic distributions

[Kullback et Leibler, 1951]

Number of represented aggregates:
 $T(\mathcal{X}) = |\mathcal{X}|$



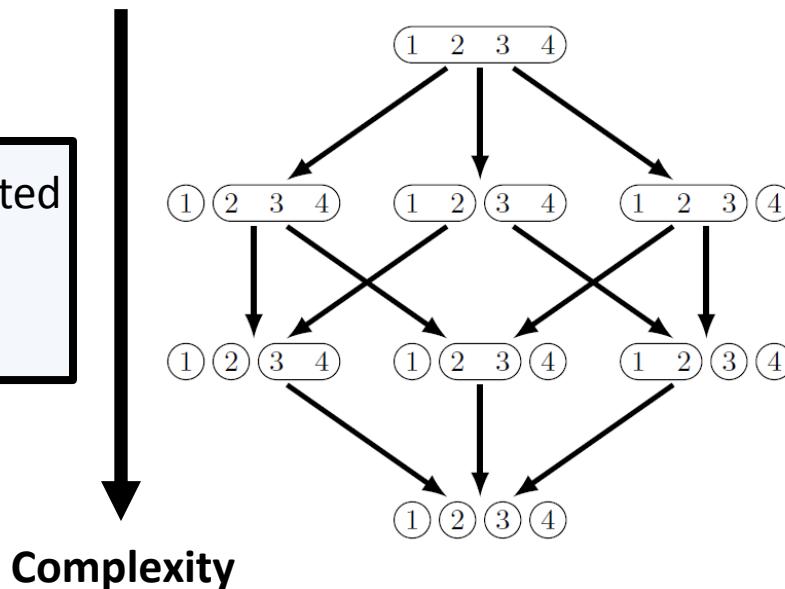
Quality Measures

[Lamarche-Perrin *et al.*, ECCS 2012]

Complexity depends on the **tasks** we want to fulfill and the **description tools** that are available to do so

[Bonabeau and Dossal, 1997]

Number of represented aggregates:
 $T(\mathcal{X}) = |\mathcal{X}|$



Information loss is measured by the **KL-divergence** between two probabilistic distributions

[Kullback et Leibler, 1951]

Information

Loss

Kullback-Leibler Divergence

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

Measures Decomposability

Number of represented aggregates:

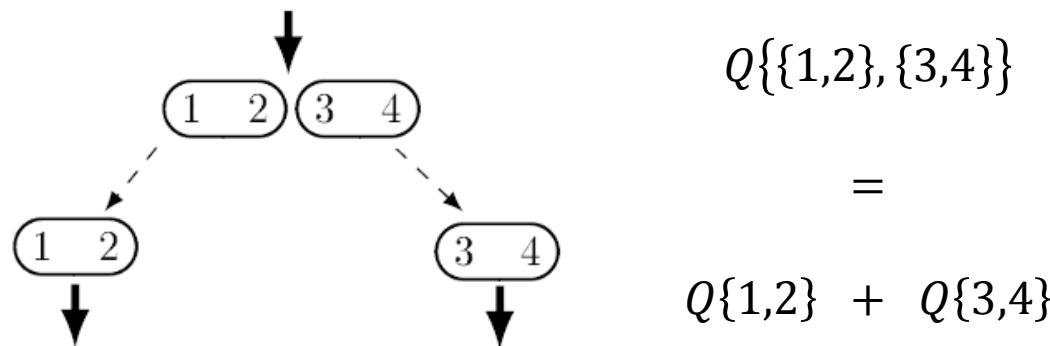
$$T(\mathcal{X}) = |\mathcal{X}|$$

Kullback-Leibler Divergence

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

Additive Decomposability: The quality of a partition is the sum of the qualities of its parts

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

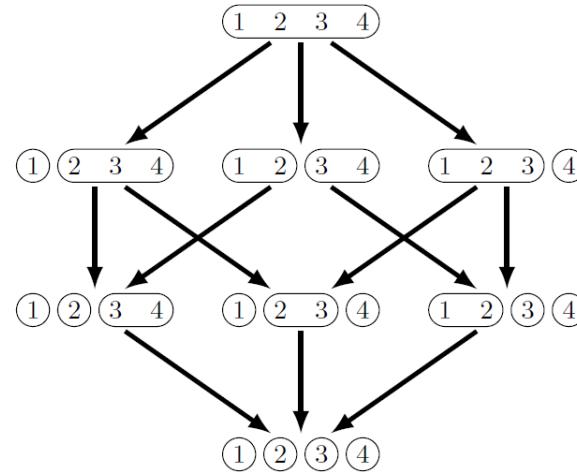
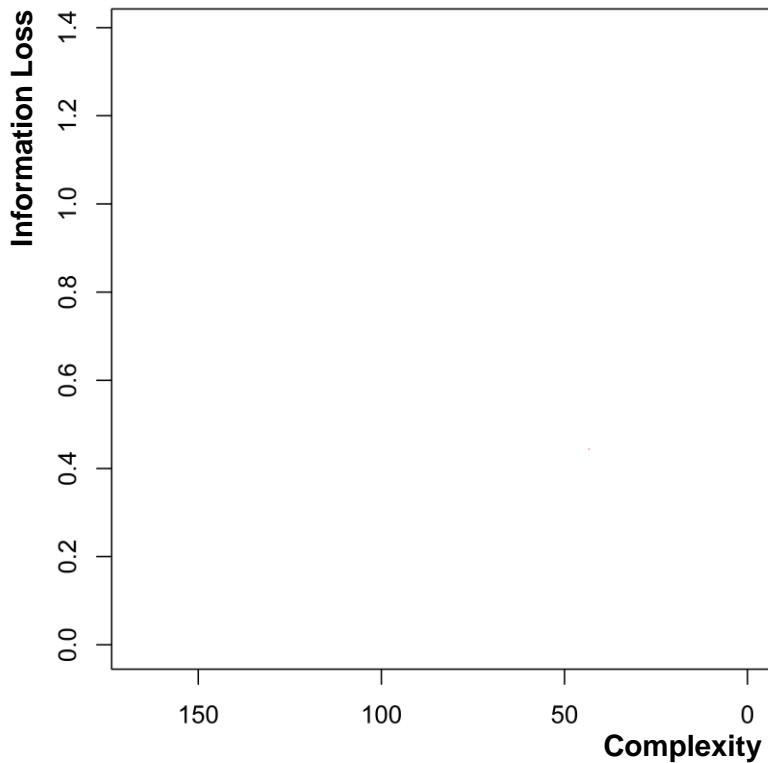


Optimizing the Partition Qualities

Two criteria that should be optimized

Compromise of Quality:

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

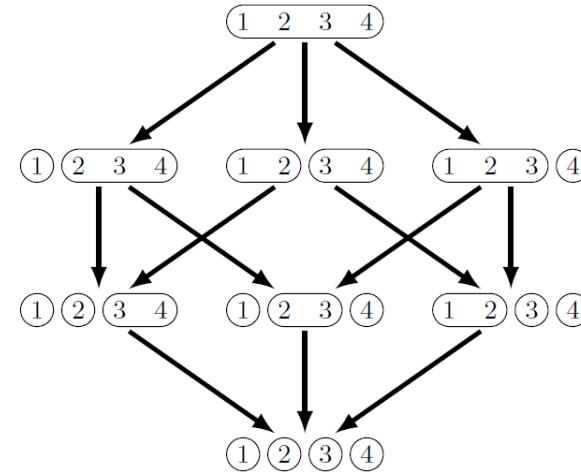
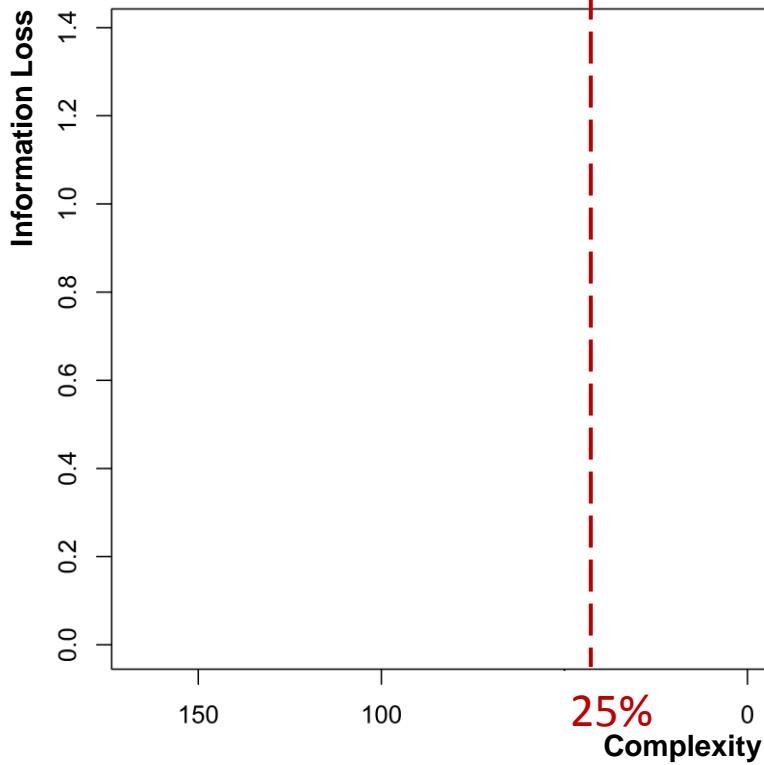


Optimizing the Partition Qualities

Two criteria that should be optimized

Compromise of Quality:

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

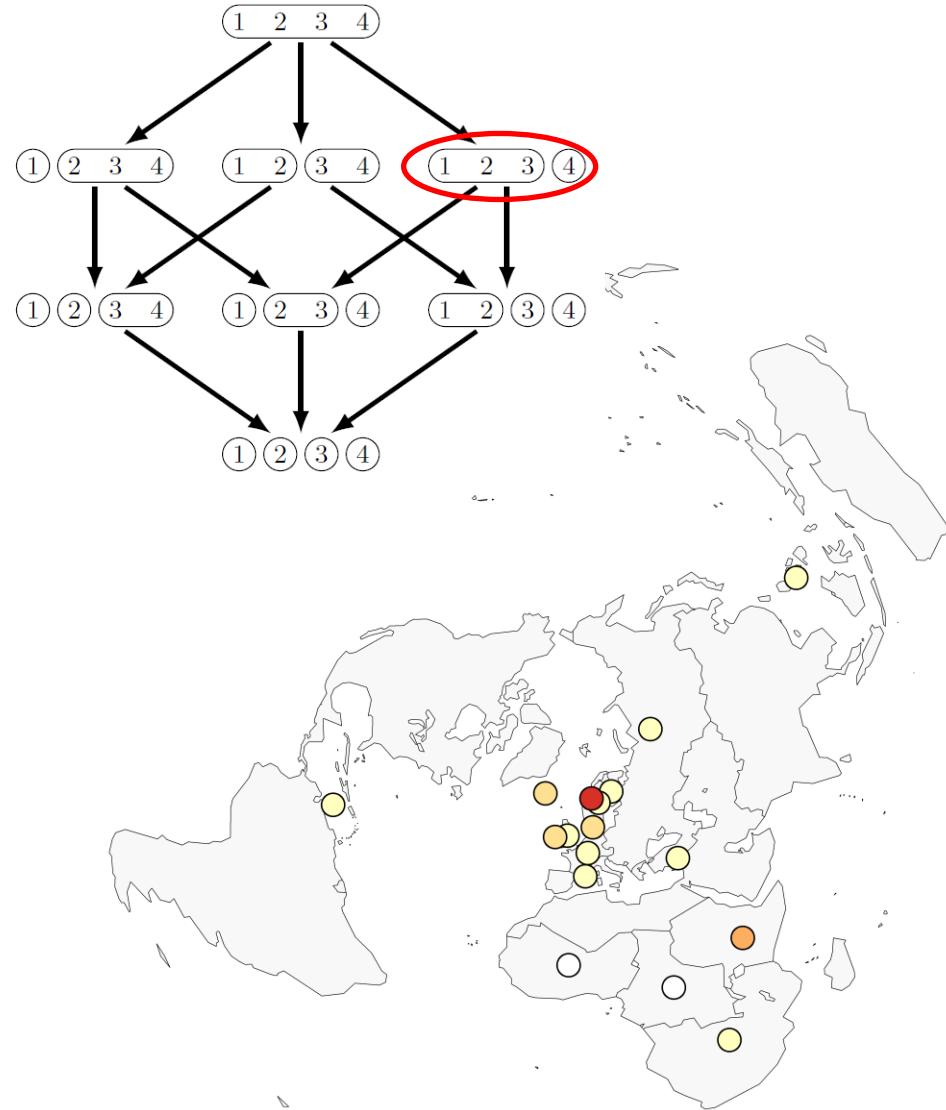
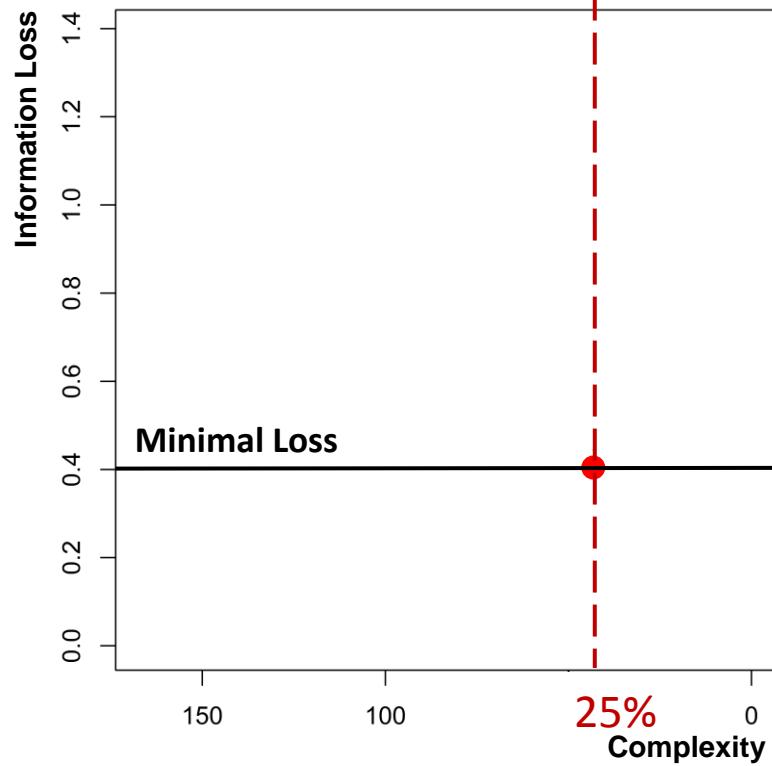


Optimizing the Partition Qualities

Two criteria that should be optimized

Compromise of Quality:

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

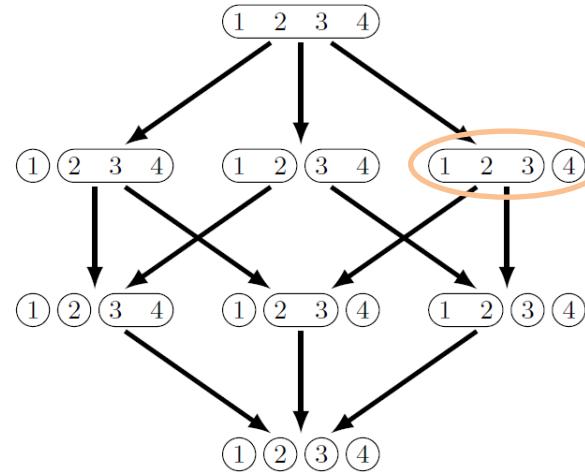
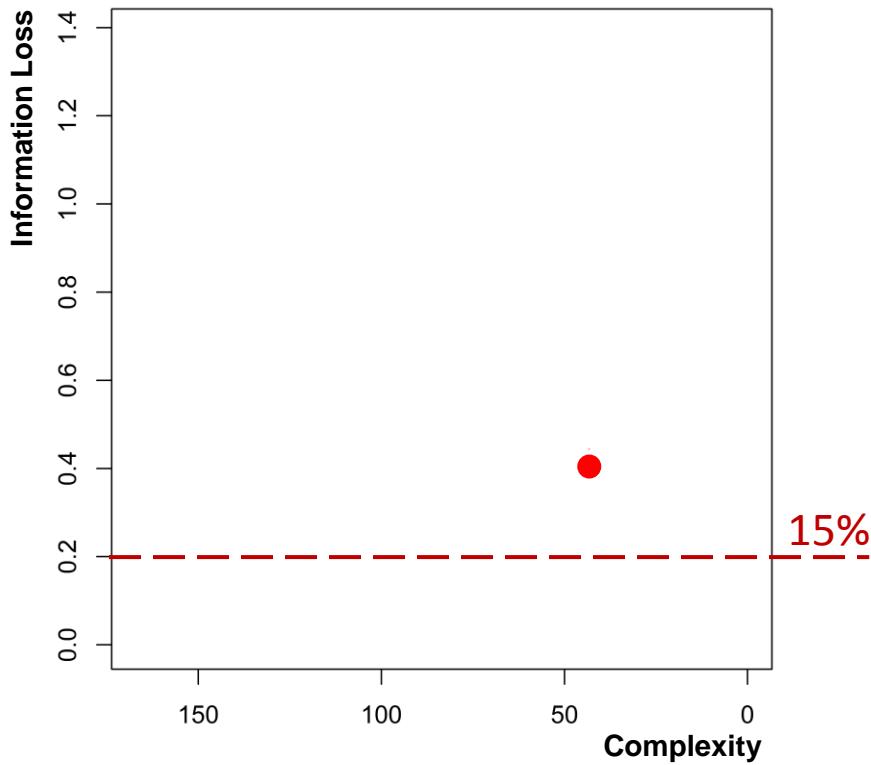


Optimizing the Partition Qualities

Two criteria that should be optimized

Compromise of Quality:

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

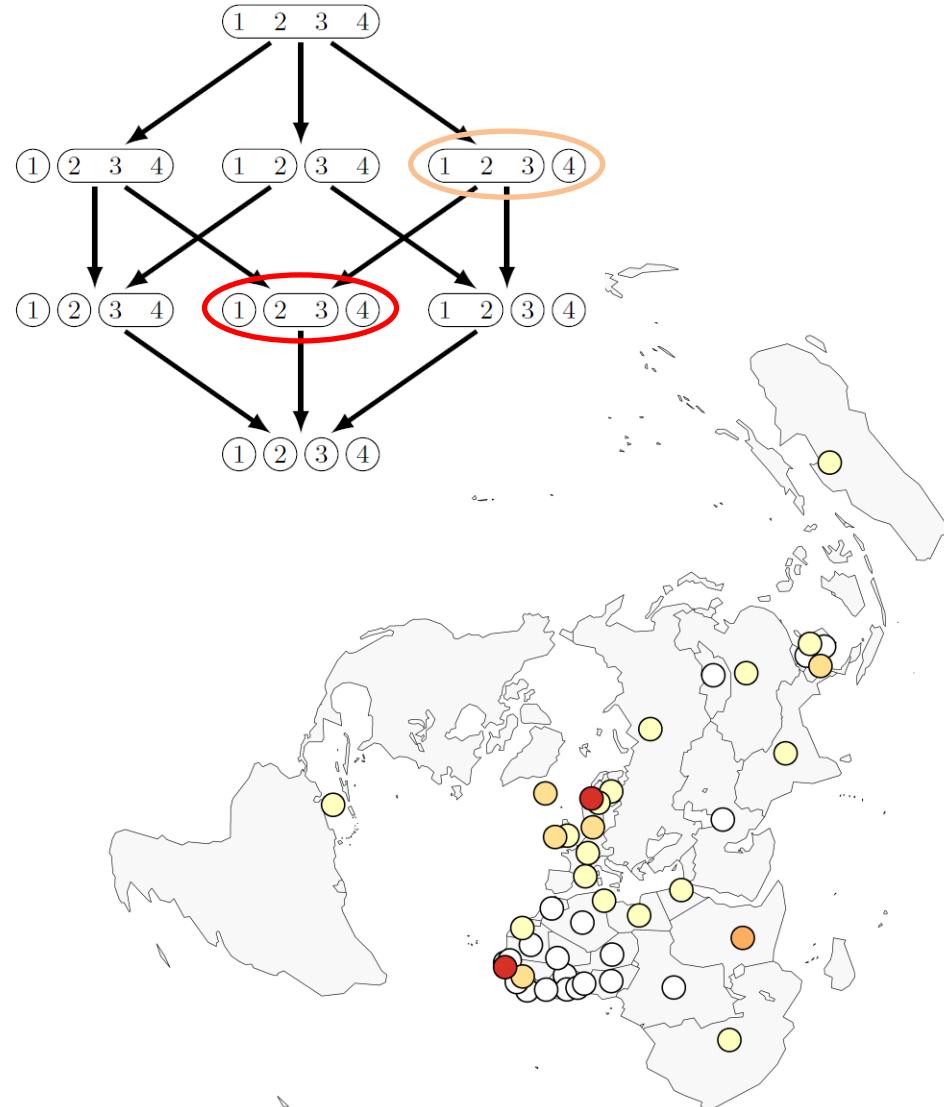
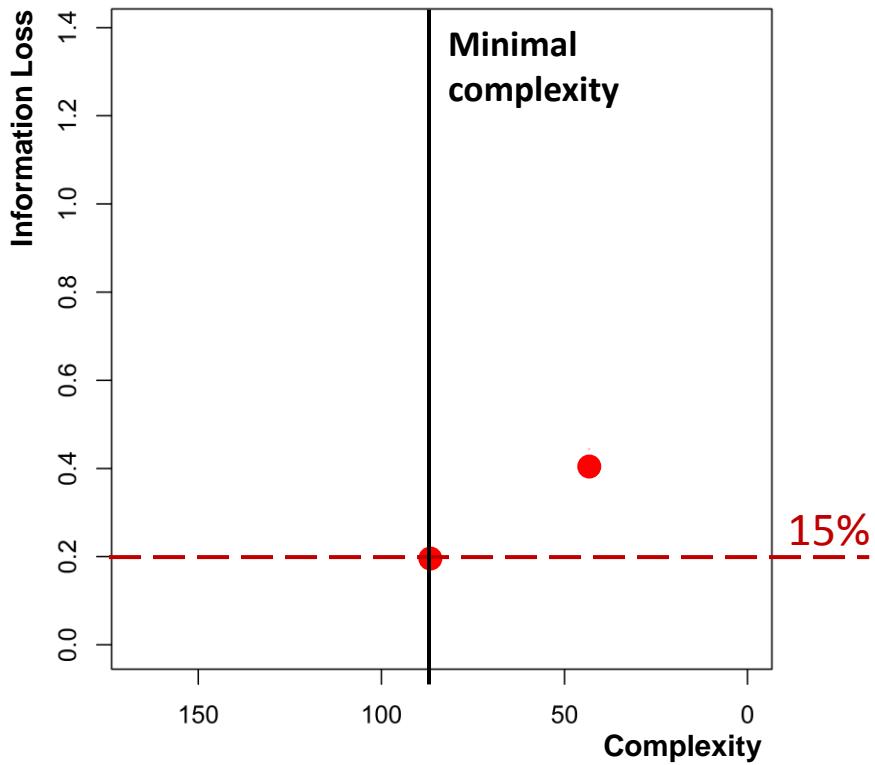


Optimizing the Partition Qualities

Two criteria that should be optimized

Compromise of Quality:

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

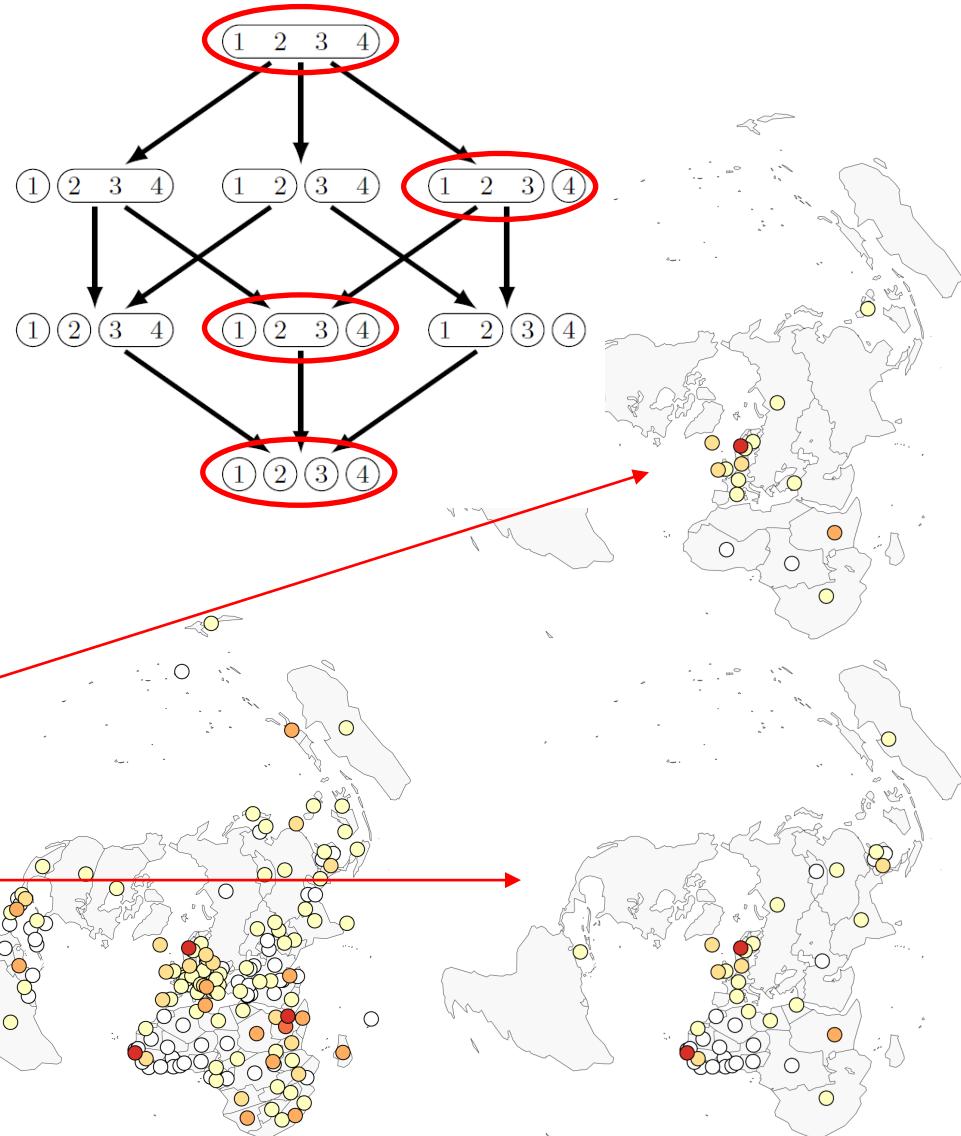
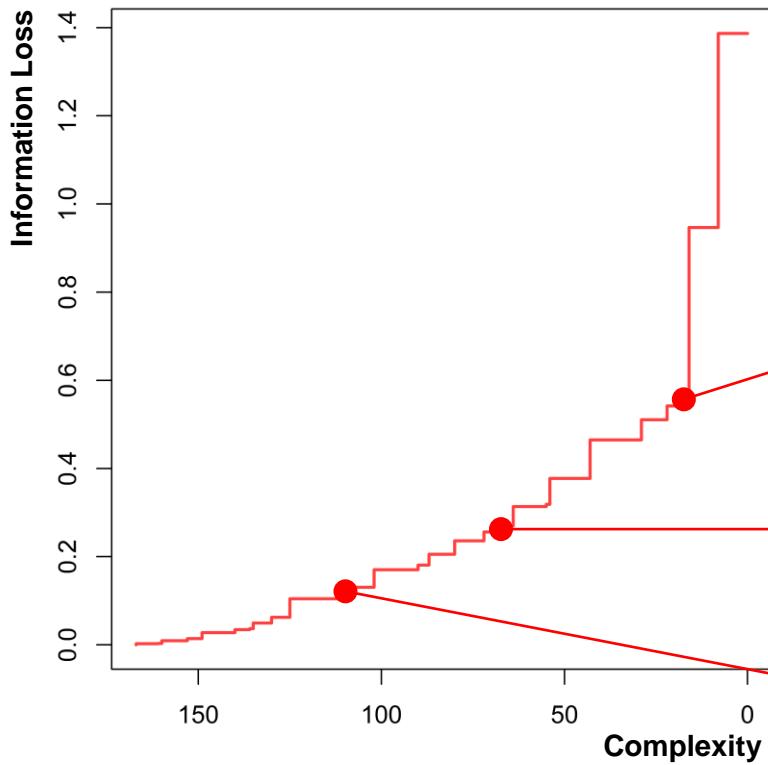


Optimizing the Partition Qualities

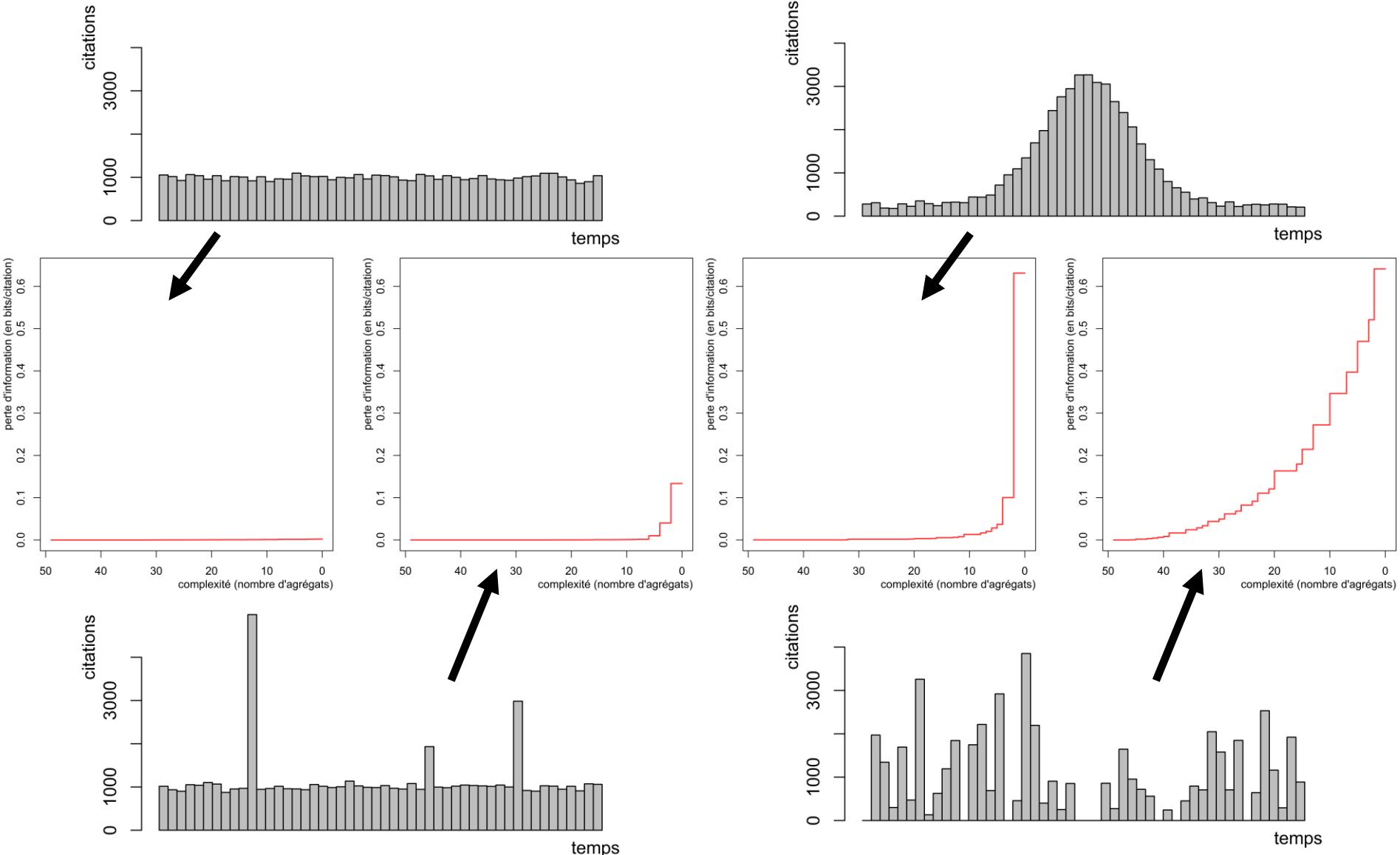
Two criteria that should be optimized

Compromise of Quality:

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

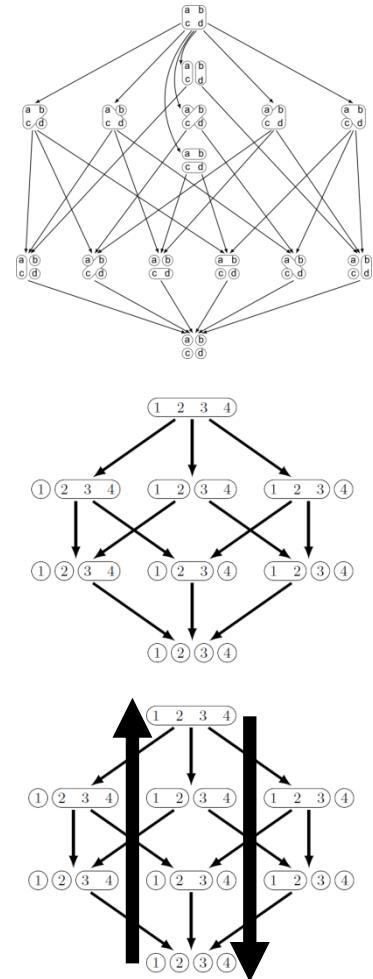


Characterizing Different Datasets



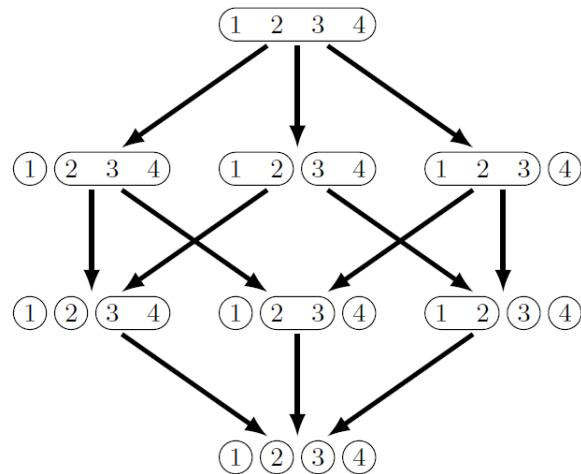
My Approach

- P0** To characterize the aggregation process
→ The algebra of possible partitions
- P1** To preserve the system's semantics
→ A constrained partitioning method
To aggregate according to several dimension
→ Some constraints expressing the system's topology
- P2** To evaluate and compare the representations
→ Some measures of complexity and information
To offer several granularity levels
→ The optimization of a compromise
- P3** To compute the best representations
→ A generic algorithm of constrained optimization



Objectives

Set of admissible partitions



Quality Measures

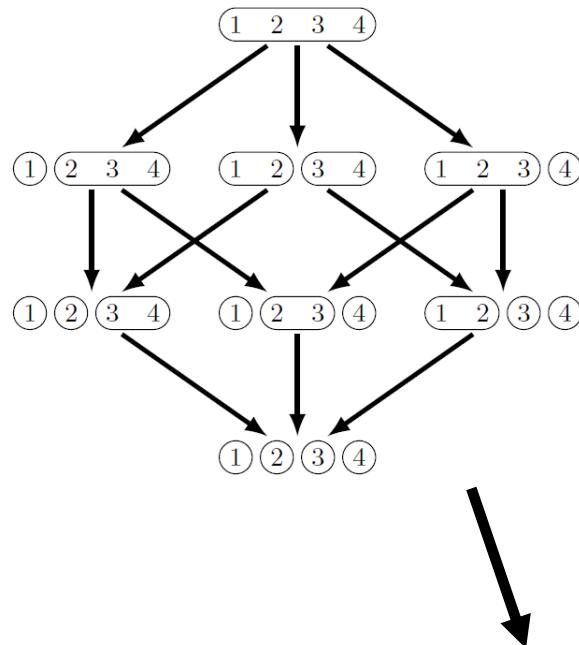
$$\Delta T(\mathcal{X}) = |\Omega| - |\mathcal{X}|$$

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

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

Objectives

Set of admissible partitions



Quality Measures

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

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

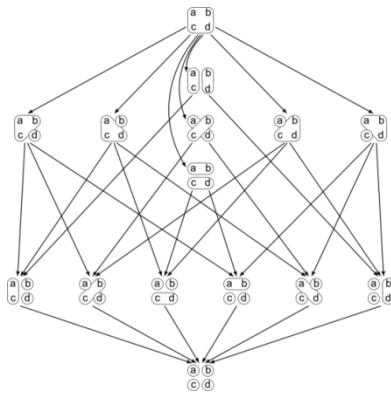
$$CQL_\alpha = \alpha \frac{\Delta T}{\Delta T_{\max}} - (1 - \alpha) \frac{D}{D_{\max}}$$



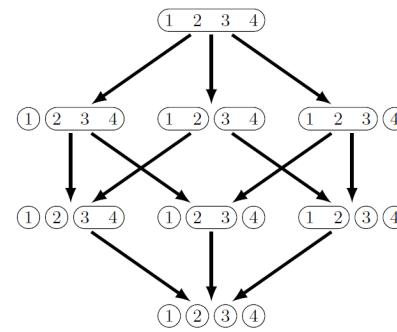
The Admissible Optimal Partitions Problem: Which admissible partitions optimize a given compromise of quality?
 → A well-known constrained optimization problem

Exponential Algorithmic Complexity

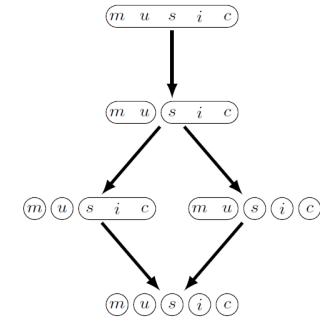
Non-constrained Set



Ordered Set



Hierarchical Set



Number of admissible partitions
(n = size of the system)

$$\Theta(e^n \log n)$$

$$\Theta(2^n)$$

$$\Theta(c^n)$$

[Berend and Tassa, 2010]

[Lamarche-Perrin *et al.*, IAT 2013]

Toward an Efficient Algorithm

- **Classical clustering techniques** uses some heuristics to find local optima [Halkidi *et al.*, 2001]

Toward an Efficient Algorithm

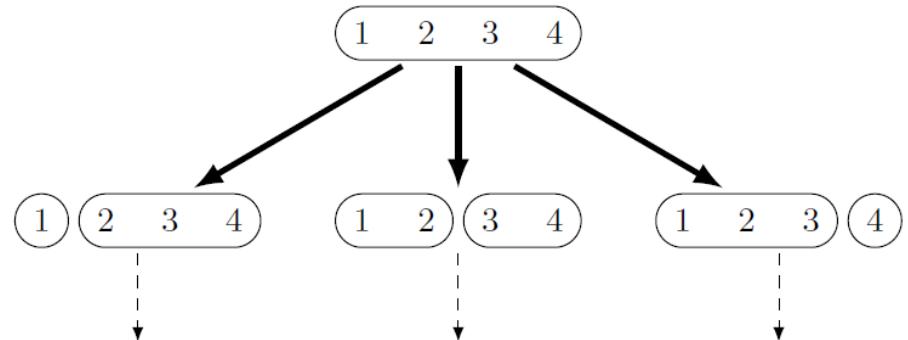
- **Classical clustering techniques** uses some heuristics to find local optima [Halkidi *et al.*, 2001]
- **In our case:**
 - The **admissibility constraints** allow to reduce the complexity of the optimization problem
 - The **algebraic properties** of the quality measures allow to efficiently compare the partitions

→ We look for the global optima

Idea of the Algorithm

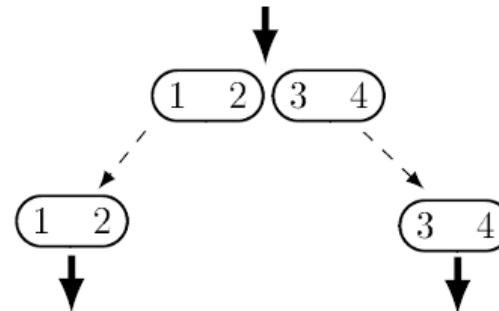
DIVIDE...

Decomposition according
to the refinement relation



...AND CONQUER

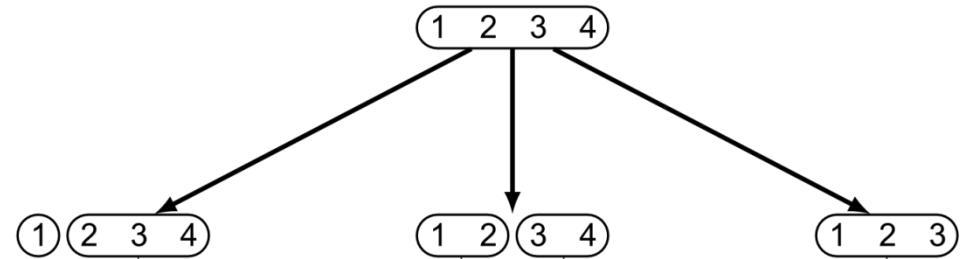
Recursion according to the
decomposability of measures



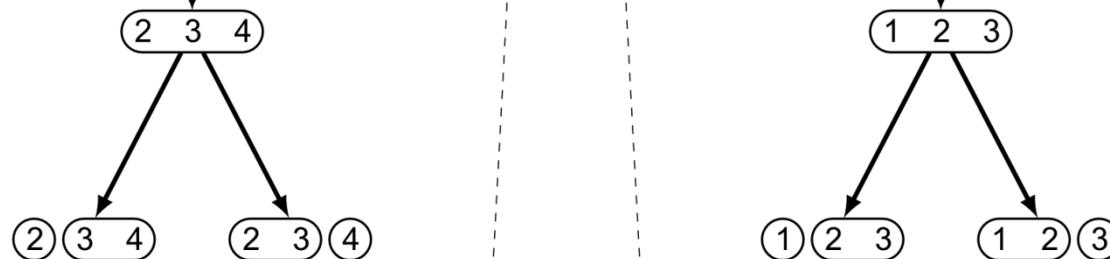
[Lamarche-Perrin *et al.*, IAT 2013]

Execution of the Algorithm

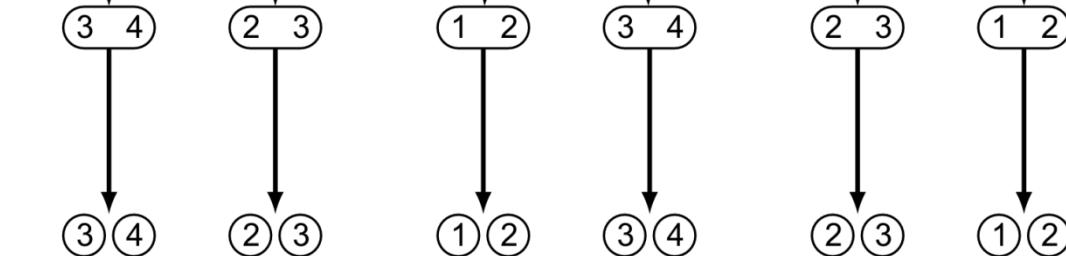
Decomposition



Recursion



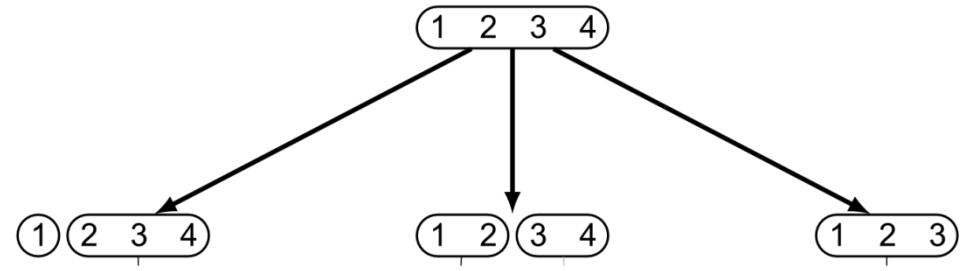
Decomposition



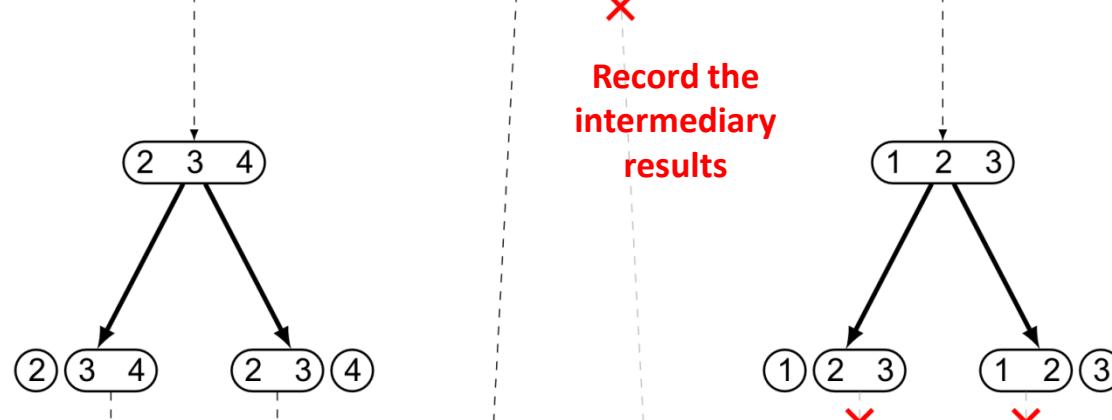
Decomposition

Execution of the Algorithm

Decomposition

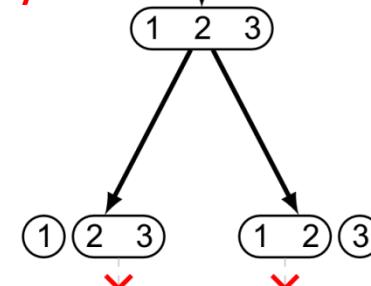


Recursion

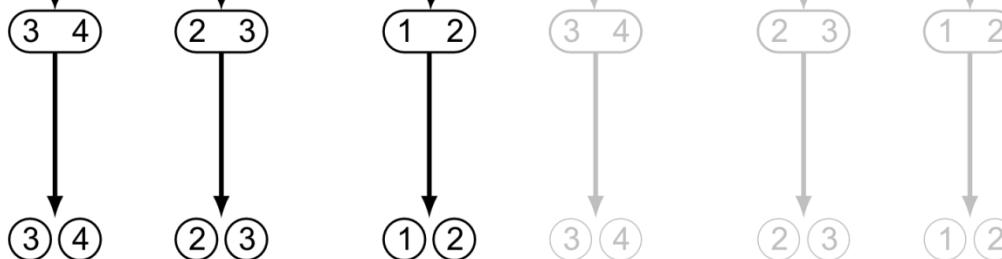


Record the intermediary results

Decomposition



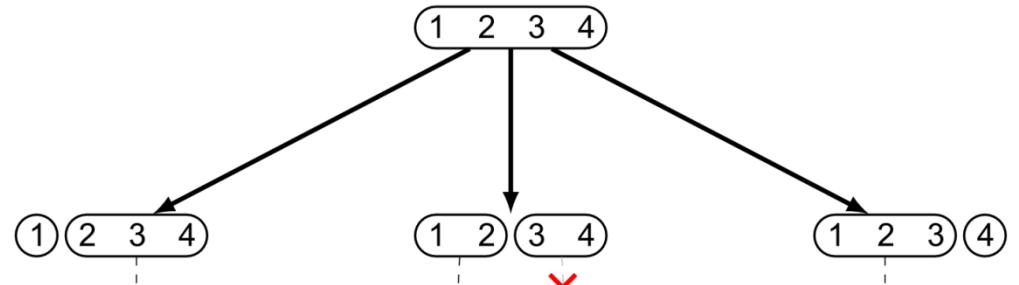
Recursion



Decomposition

Execution of the Algorithm

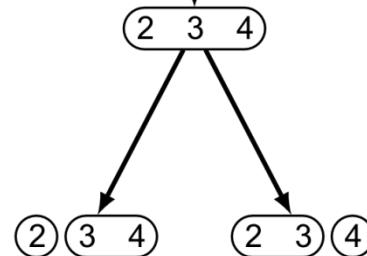
Decomposition



Recursion

Record the intermediary results

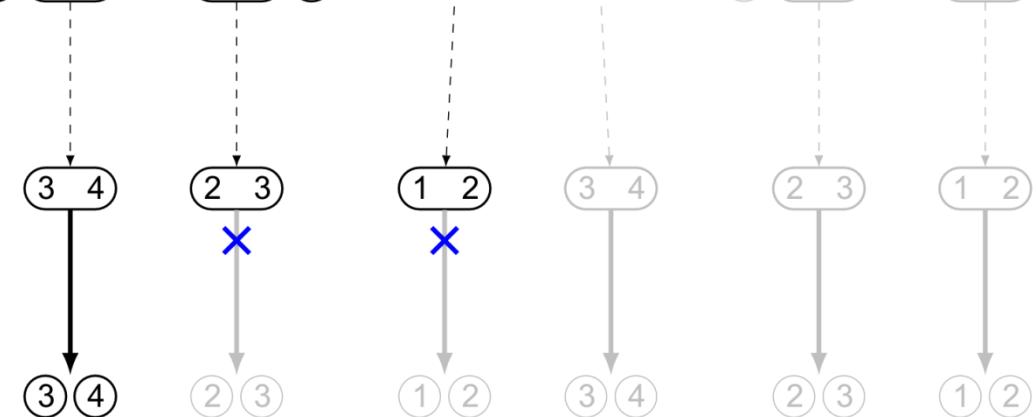
Decomposition



Recursion

Avoid redundant evaluation within the algebraic structure

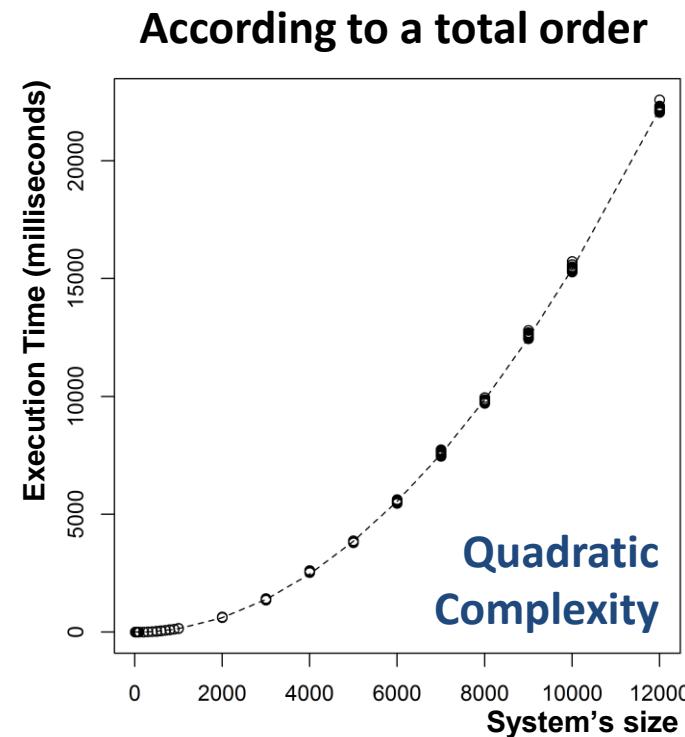
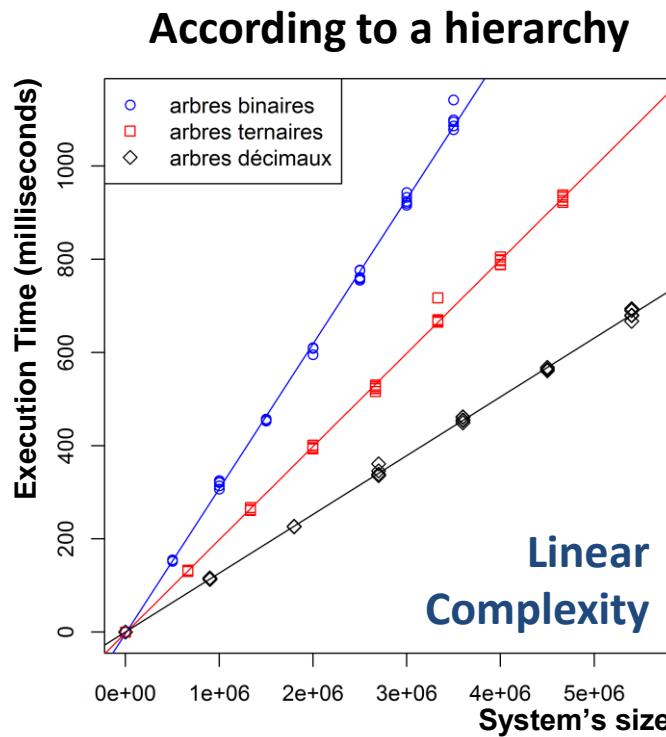
Decomposition



Complexity of the Algorithm

The algorithm spatial and temporal complexity directly depend on the algebraic properties of the set of admissible partitions

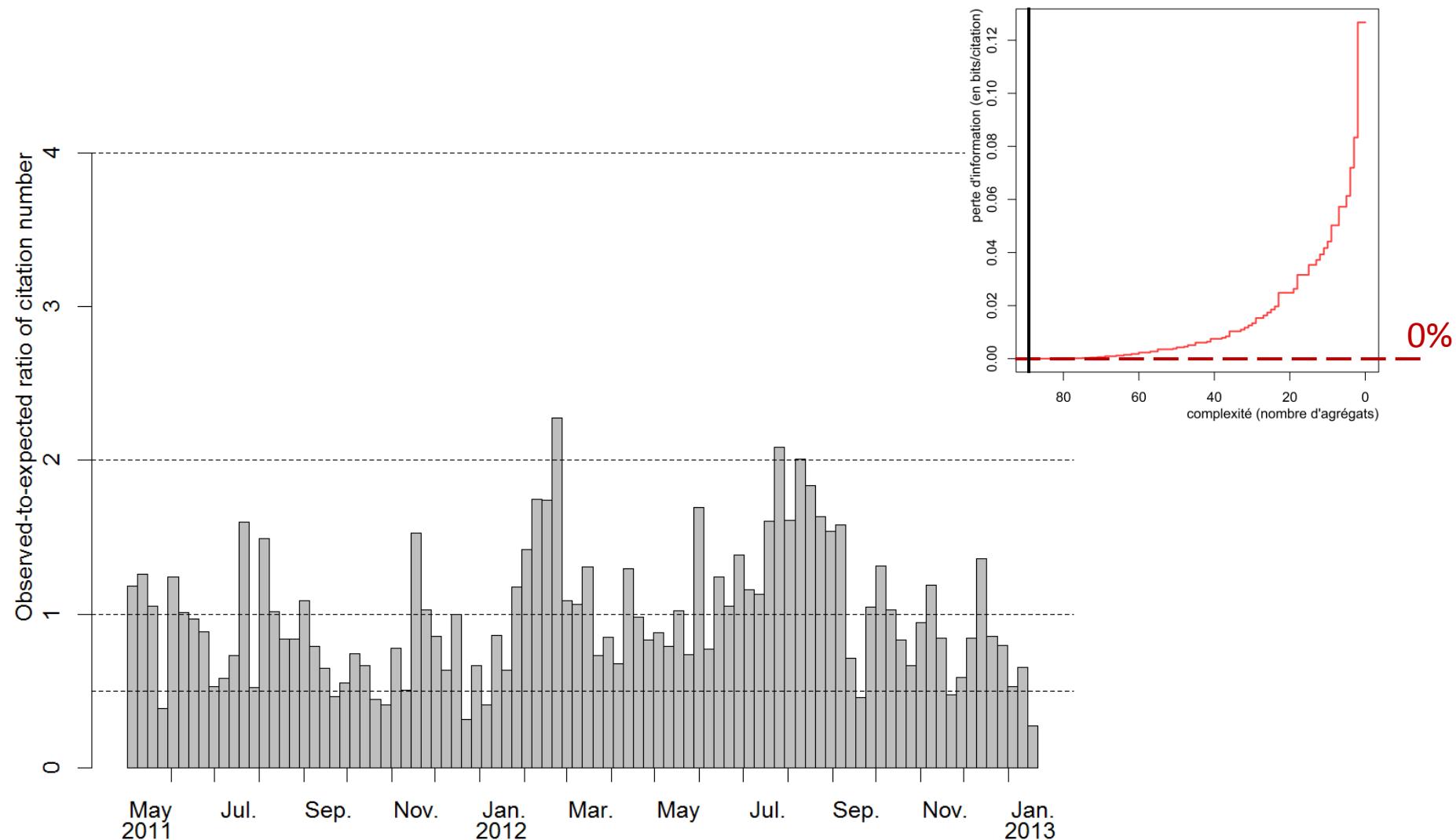
the more constrained, the less complex



EXPERIMENTS AND RESULTS

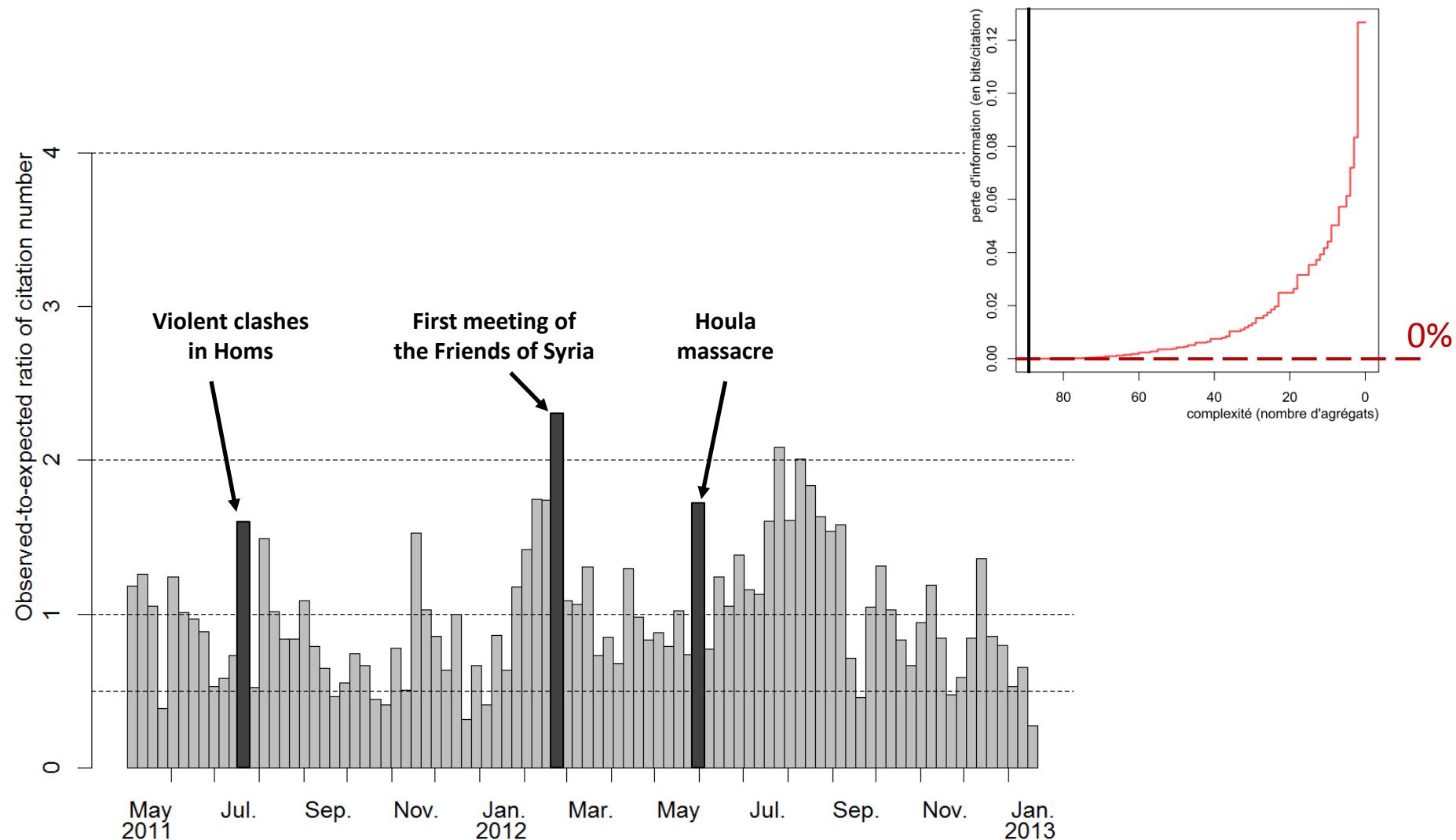
The Syrian civil war according to LE MONDE

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



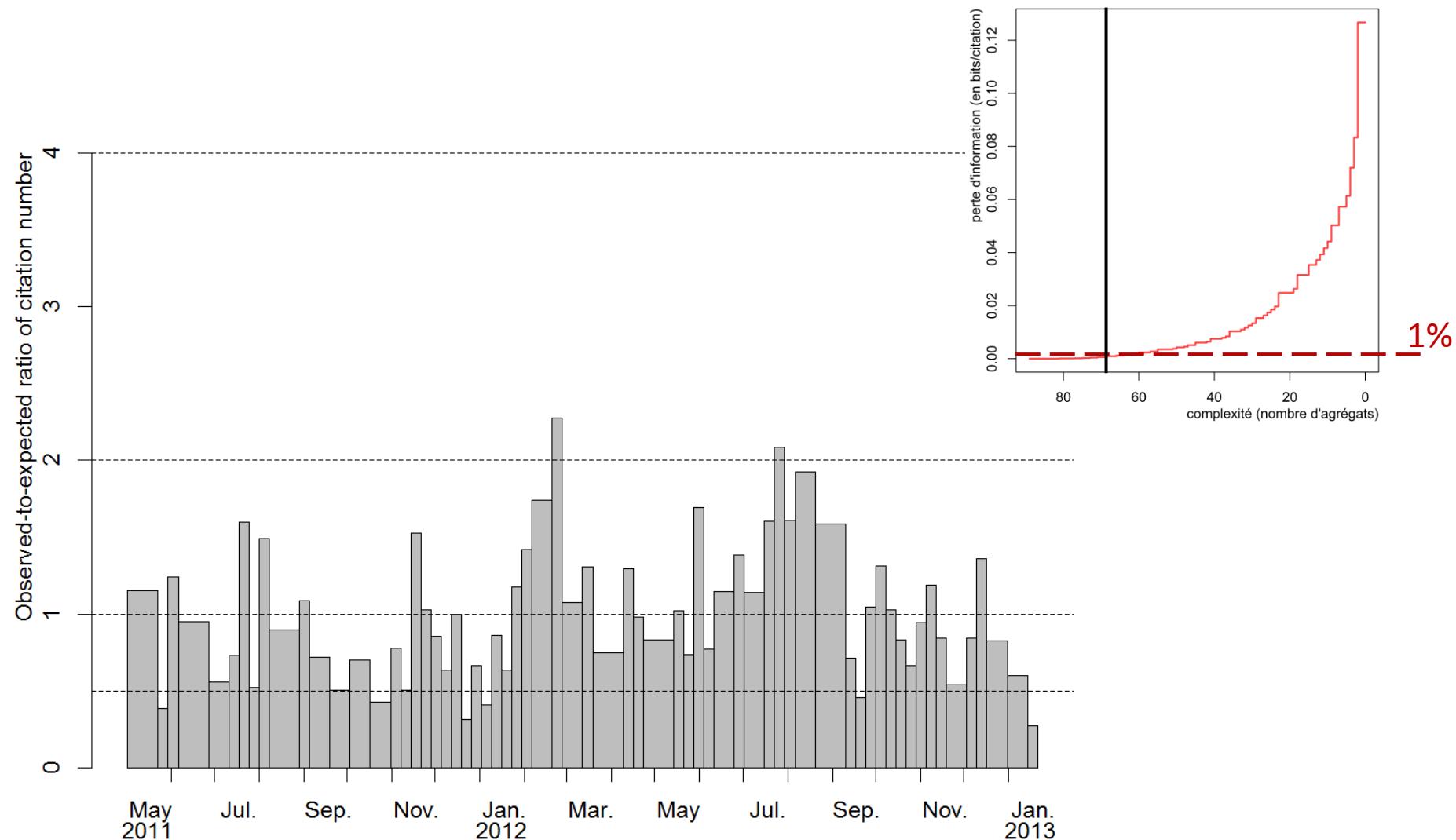
The Syrian civil war according to LE MONDE

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



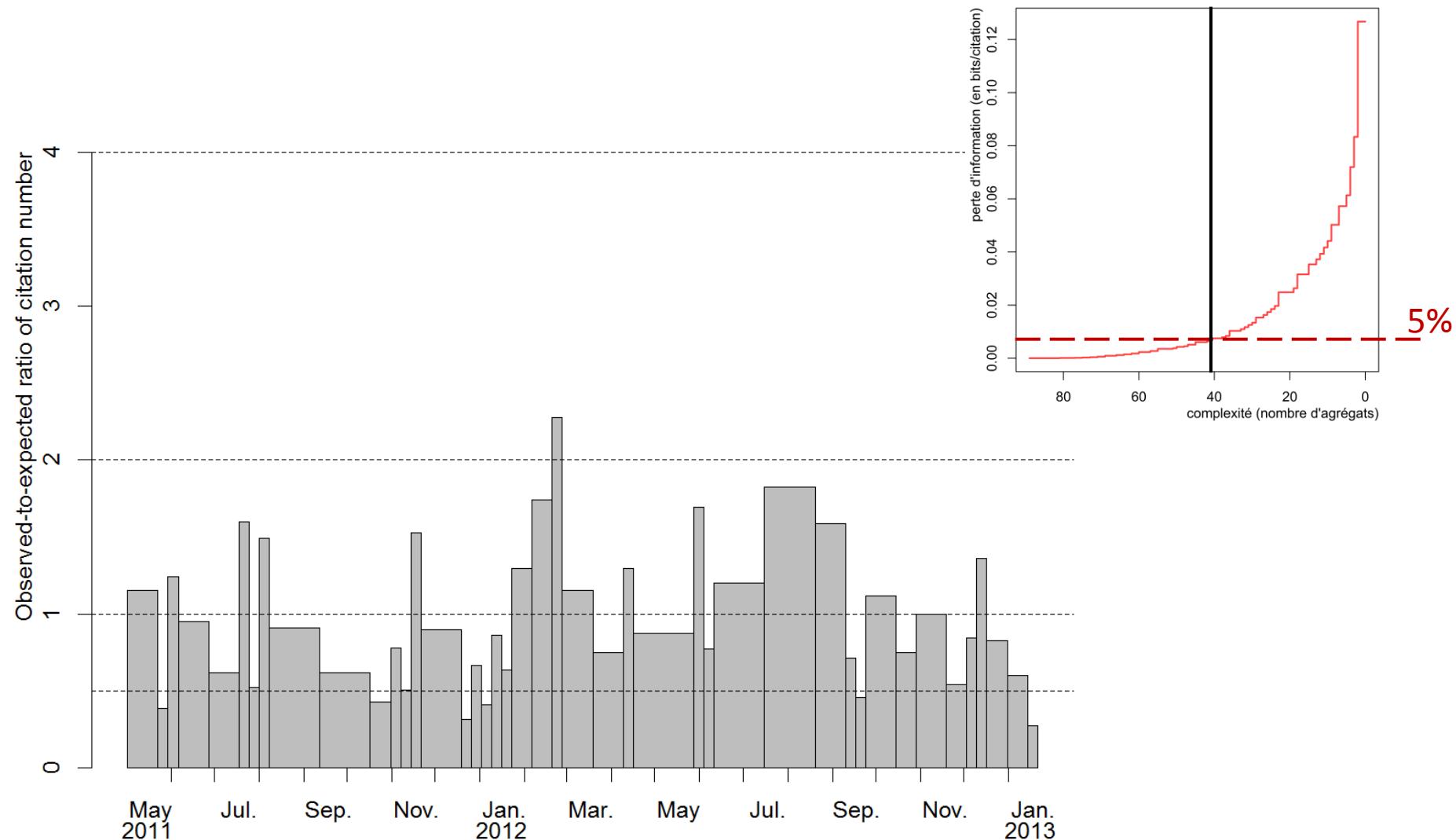
The Syrian civil war according to LE MONDE

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



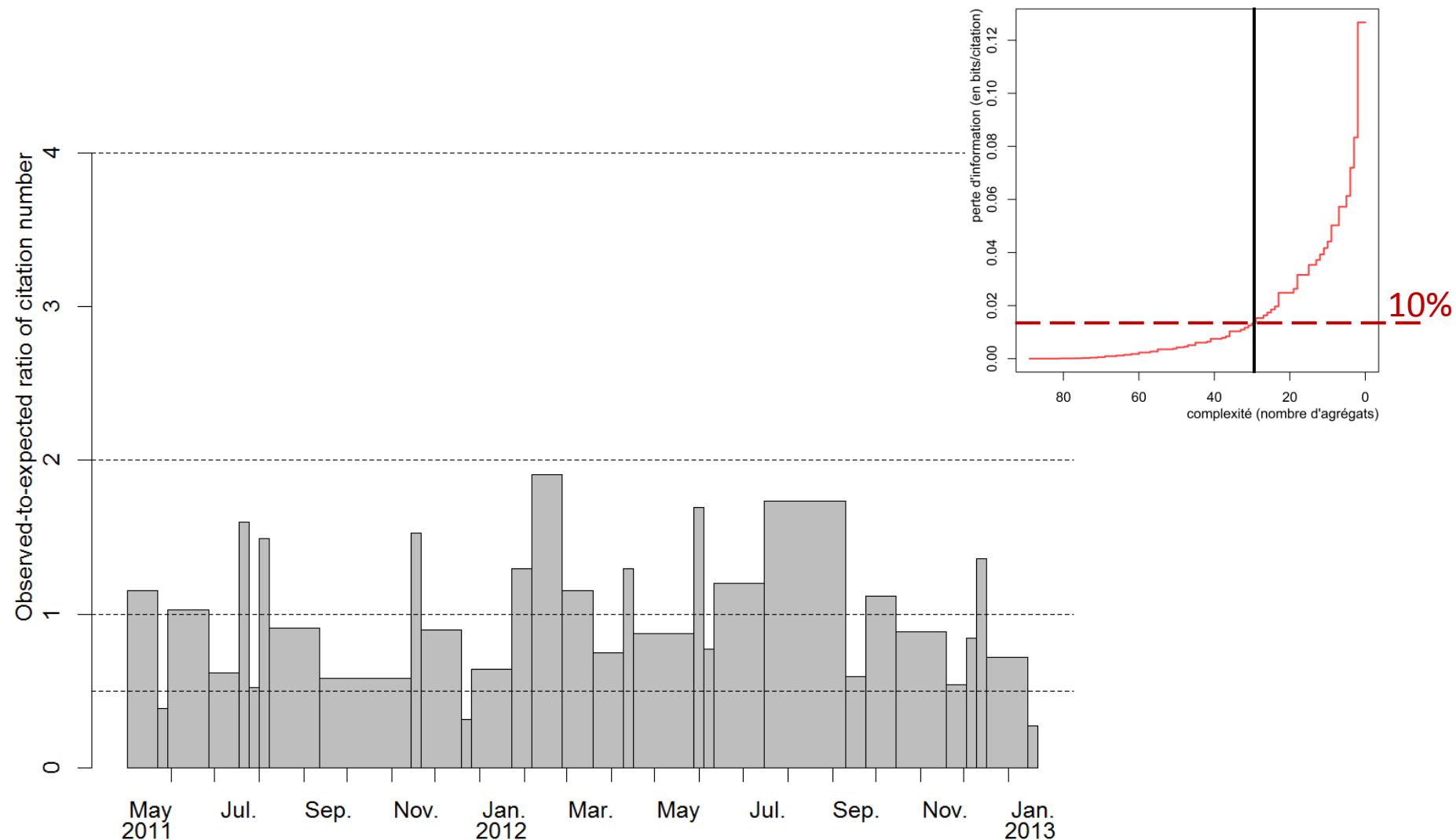
The Syrian civil war according to LE MONDE

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



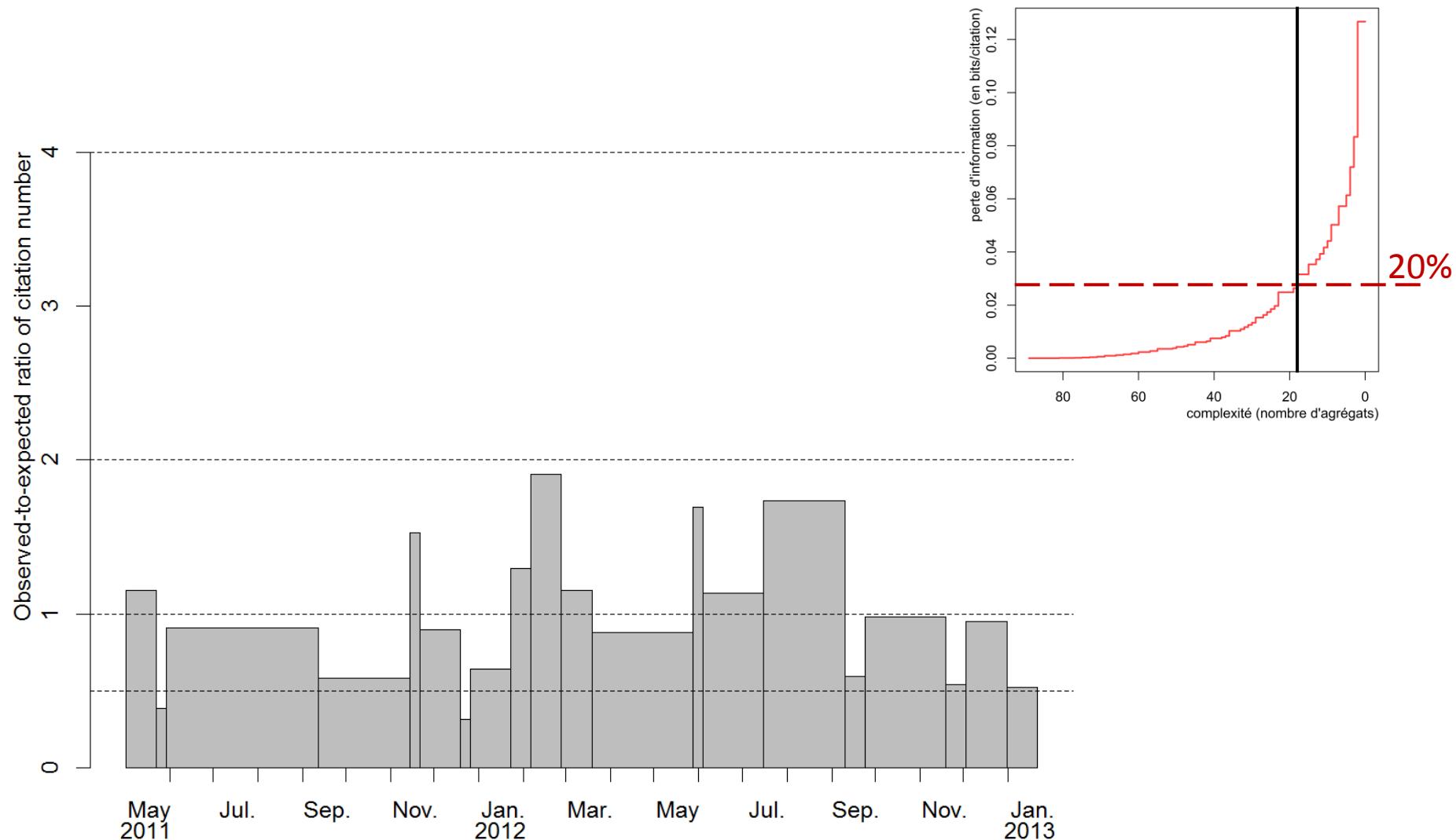
The Syrian civil war according to LE MONDE

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



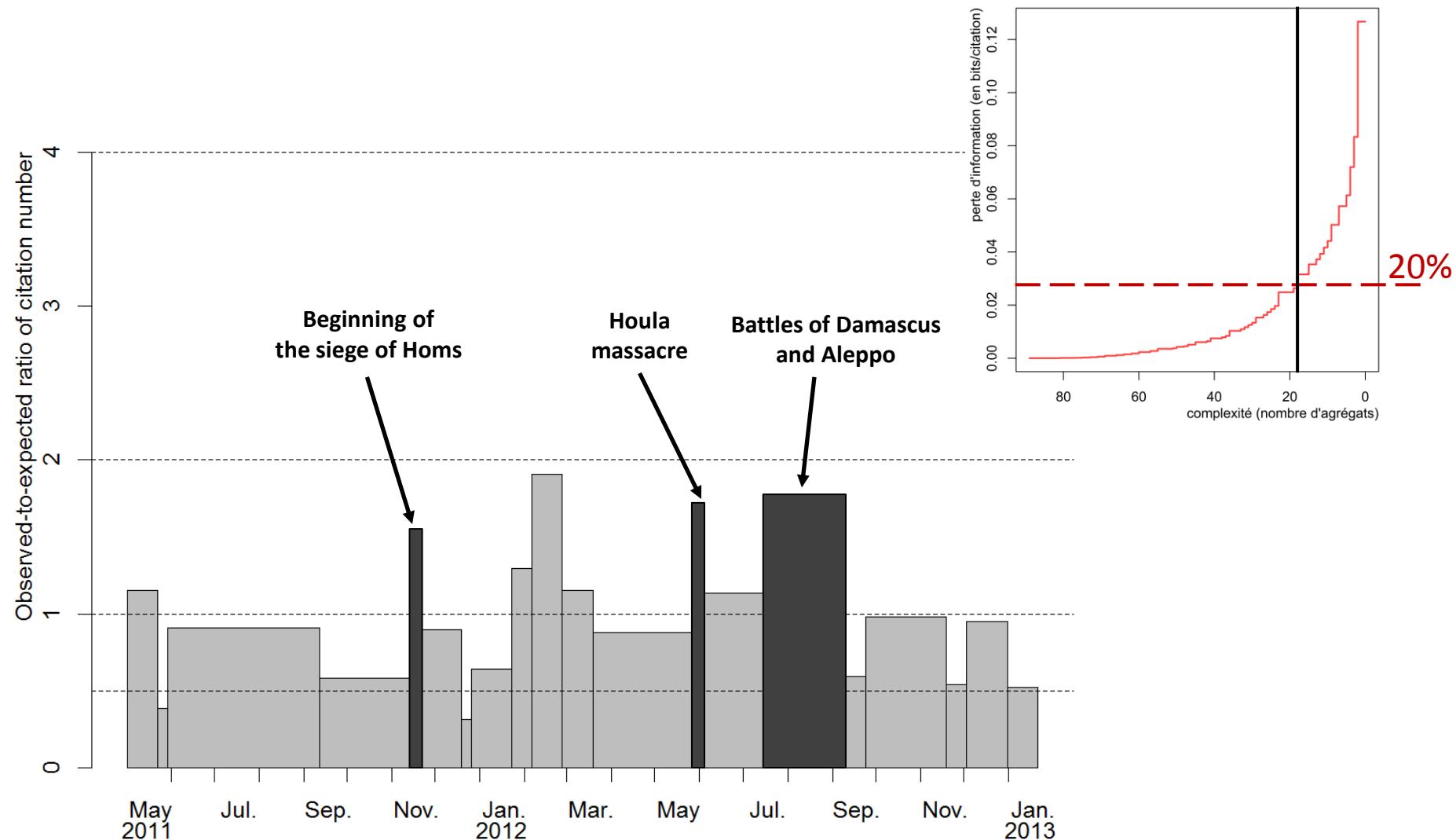
The Syrian civil war according to LE MONDE

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



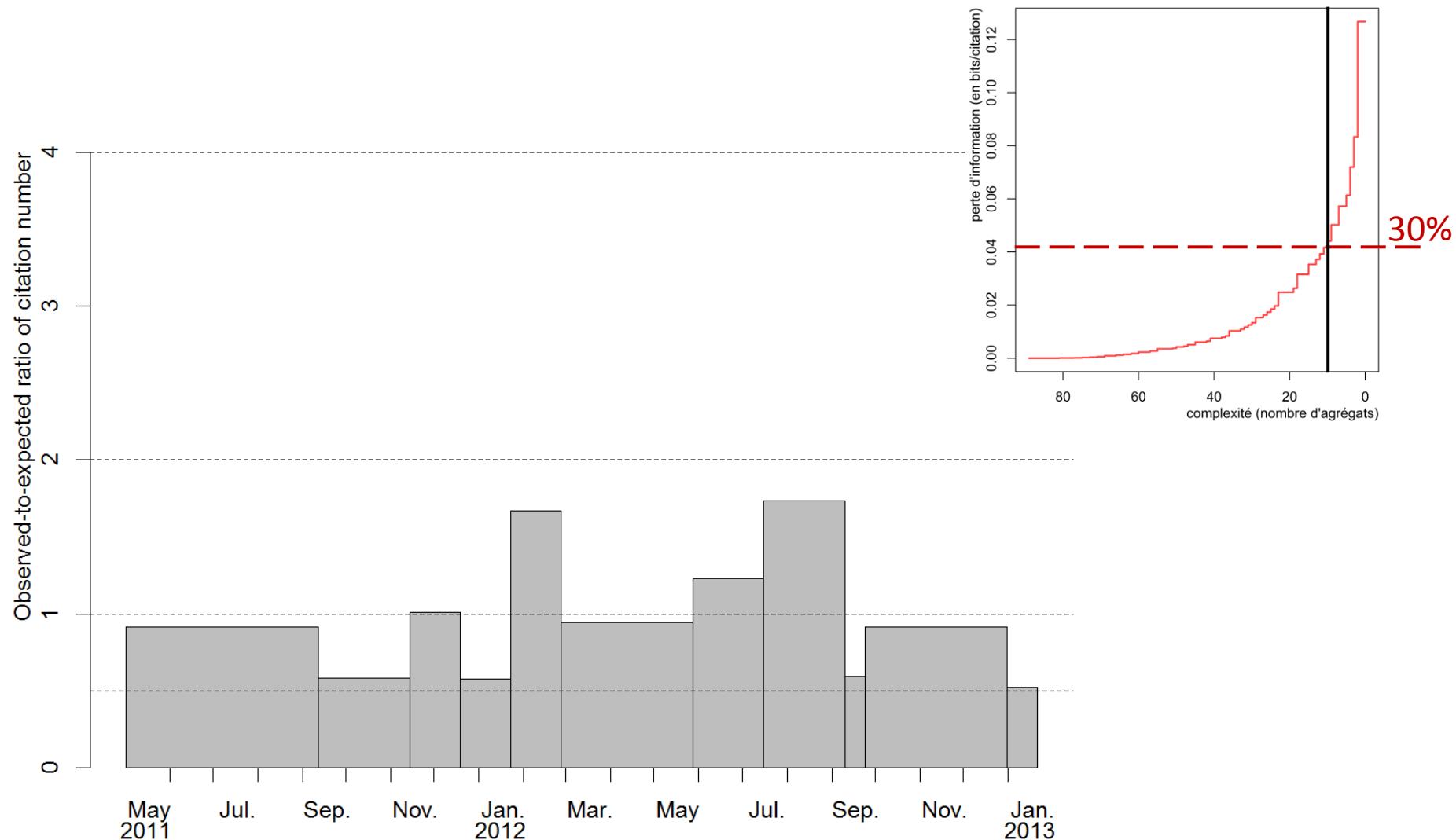
The Syrian civil war according to LE MONDE

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



The Syrian civil war according to LE MONDE

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

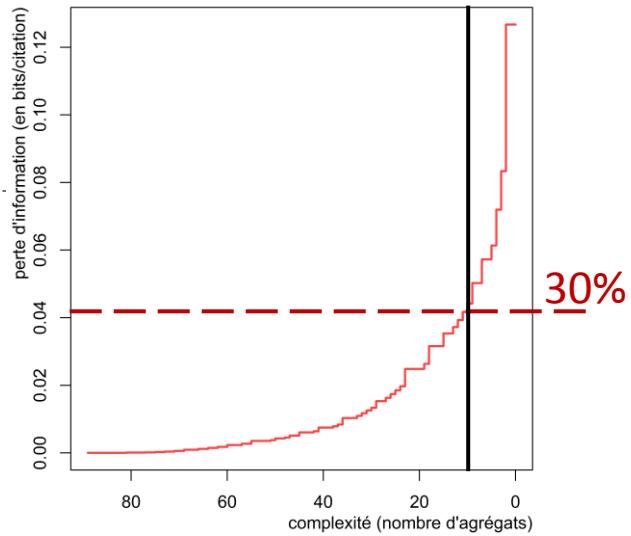
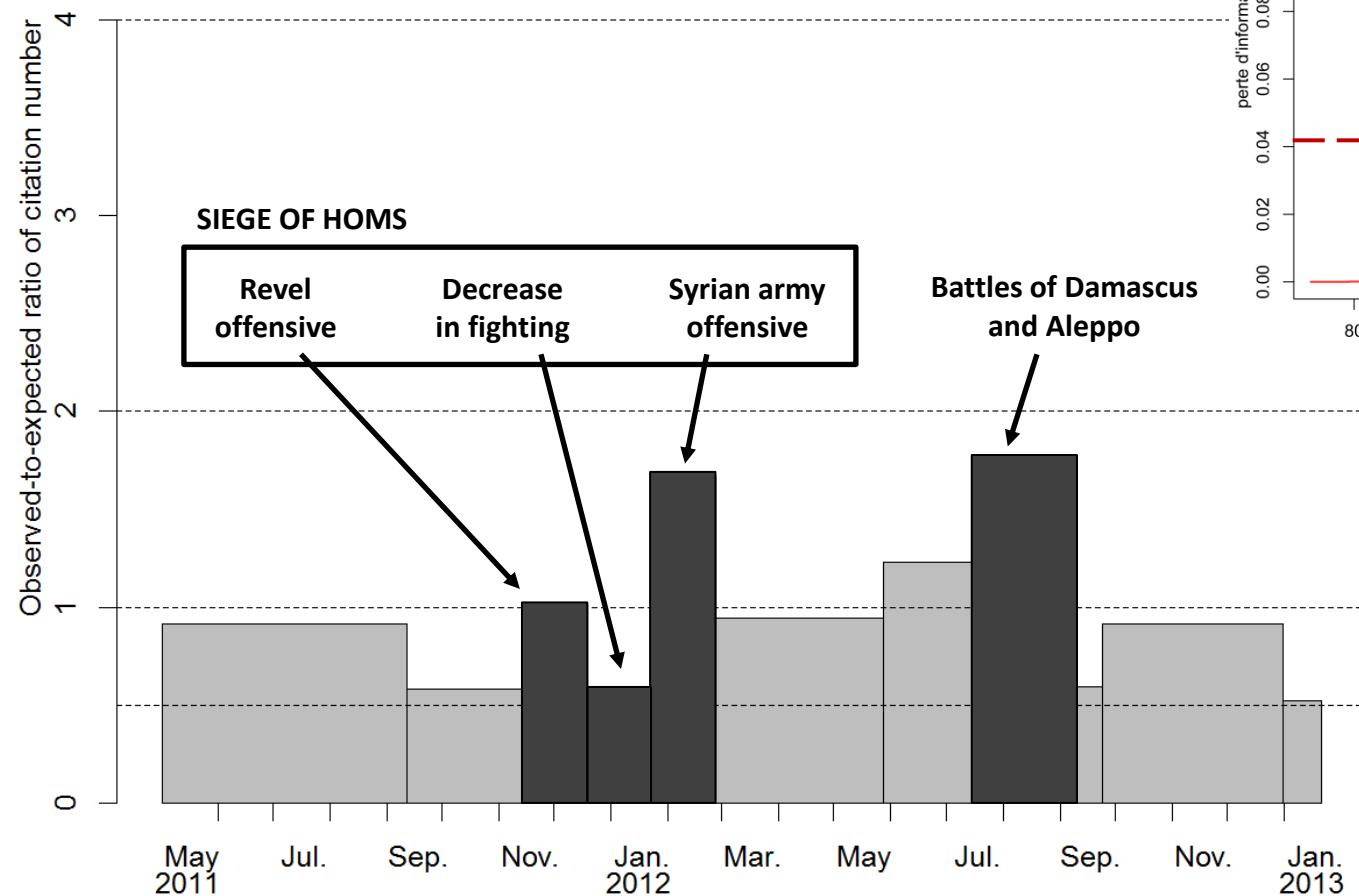


The Syrian civil war according to LE MONDE

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

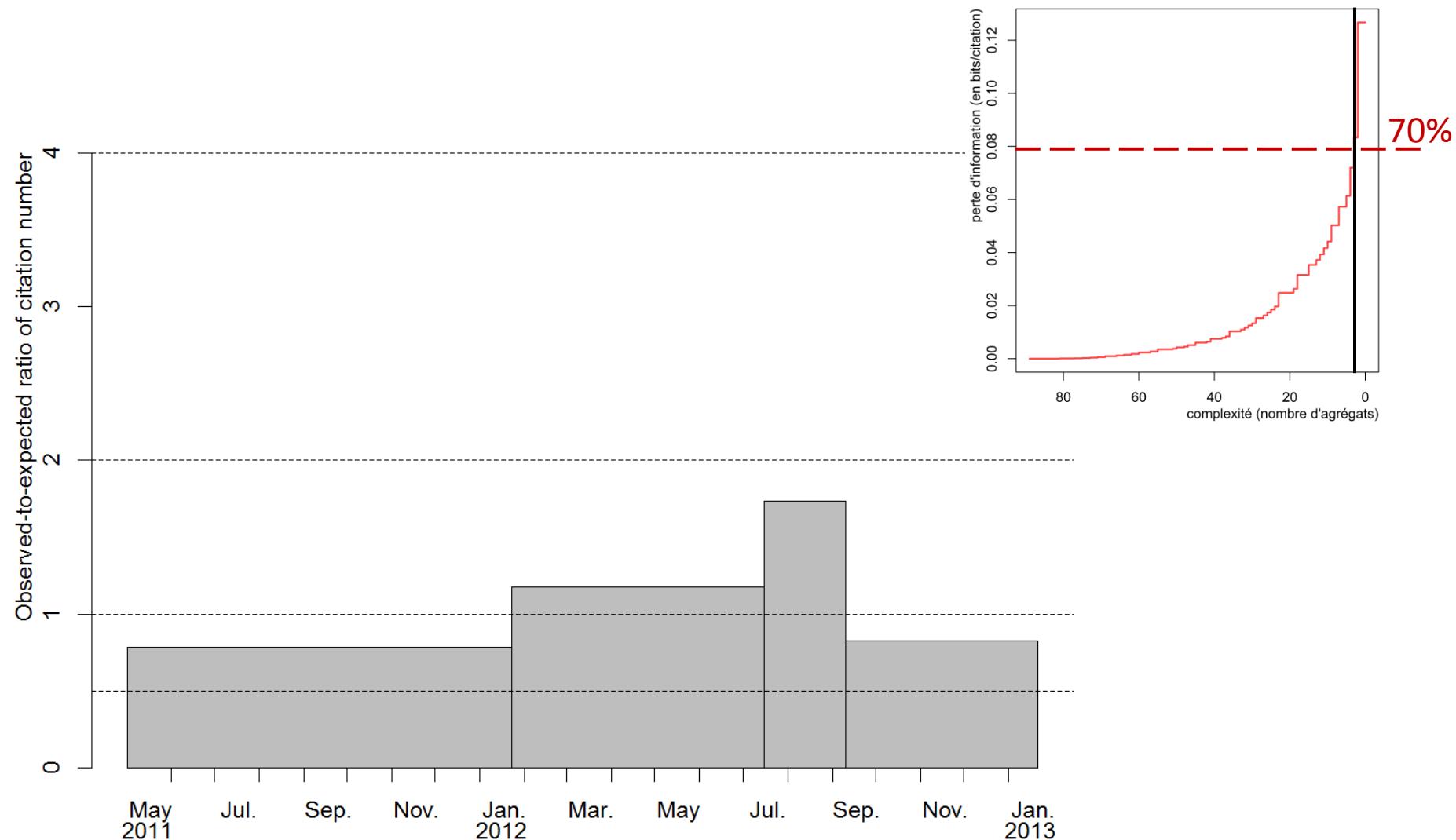
Source: Wikipedia

Timeline of the Syrian civil war
Siege of Homs



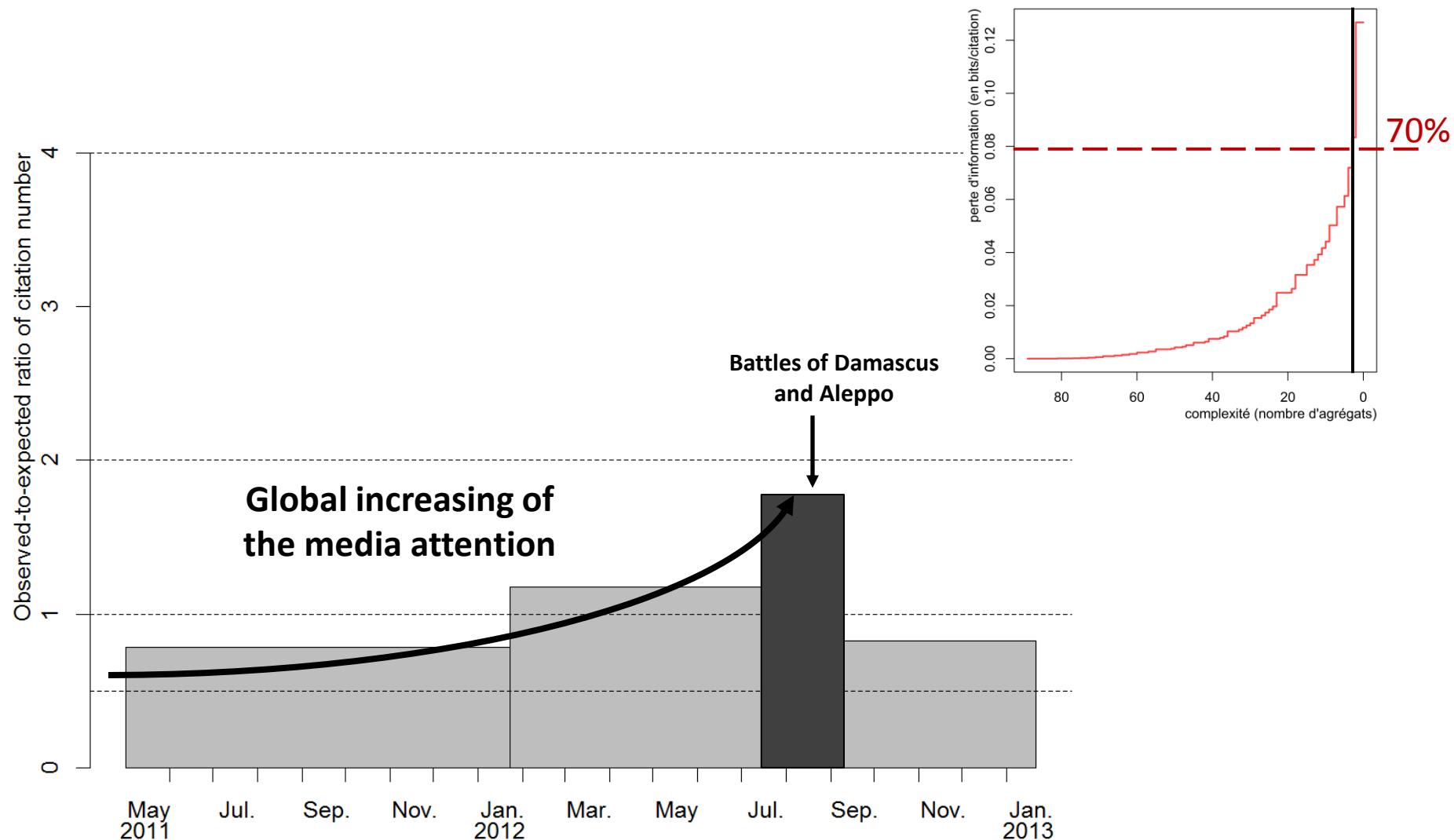
The Syrian civil war according to LE MONDE

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



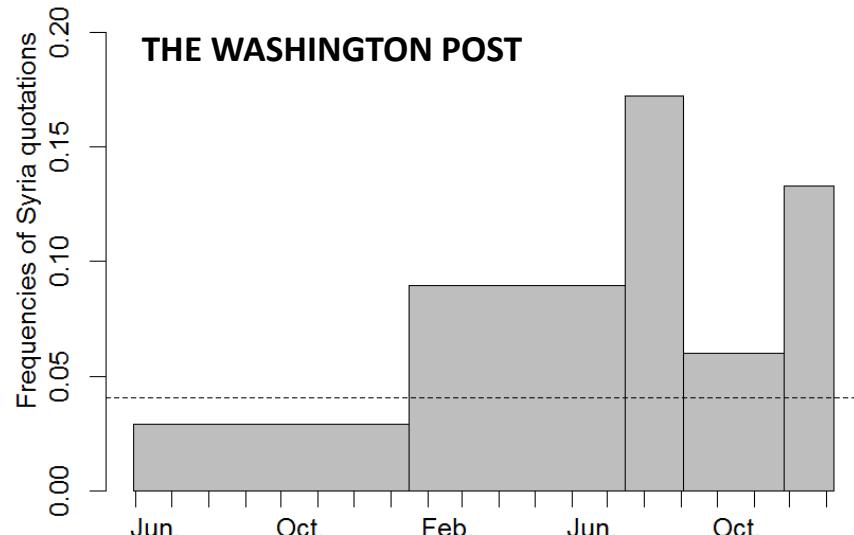
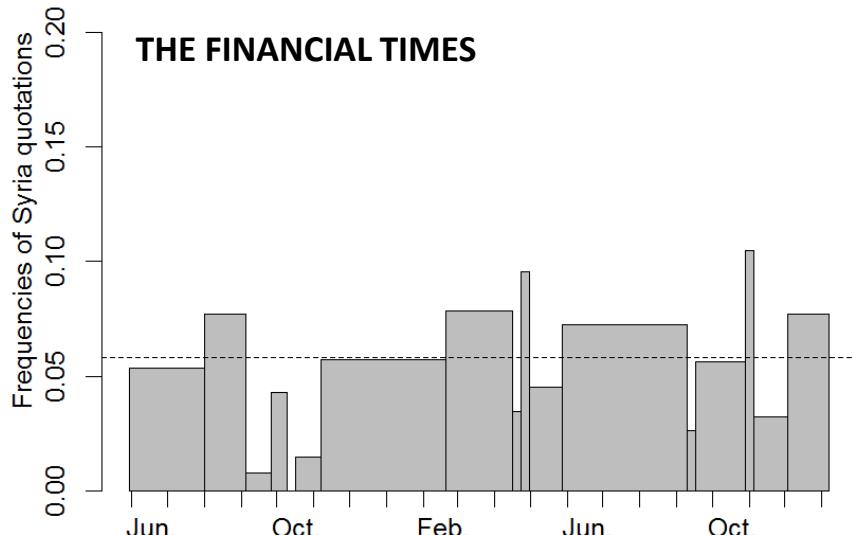
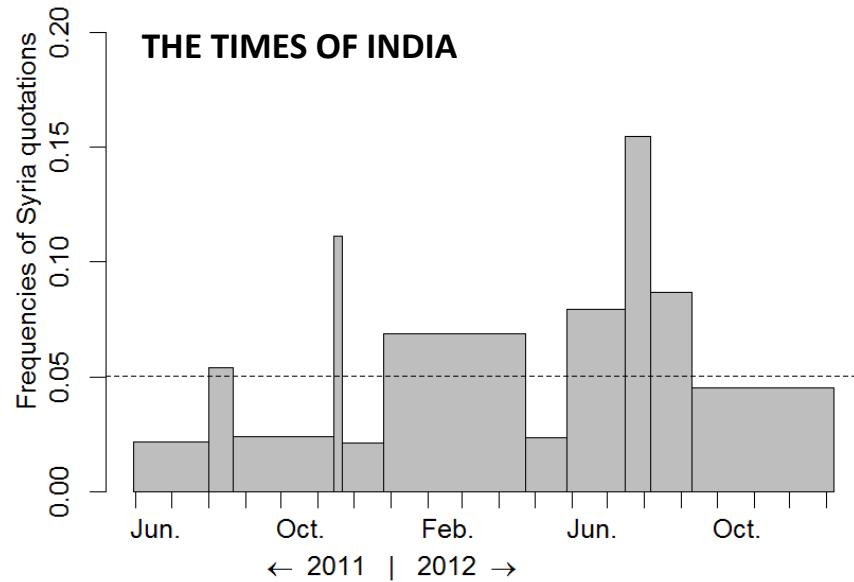
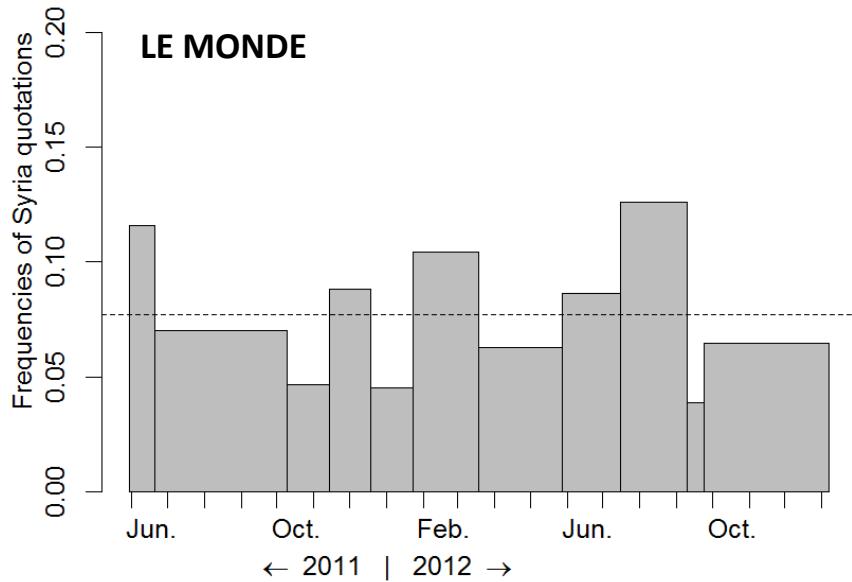
The Syrian civil war according to LE MONDE

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



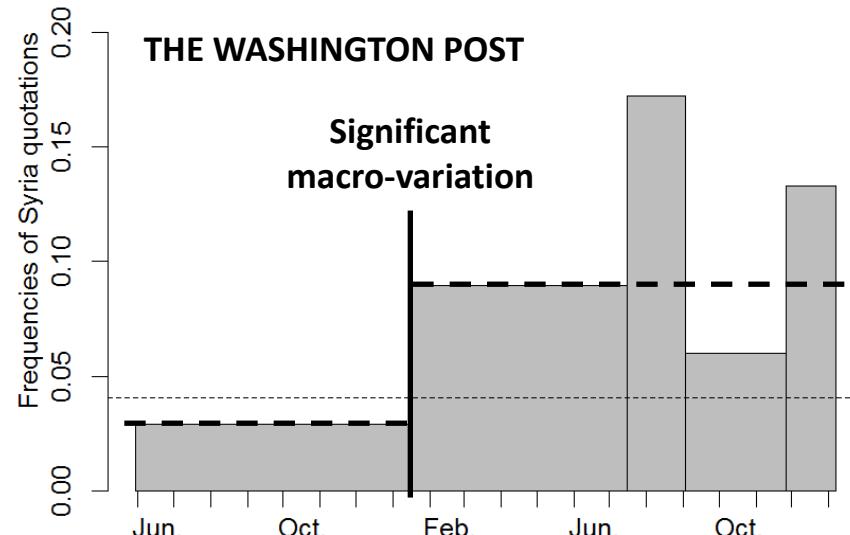
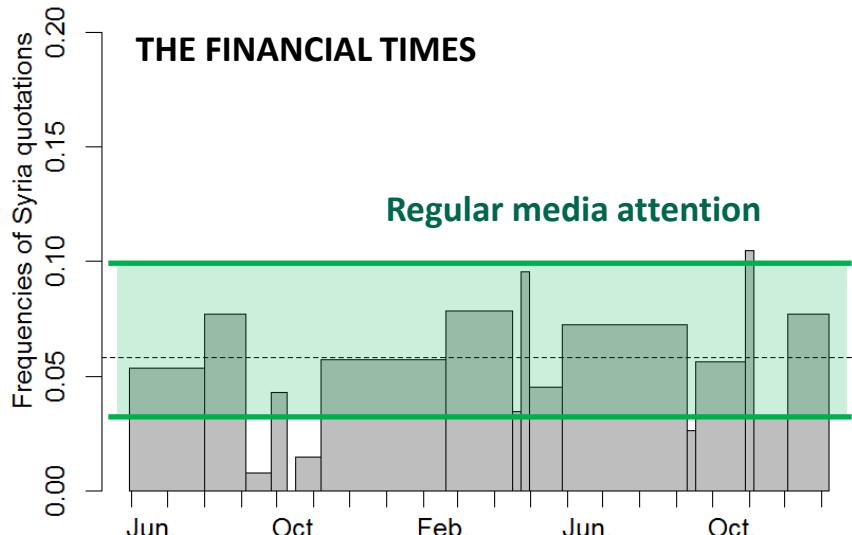
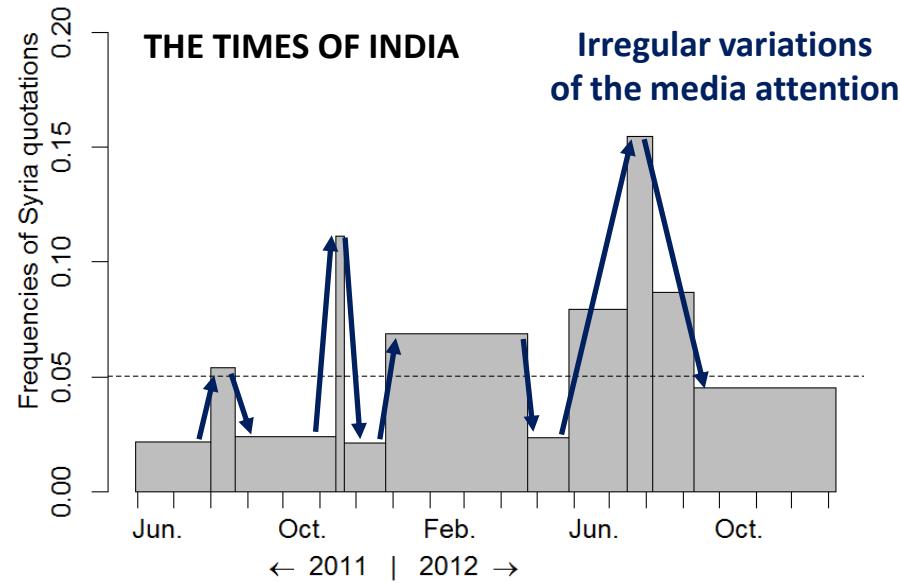
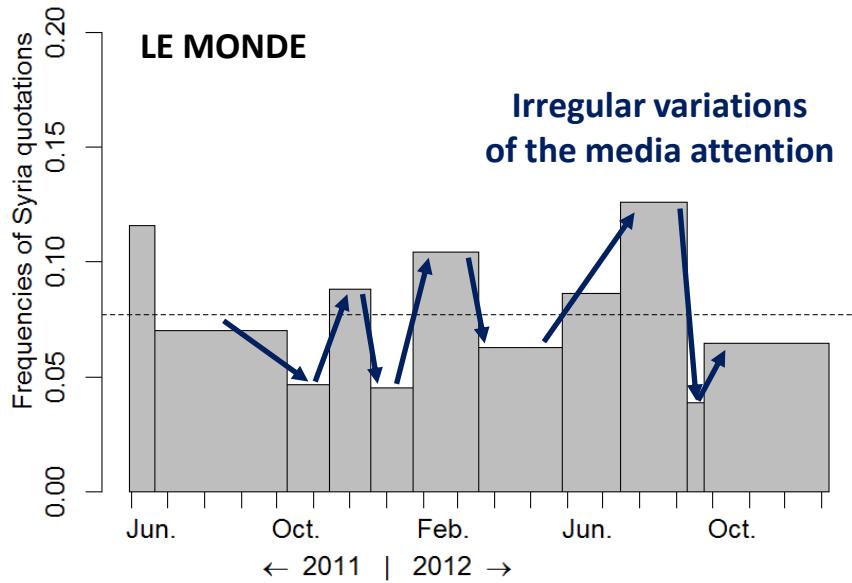
The Syrian civil war according to 4 newspapers

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



The Syrian civil war according to 4 newspapers

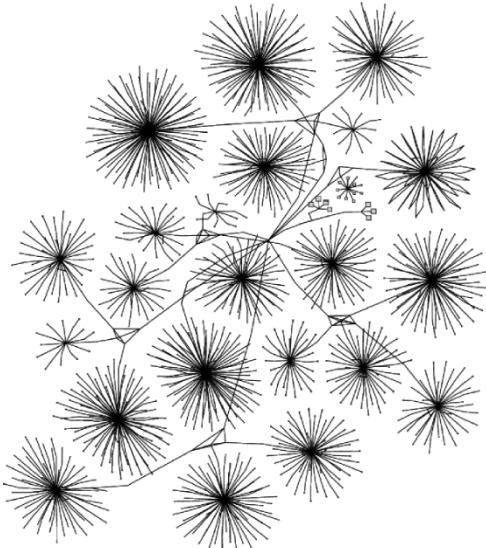
[Giraud, Grasland, Lamarche-Perrin *et al.*, ECTQG 2013]



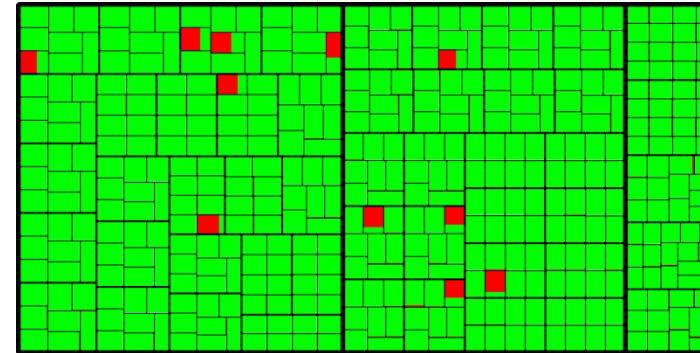
Aggregation of Execution Traces

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

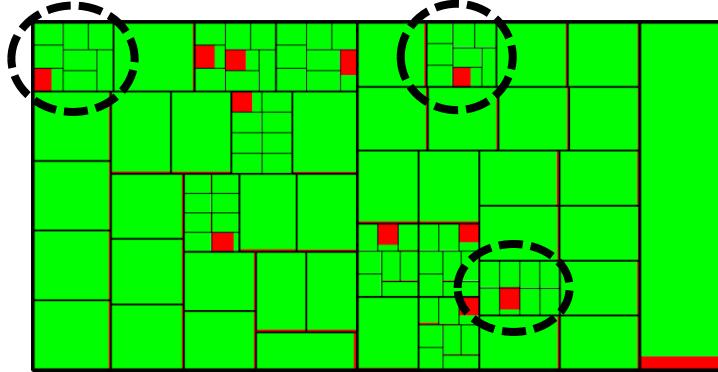
Hierarchical structure
of the grid computing
[Schnorr *et al.*, 2013]



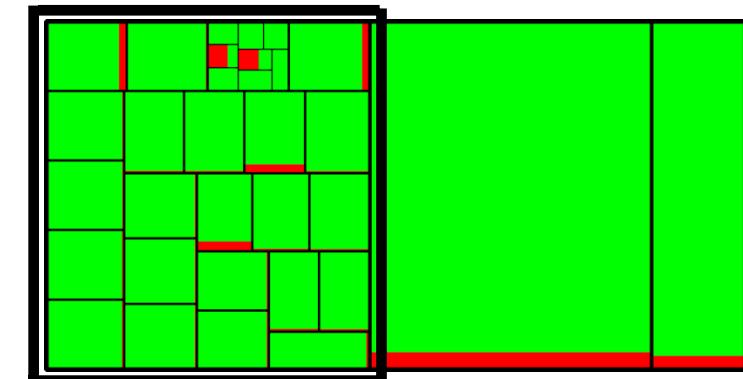
Microscopic treemap
representation



Detection of
multi-scale
anomalies

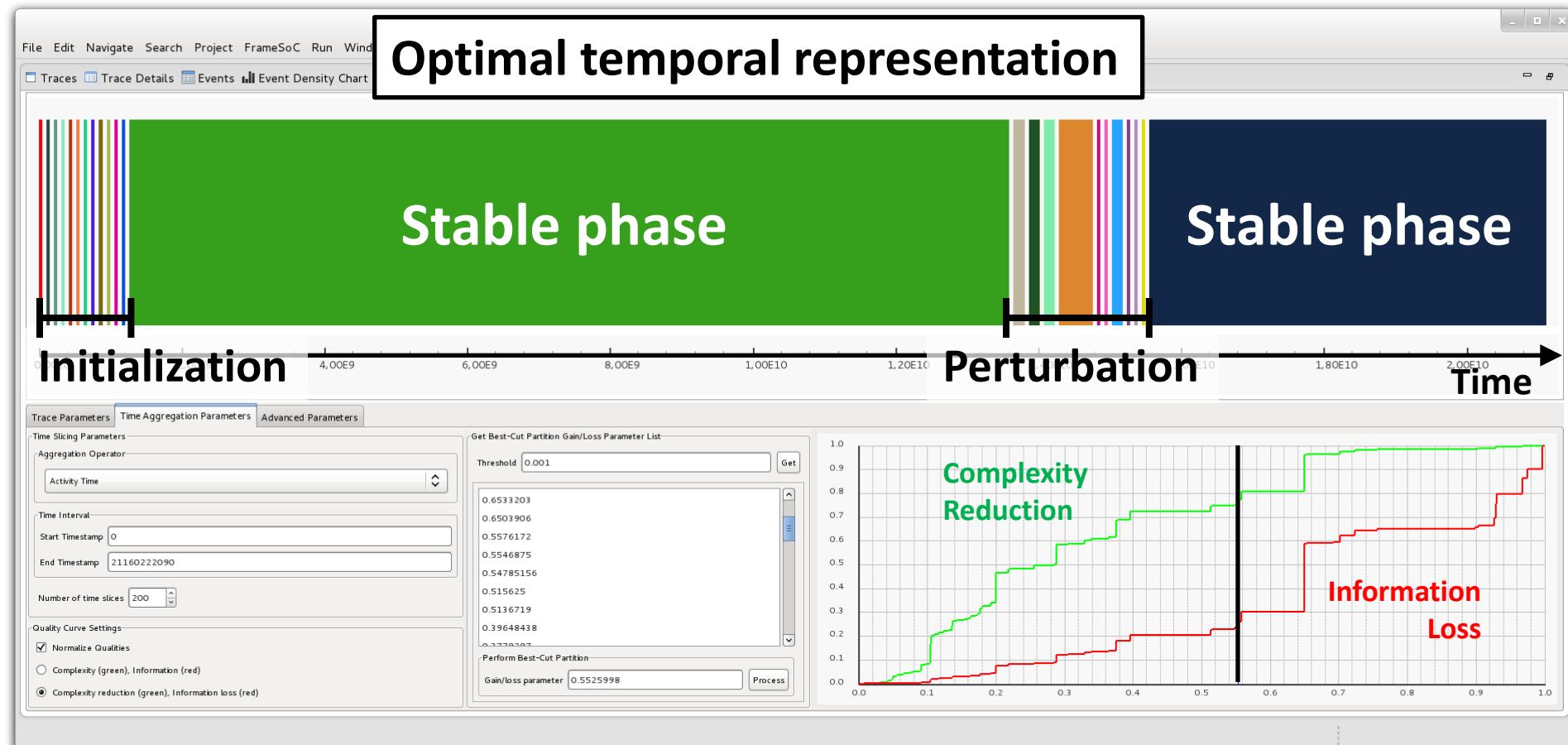


Aggregated treemap representations



Aggregation of Execution Traces

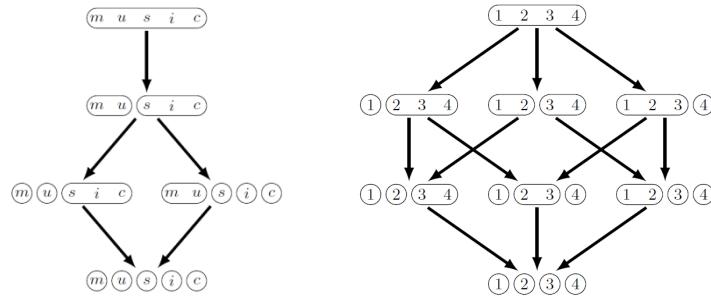
[Pagano, Dosimont *et al.*, 2013]



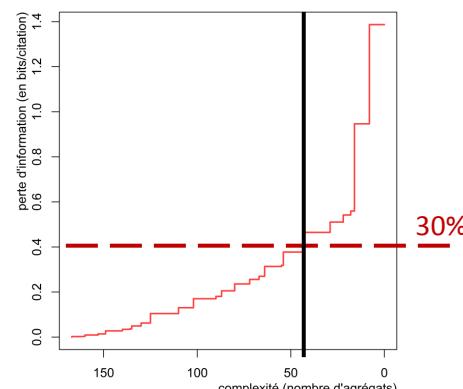
OUTCOMES AND PERSPECTIVES

Summary of the Contributions

P1 Algebraic structures that express the system's semantic properties



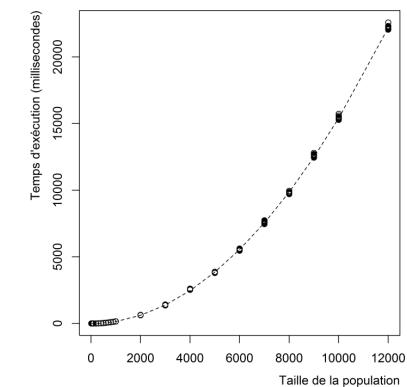
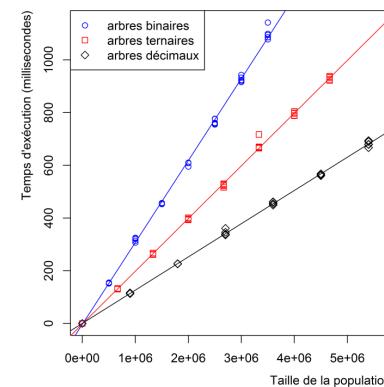
P2 Graphs of quality to choose the representations granularity



P2 A compromise of quality to generate multi-resolution representations

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

P3 A generic aggregation algorithm with polynomial complexity



Summary of the Hypotheses

Microscopic observation tools

- P0** Aggregation according to **partitions**

- P1** Hierarchical or ordered systems

- P2** Decomposable quality measures

Theoretical Perspectives

Aggregation of other structures

- Multidimensional aggregation and commutativity
- Aggregation of causal relations, of lexical networks
- Aggregation of the state space of a Markov process

Some measures to evaluate representations

- Measures from information theory and complexity science
- Measures of cognitive efficiency [Tufte, 1983]
- Other criteria to evaluate representations...

Application Perspectives

In social science

- Urban structures, micro/macro economics, social networks...
- **THE GEOMEDIA DATABASE IS AVAILABLE!**

In computer science and AI

- Exascale computing, embedded systems, sensor networks, multi-agent systems, swarm intelligence...

In other fields?

- Biology, neuroscience, physics...

Epistemological Perspectives

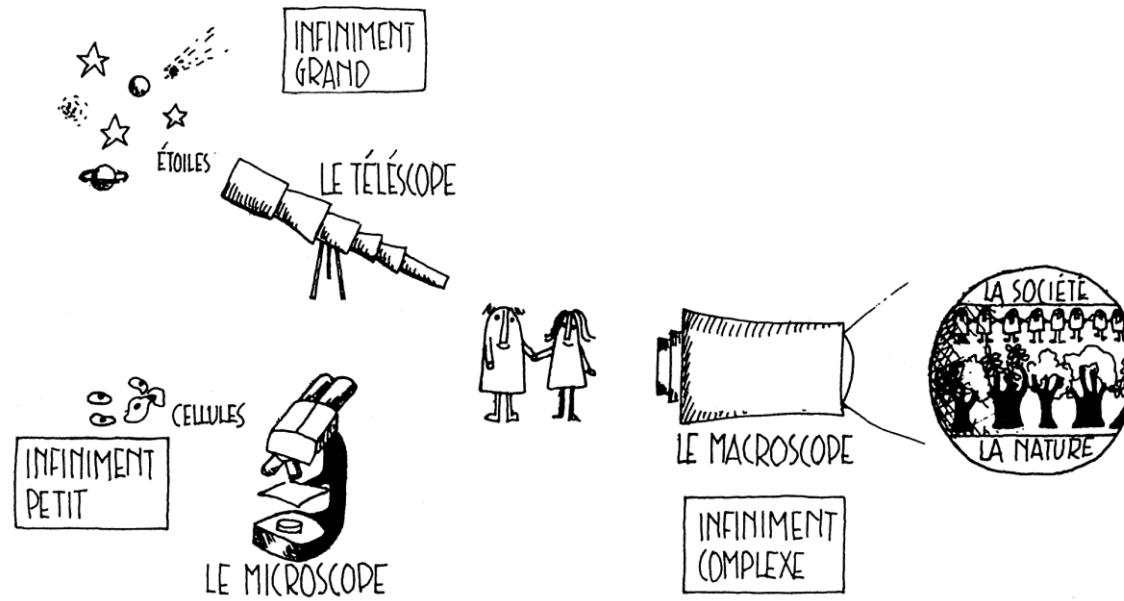
What is a macroscopic phenomenon?

- Is it a real object or just an abstraction?
- Are upward and downward causations possible?
- What is the concept of epistemological emergence?

What is the place of the observer?

- Internal or external observation?
- What is the probe effect?
- Is a self-observation possible?

THANKS FOR YOUR ATTENTION



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

Joël de Rosnay, *Le macroscope*, 1975

LIST OF PUBLICATIONS

Peer-reviewed Journals (being accepted)

Lamarche-Perrin, Demazeau et Vincent. **Building the Best Macroscopic Representations of Complex Multi-Agent Systems.** *Transaction on Computational Collective Intelligence (TCCI)*, 2014.

Lamarche-Perrin, Schnorr, Vincent et Demazeau. **Agrégation de traces pour la visualisation de grands systèmes distribués.** *Technique et Science Informatiques (TSI)*, 2014.

International Peer-reviewed Conferences with Proceedings

Lamarche-Perrin, Demazeau et Vincent. **The Best-partitions Problem: How to Build Meaningful Aggregations.** *Intelligent Agent Technology (IAT)*, Atlanta, 2013.

Giraud, Grasland, Lamarche-Perrin, Demazeau et Vincent. **Identification of International Media Events by Spatial and Temporal Aggregation of RSS Flows of Newspapers.** *European Colloquium in Theoretical and Quantitative Geography (ECTQG)*, Dourdan, 2013.

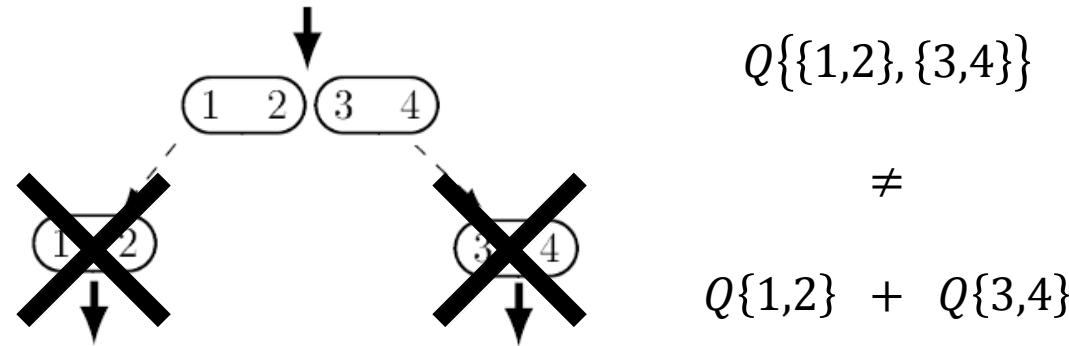
Lamarche-Perrin, Demazeau and Vincent. **How to Build the Best Macroscopic Description of your Multi-agent System? Practical Applications of Agents and Multi-Agent Systems (PAAMS)**, Salamanca, 2013.

French Peer-reviewed Conferences with Proceedings

Lamarche-Perrin, Demazeau et Vincent. **Organisation, agrégation et visualisation d'informations médiatiques.** *Colloque annuel du Collège des Sciences du Territoire*, Paris, 2011.

Lamarche-Perrin, Demazeau et Vincent. **Observation macroscopique et émergence dans les SMA de très grande taille.** *Journées Francophones des Systèmes Multi-Agents (JFSMA)*, Valenciennes, 2011.

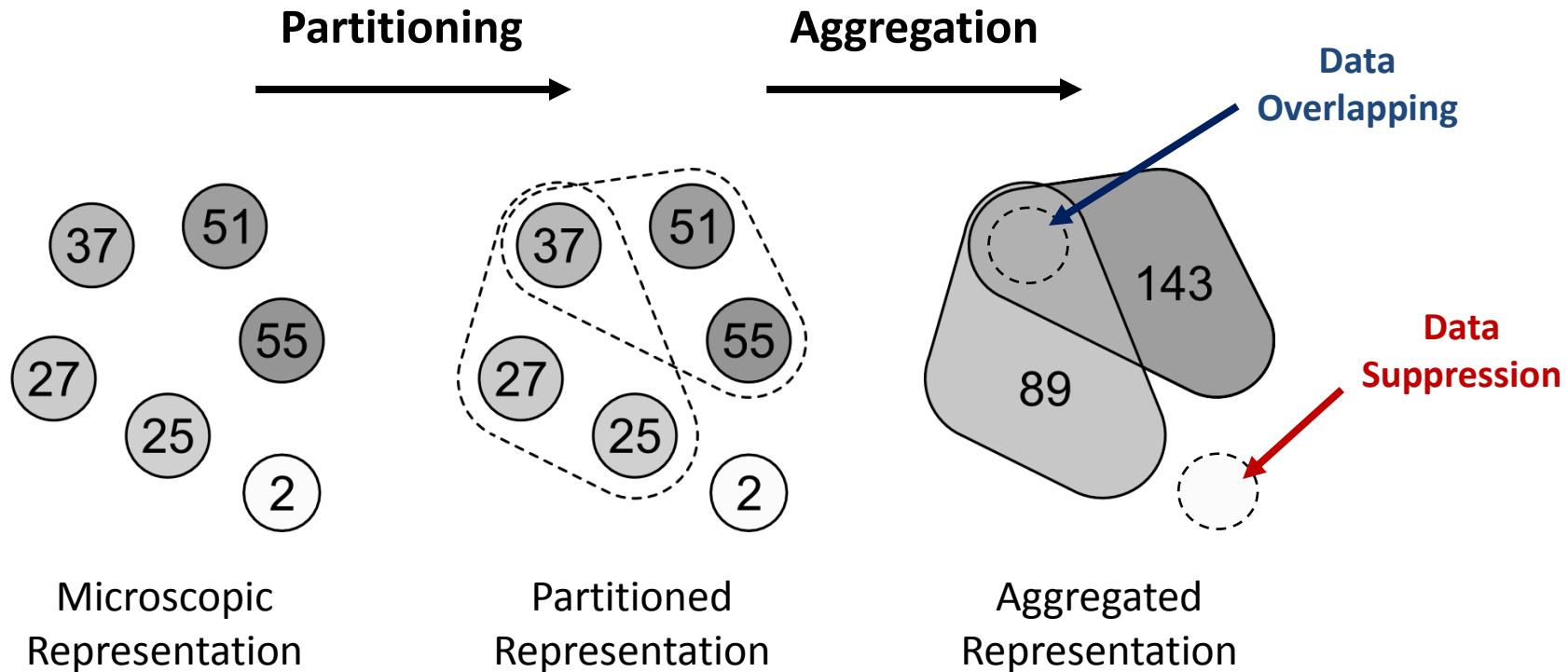
Non-decomposable Measures



- The recursive approach is no longer possible
- The quality of a part cannot be defined

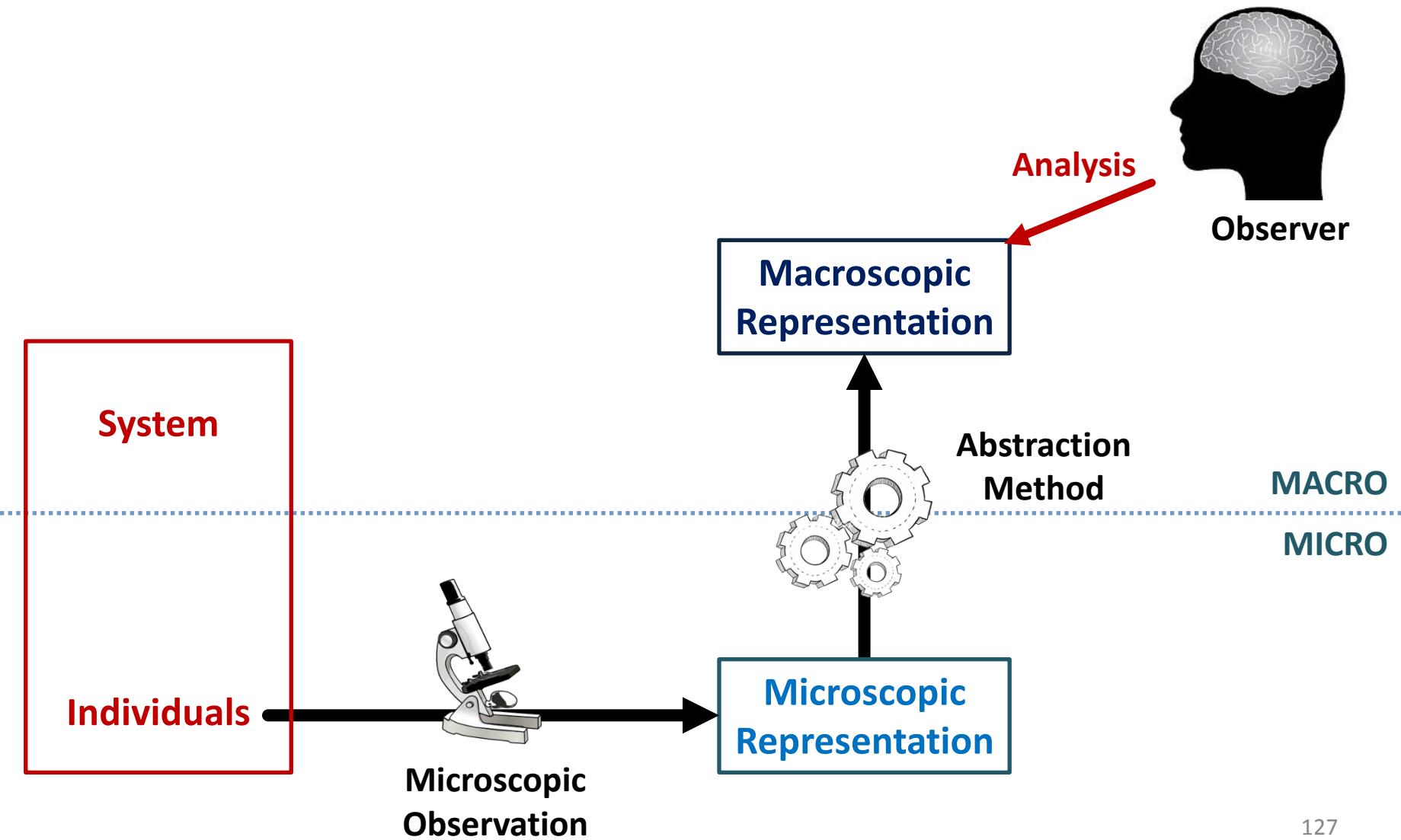
→ Do non-decomposable quality measures have a meaning to evaluate the aggregation process?

Non-disjoint and Non-covering Parts

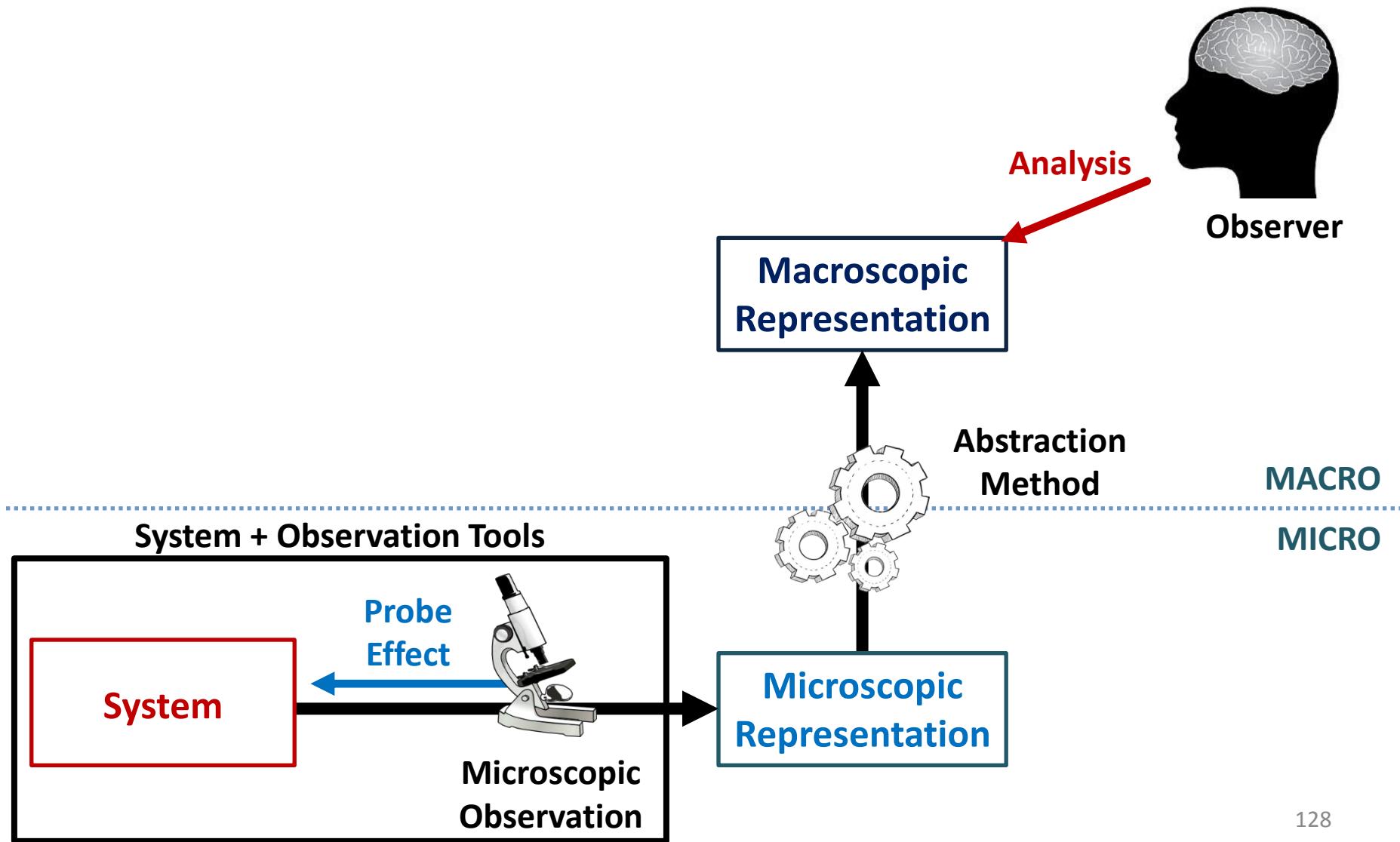


→ Effect on the algorithmic complexity?

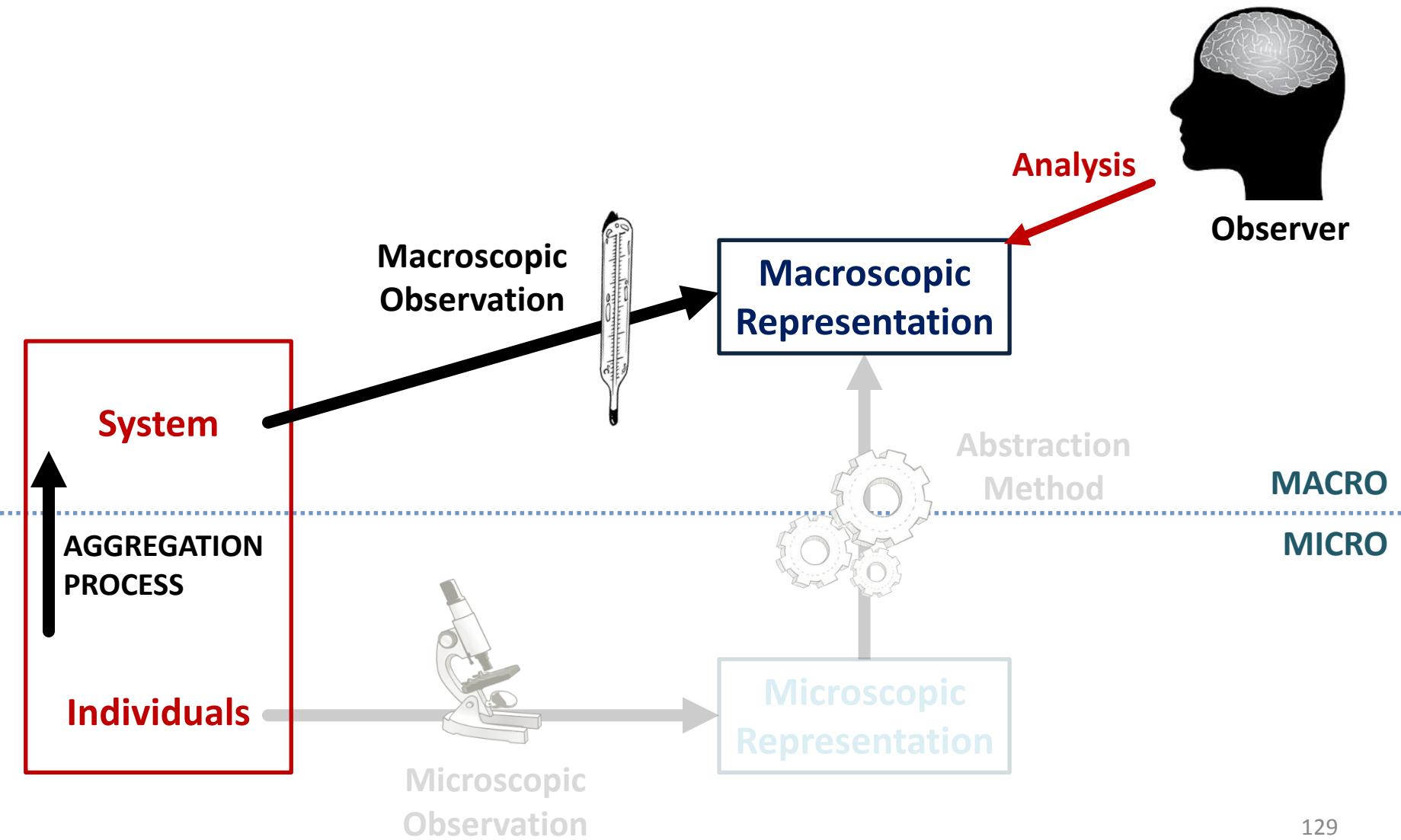
Microscopic Observation



The Probe Effect



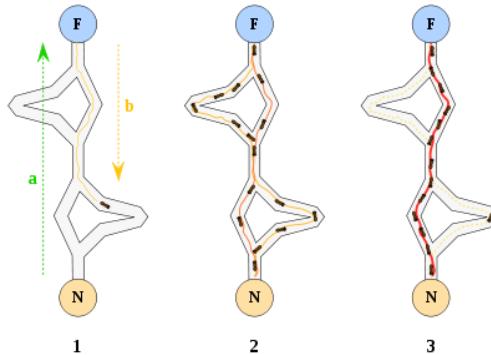
Macroscopic Observation



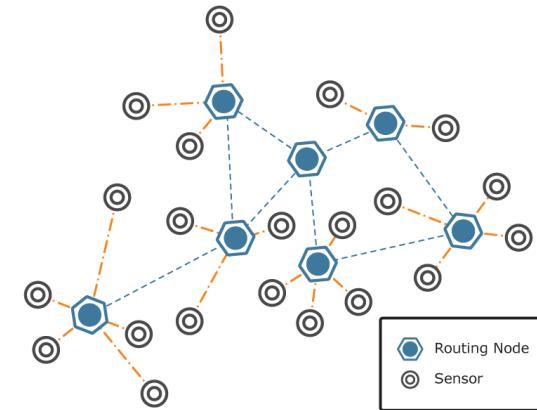
Application Perspectives

Multi-agent Systems

[Lamarche-Perrin *et al.*, JFSMA 2011]

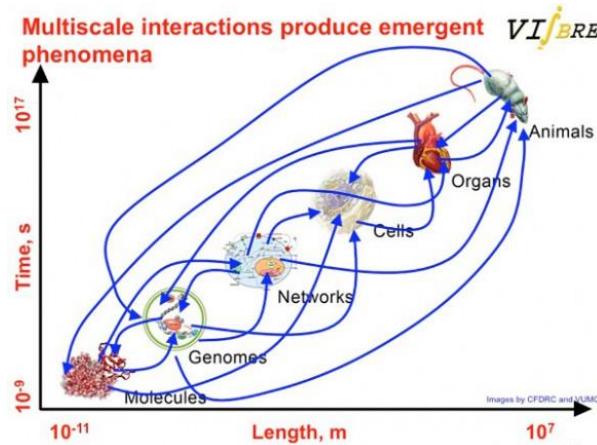


Sensor Networks

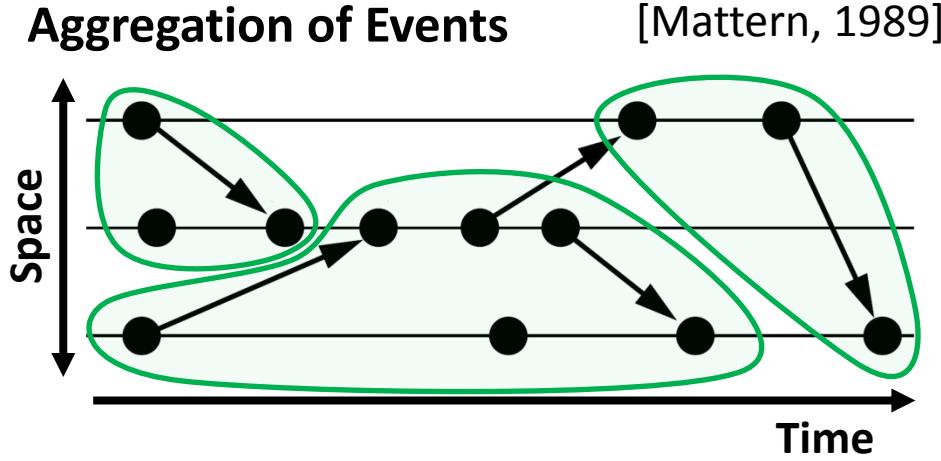


Multi-scale simulation

[Gil-Quijano *et al.*, 2012]



Other Topological Structures

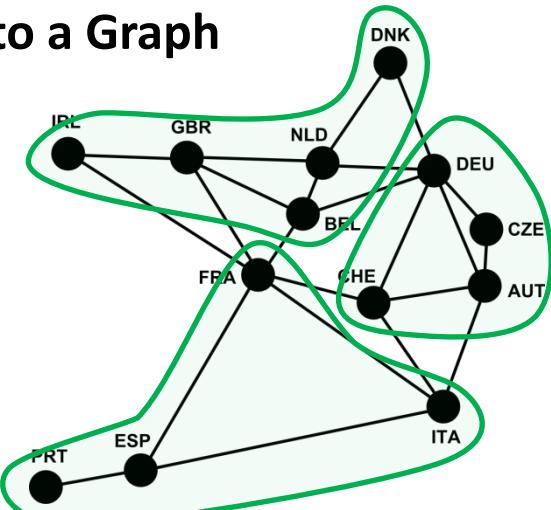


Aggregation of an Interaction Matrix

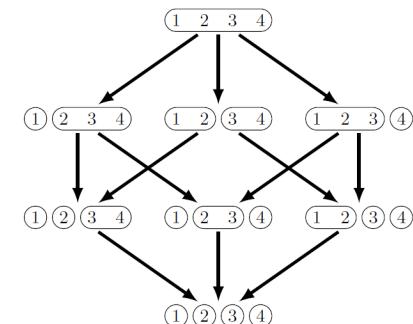
	ESP	FRA	GBR	BEL	CHE
ESP	X	12	11	10	4
FRA	14	X	12	12	5
GBR	20	11	X	6	9
BEL	15	9	6	X	5
CHE	10	16	17	9	X

[Lamarche-Perrin *et al.*, CIST 2011]

Aggregation according to a Graph



Corresponding Algebras



Algorithmic Complexity

Find the complexity classes that are associated to other dimensions, other semantics, other topologies, *etc.*

Set	Admissible Parts	Admissible Partitions	Temporal Complexity	Spatial Complexity
Non-constrained	$\Theta(2^n)$	$\Theta(e^{n \log n})$	$\Theta(3^n)$	$\Theta(2^n)$
Ordered	$\Theta(n^2)$	$\Theta(2^n)$	$\Theta(n^2)$	$\Theta(n^2)$
Hierarchical	$O(n)$	$O(c^n)$	$O(n)$	$O(n)$
Other topologies	?	?	?	?

Multidimensional Aggregation

$\{a, b, c, d, e\}$				
$\{a, b\}$	$\{c, d, e\}$			
$\{a\}$	$\{b\}$	$\{c\}$	$\{d\}$	$\{e\}$

$(a, 1)$	$(b, 1)$	$(c, 1)$	$(d, 1)$	$(e, 1)$
$(a, 2)$	$(b, 2)$	$(c, 2)$	$(d, 2)$	$(e, 2)$
$(a, 3)$	$(b, 3)$	$(c, 3)$	$(d, 3)$	$(e, 3)$
$(a, 4)$	$(b, 4)$	$(c, 4)$	$(d, 4)$	$(e, 4)$

$\{\underline{1}\}$
$\{\underline{2}\}$
$\{\underline{3}\}$
$\{\underline{4}\}$

$\{a, b, c, d, e\}$		
$\{a, b\}$		
$\{a\}$	$\{b\}$	$\{c, d, e\}$

$(a, 1)$	$(b, 1)$	$(\{c, d, e\}, \{1\})$
$(\{a\}, \{2, 3, 4\})$	$(\{b\}, \{2, 3, 4\})$	$(\{c, d, e\}, \{2, 3, 4\})$

$\{\underline{1}\}$
$\{2, 3, 4\}$

Microscopic Representation

Aggregated Representation

Aggregation of Causal Structures

