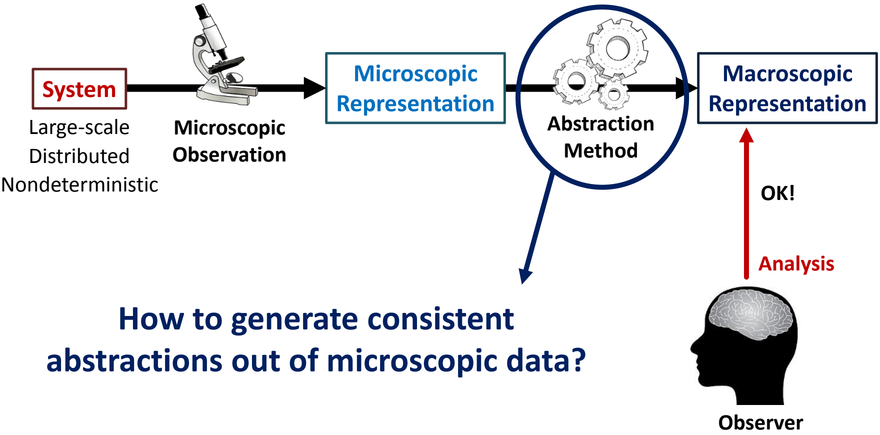
Multiscale Analysis of Social Systems

Robin Lamarche-Perrin

In collaboration with: Sven Banisch, Yves Demazeau, Damien Dosimont, Claude Graland, Eckehard Olbrich, Hugues Pecout, and Jean-Marc Vincent

Macroscopic Analysis of Complex Systems



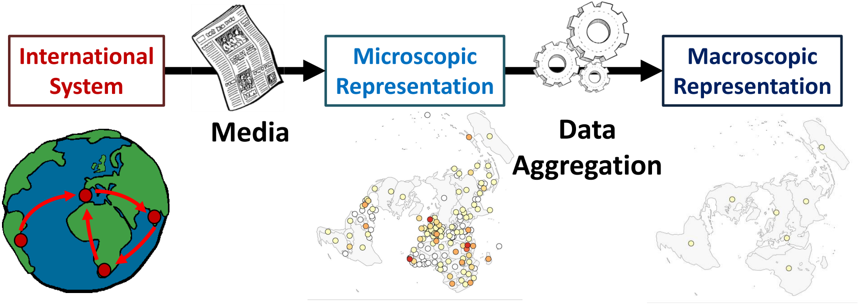
How to generate consistent abstractions out of microscopic data?

Macroscopic Analysis of Media Flows

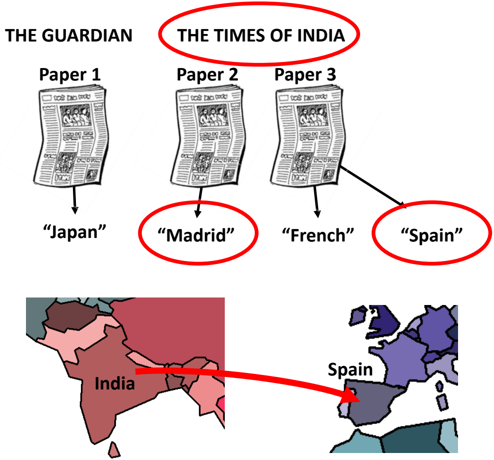
Hypothesis: media constitute an adequate instrument to observe the national level

[Grasland *et al.*, 2011]

Geographer



Raw Data



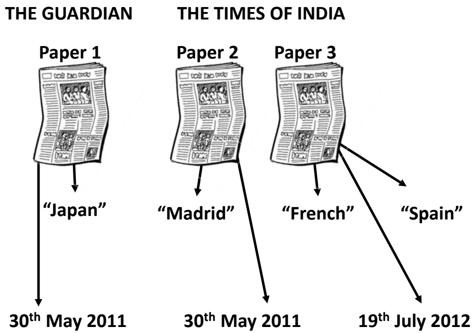
The GEOMEDIA Database
(ANR CORPUS GUI-AAP-04)

150 newspapers

1,944,000 papers

GEOGRAPHIC INFORMATION
193 countries (UN members)

Raw Data



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)

Raw Data

Space

Time

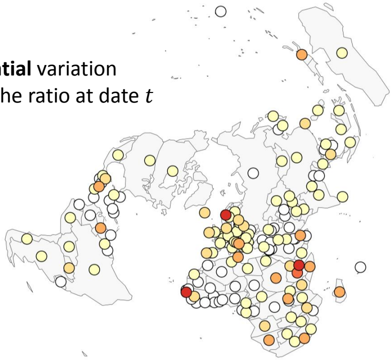
t

	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

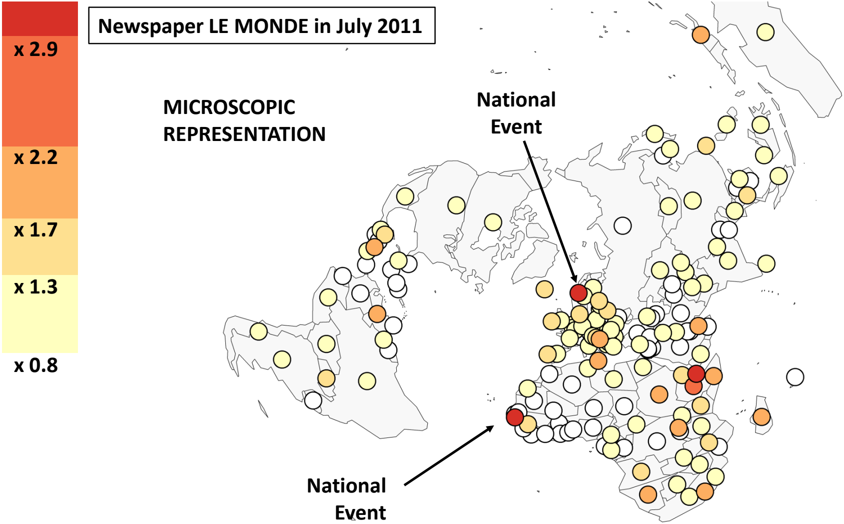
Observed-to-expected ratio of citation number

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

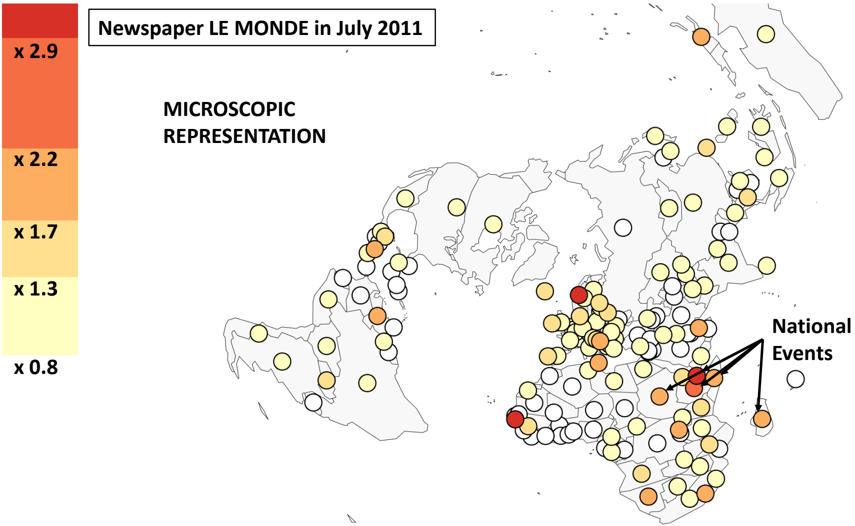
Spatial variation of the ratio at date t



Detection of Media Events



Detection of Media Events

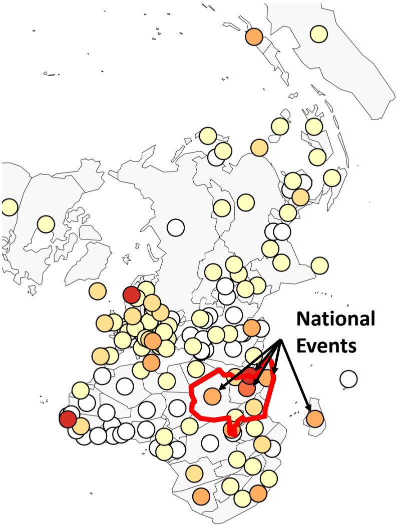


Detection of Media Events

π_1 π_2 π_3 **Space**

Time t

	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**

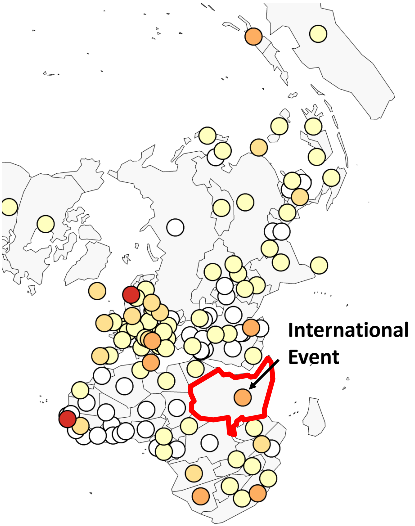


Detection of Media Events

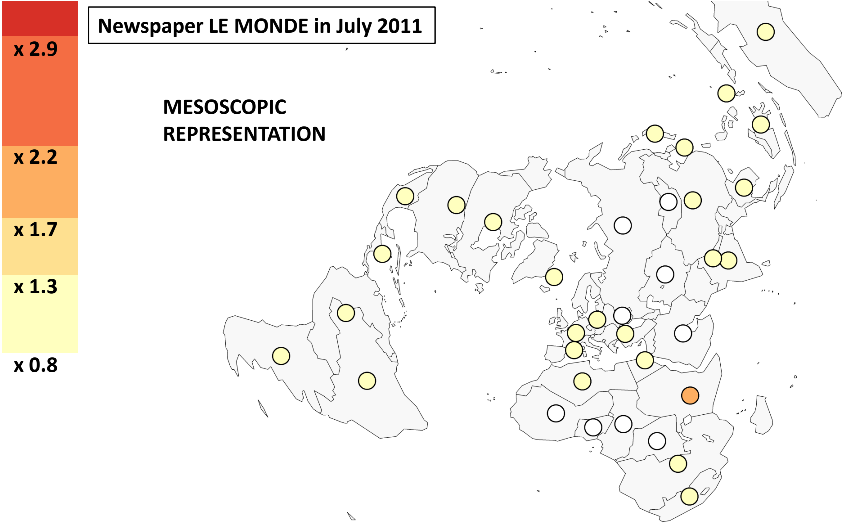
π_1 π_2 π_3 **Space**

Time t

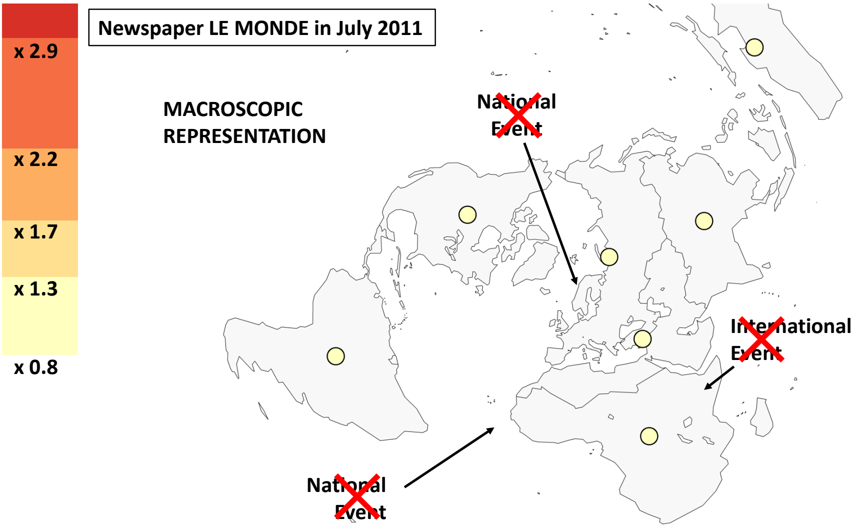
	USA	Aggregate	Israel	...	Total
2 May	25	13+11+10	4	...	142
9 May	14	6+12+12	5	...	108
16 May	20	11+12+6	9	...	142
23 May	15	9+6+13	5	...	120
30 May	10	16+17+9	4	...	137
6 June	14	16+16+9	4	...	114
13 June	15	14+17+9	6	...	119
20 June	17	13+12+12	7	...	123
27 June	7	6+7+20	2	...	103
4 July	12	13+8+10	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



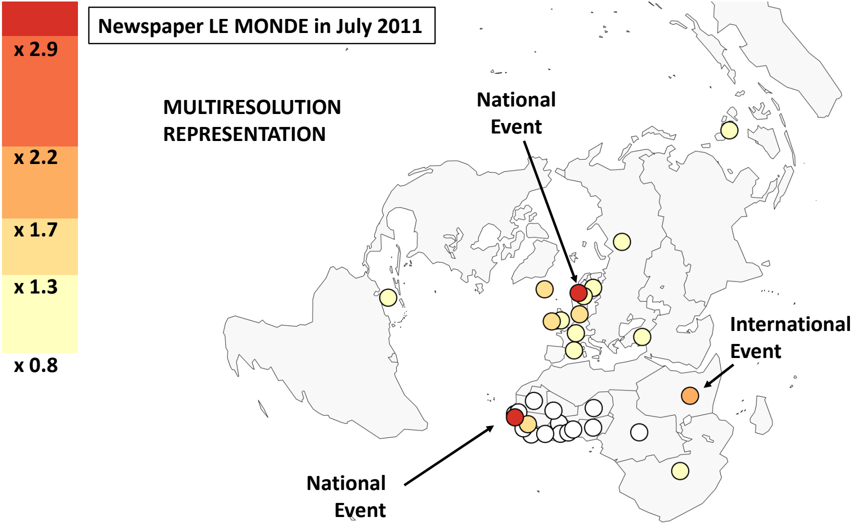
Problem I: Granularity of Geographical Information



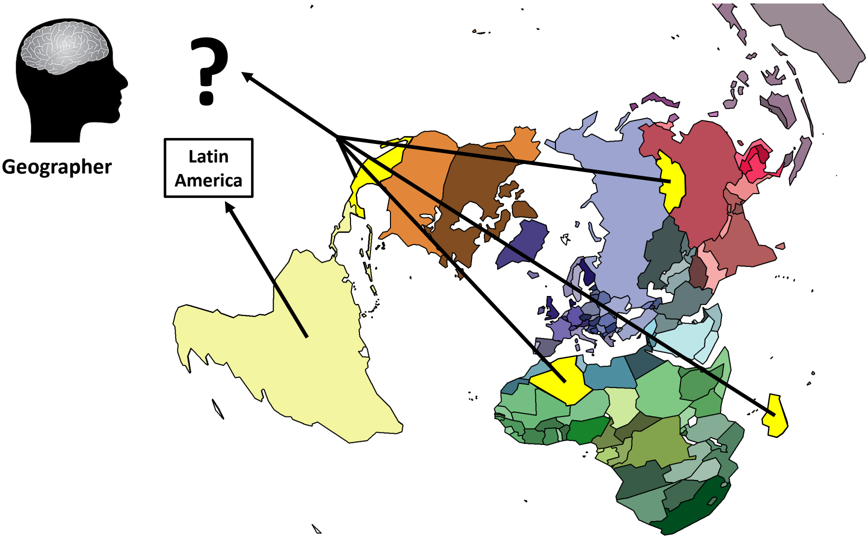
Problem I: Granularity of Geographical Information



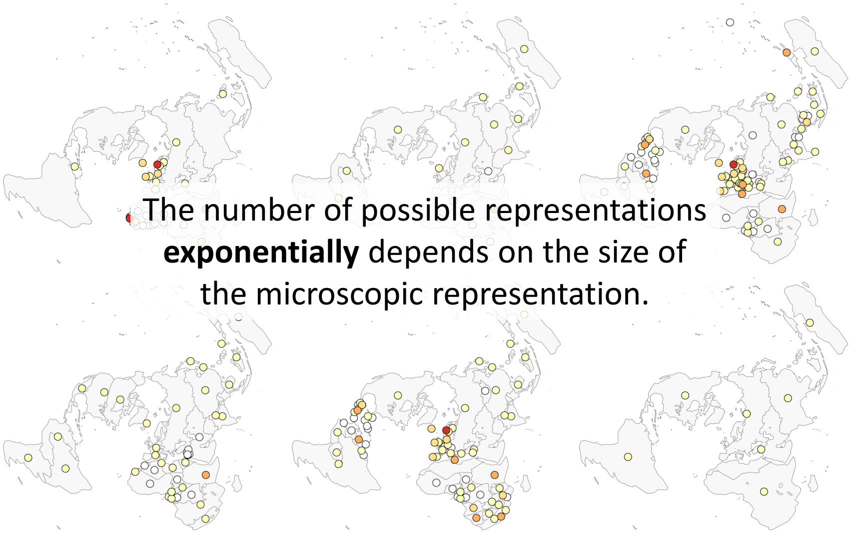
Problem I: Granularity of Geographical Information



Problem II: Semantics of Geographical Aggregates



Problem III: Computation of Optimal Aggregates



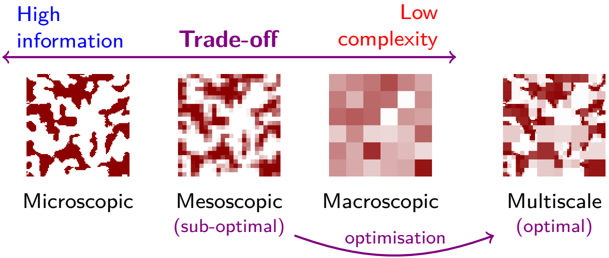
The number of possible representations **exponentially** depends on the size of the microscopic representation.

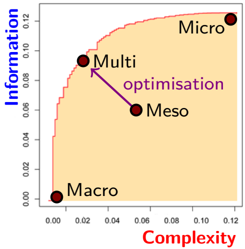
Outline

Part I

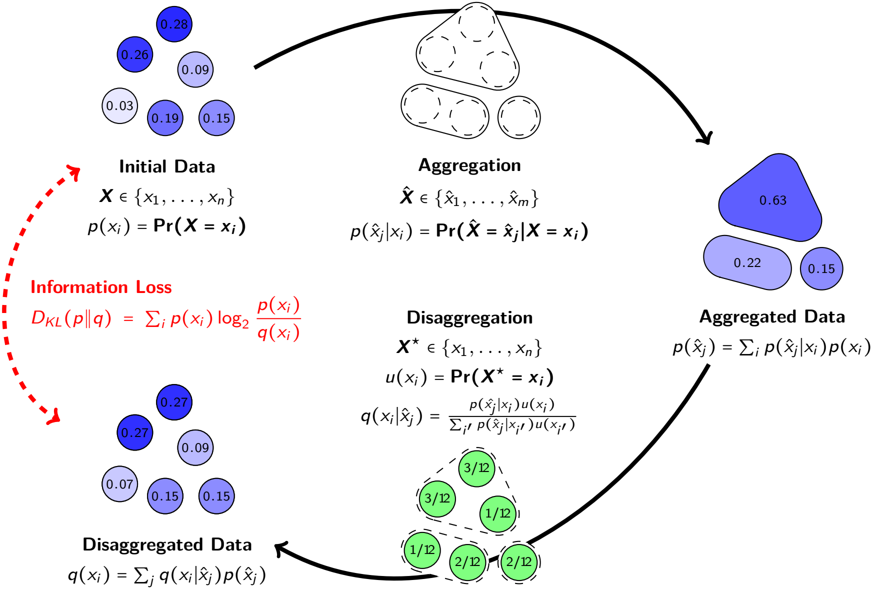
Measuring Scales with Information Theory

Objectives: Description / Explanation / Prediction of complex systems

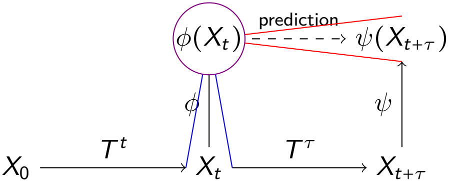




Kullback-Leibler Divergence for Information Loss



Information Bottleneck for Optimal Prediction



$T(X_{t+1}\|X_t)$	Markovian Kernel
$X_0 \in \Sigma$	Initial State
$X_t \in \Sigma$	Current State
$X_{t+\tau} \in \Sigma$	Future State
$\psi : \Sigma \to \mathcal{S}_\psi$	Post-measurement
$\phi : \Sigma \to \mathcal{S}_\phi$	Pre-measurement

Applying the Information Bottleneck Method: [Tishby *et al.*, 1999]

- Maximising the **Predictive Capacity** $\max_\phi \ \mathrm{I}(\phi(X_t); \psi(X_{t+\tau}))$
- Minimising the **Measurement Complexity** $\min_\phi \ \mathrm{I}(X_t; \phi(X_t))$

- Minimising the **Bottleneck Variational**

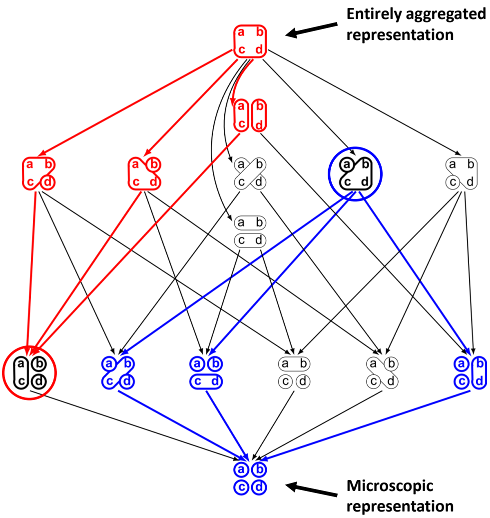
$$\min_\phi \quad \mathrm{I}(X_t; \phi(X_t)) \ - \ \beta \ \mathrm{I}(\phi(X_t); \psi(X_{t+\tau})) \quad \text{with } \beta \in \mathbb{R}^+$$

where $\mathrm{I}(X; Y)$ is the **Mutual Information** between X and Y

Part II

Formalising Semantics with Combinatorial Constraints

Solution Space: Set of Partitions



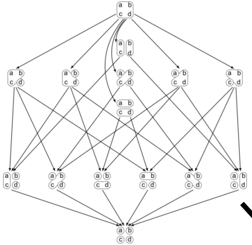
Algebraic Structure

A partial order on the set of possible partitions

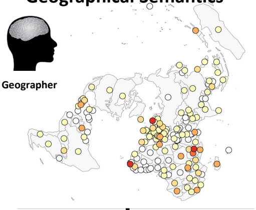
→ **the refinement relation**

Integrating the Semantics into the Aggregation Process

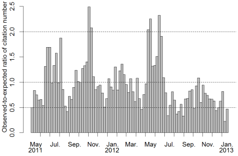
Set of <u>possible</u> partitions



Geographical Semantics

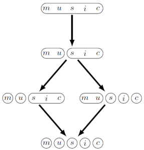


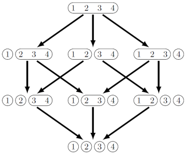
Temporal Semantics



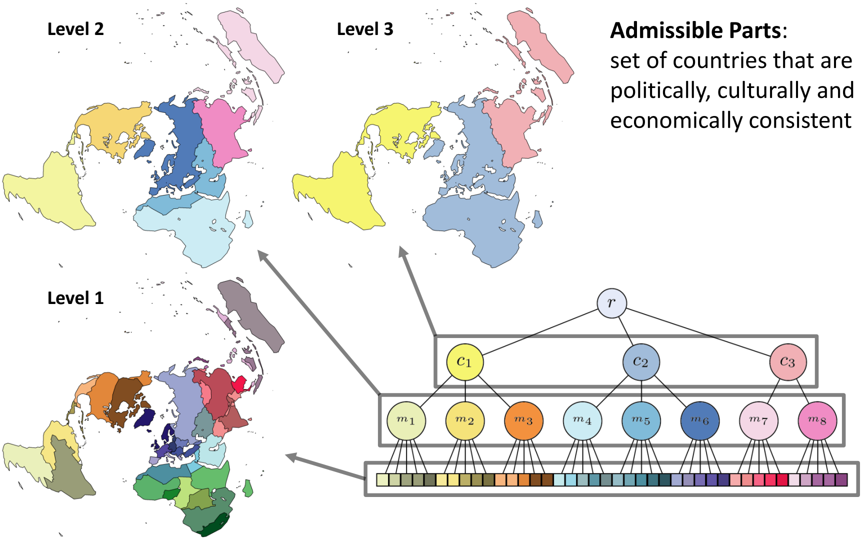
Constraints

Set of <u>admissible</u> partitions



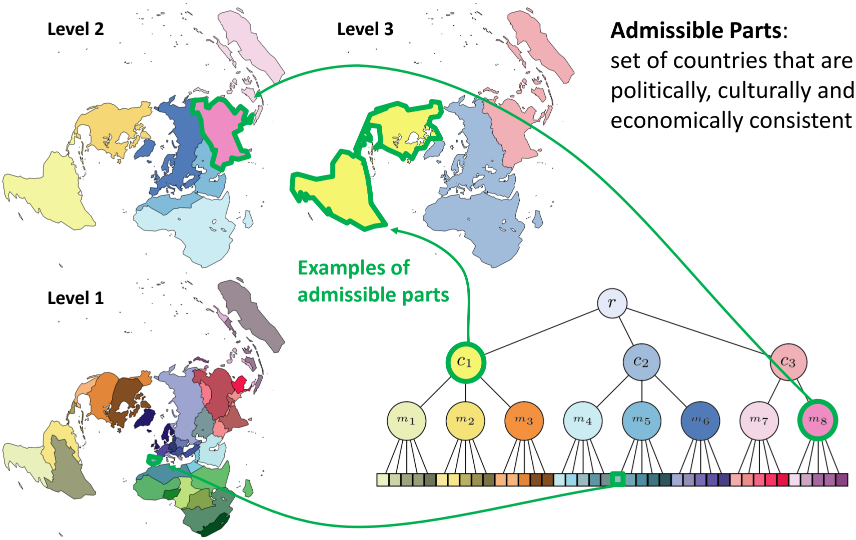


Preserving a Hierarchical Structure



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

Preserving a Hierarchical Structure

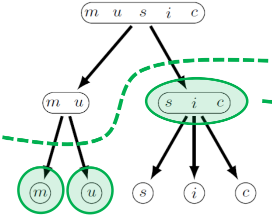


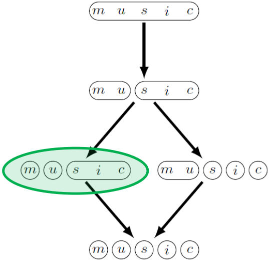
Admissible Parts:
set of countries that are politically, culturally and economically consistent

Preserving a Hierarchical Structure

Admissible Parts
(nodes of the hierarchy)

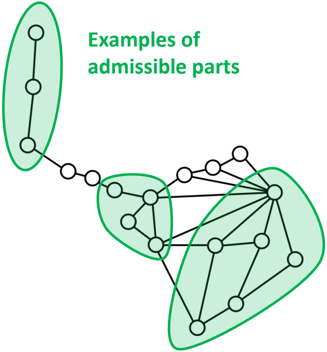
Admissible Partitions
(cuts of the hierarchy)



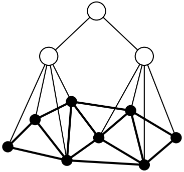


Preserving a Neighbourhood Relation

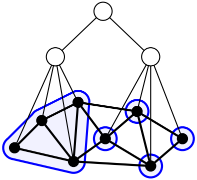
Admissible Parts: set of connected countries regarding the adjacency graph



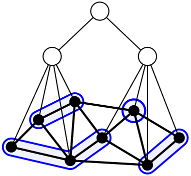
Preserving Mixed Structures



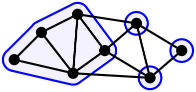
Mixed Structure
(graph + hierarchy)



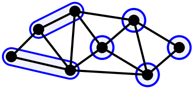
Hierarchical Aggregation
(consistent with the graph)



Graph Aggregation
(not consistent with the hierarchy)

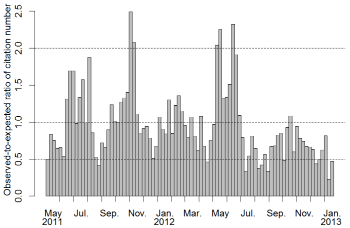


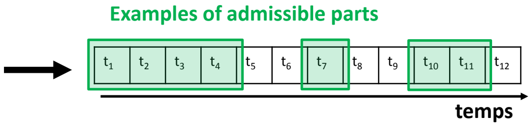
Union of Aggregates
(locally extends a node of the hierarchy)



Intersection of Aggregates
(locally subdivides a node of the hierarchy)

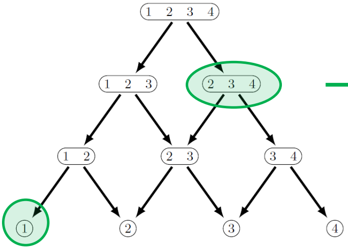
Admissible Parts:
time intervals



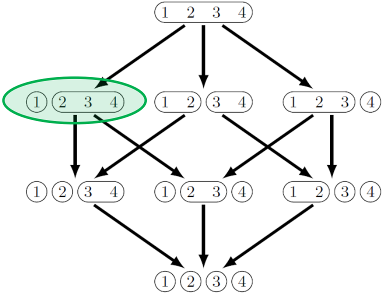


Preserving a Temporal Order

Admissible Parts
(time intervals)

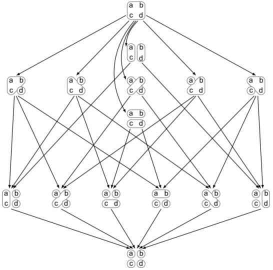


Admissible Partitions
(interval sequences)

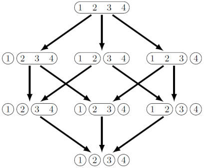


Complexity of Algebraic Structures

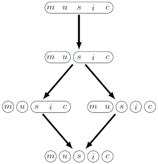
Non-constrained partitions



Admissible partitions according to a **total order**



Admissible partitions according to a **hierarchy**



Less constrained ⟵⟶ **More constrained**

More complex ⟵⟶ **Less complex**

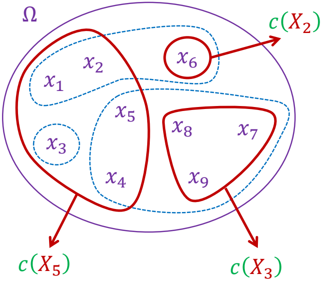
Part III

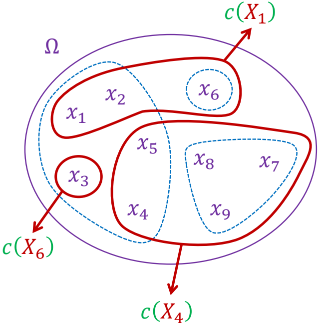
Solving the Optimal Partition Problem

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 $c : \mathcal{P} \to \mathbb{R}$
- the corresponding set of admissible partitions $\mathfrak{P} = \{\mathcal{X} \subset \mathcal{P}$ such that $\mathcal{X}$ is a partition of $\Omega\}$



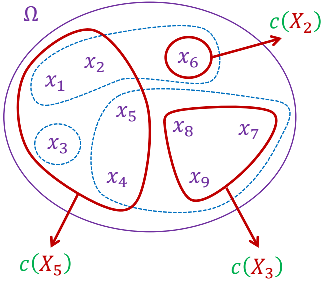


The Set Partitioning Problem

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 $c : \mathcal{P} \to \mathbb{R}$
- the corresponding set of admissible partitions $\mathfrak{P} = \{\mathcal{X} \subset \mathcal{P}$ such that $\mathcal{X}$ is a partition of $\Omega\}$

Additional assumptions



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

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

→ NP-complete!

Special Versions of the Set Partitioning Problem

Multilevel Geographical Analysis

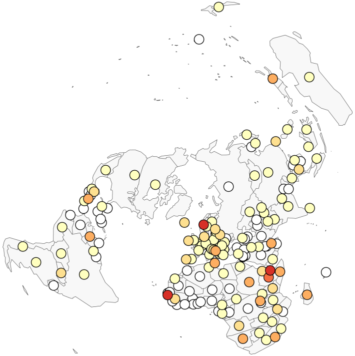
- $\Omega =$ territorial units
- $\mathcal{P} =$ admissible aggregates
- $c =$ compression rate
- $\mathfrak{P} =$ aggregated representations

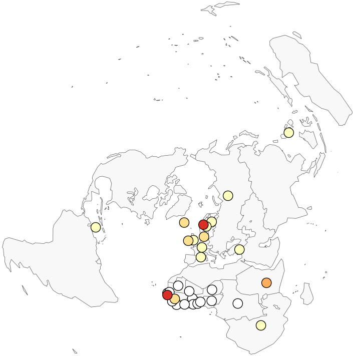
Hierarchical SPP

- 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]





Multilevel Geographical Analysis

Time Series Analysis

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

Hierarchical SPP

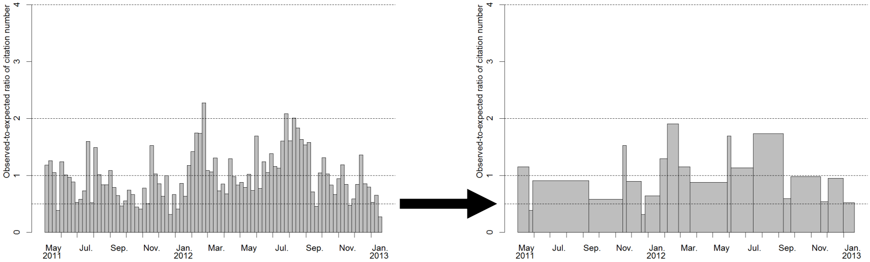
- Assumption: $\mathcal{P}$ forms a hierarchy
- Result: $\mathcal{O}(n)$ depth-first search
  [Pons *et al.*, 2011] [Lamarche-Perrin *et al.*, 2014]

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]



Special Versions of the Set Partitioning Problem

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]

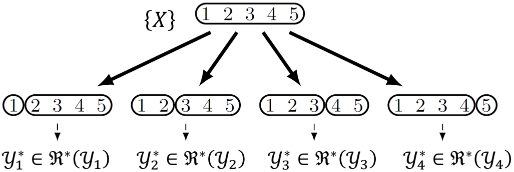
SPP with Size Bounds [Rothkopf *et al.*, 1998]

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

Dynamic Programming

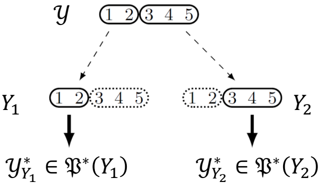
DIVIDE…

Branching the state space according to the covering relation

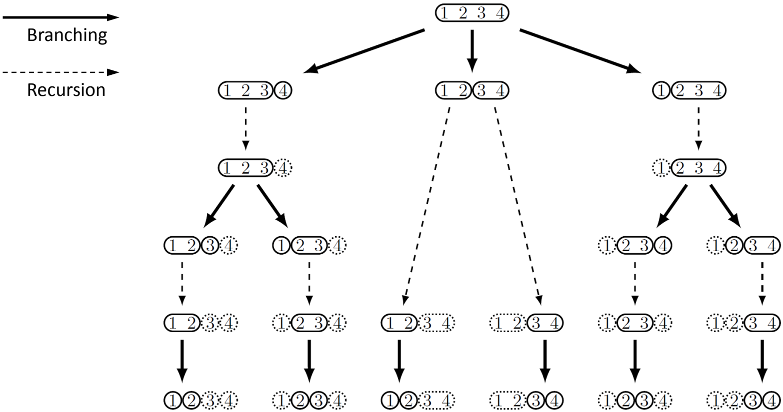


…AND CONQUER

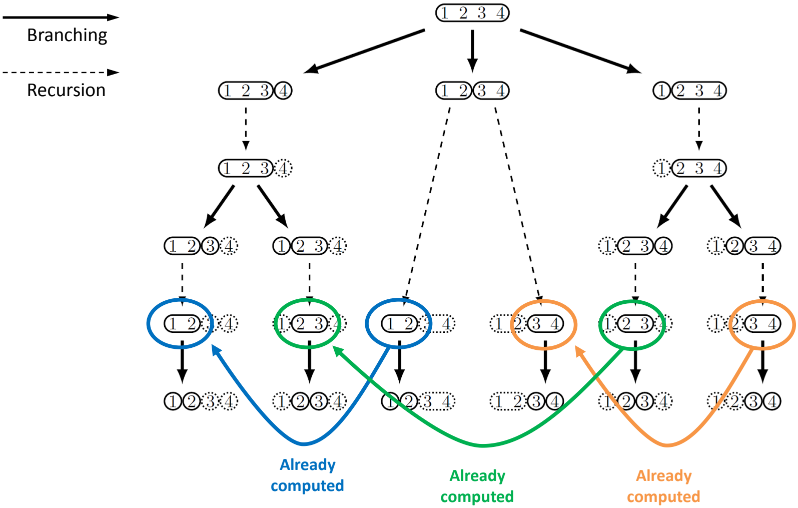
Recursion according to the principle of optimality



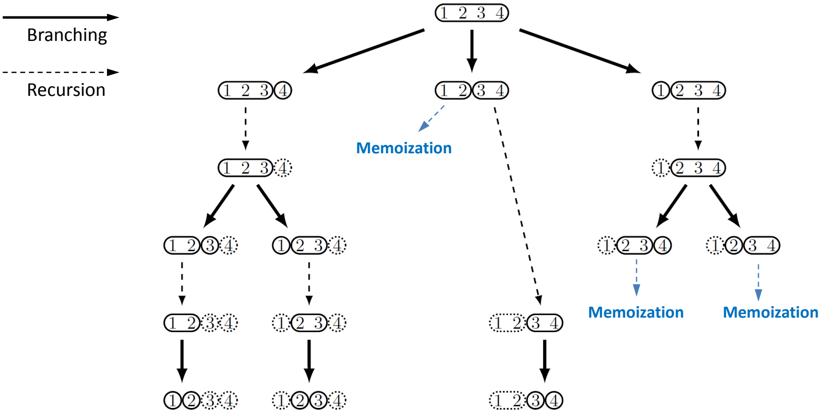
Dynamic Programming



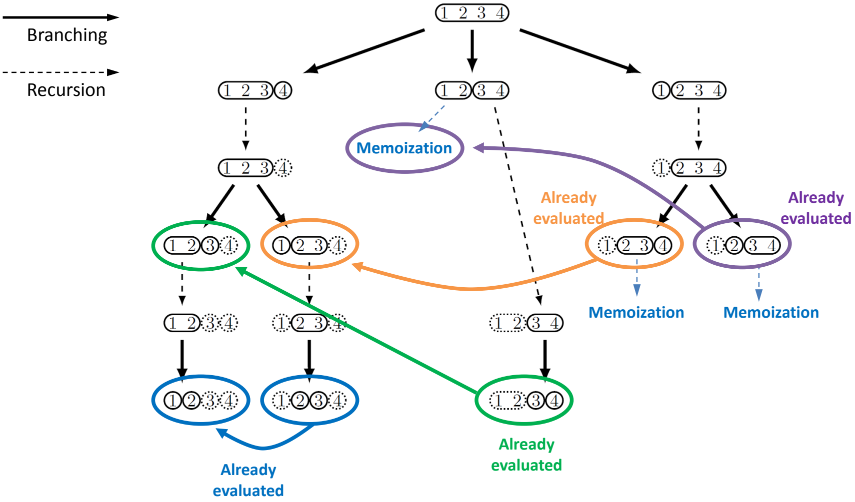
Dynamic Programming



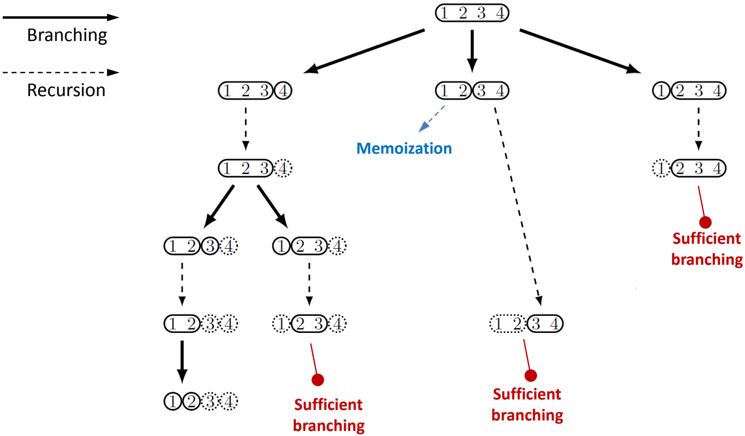
Dynamic Programming



Dynamic Programming



Dynamic Programming



Generic Algorithmic Framework

A Generic Algorithm to Solve the SPP

Global Inputs:

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

Local Inputs:

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

Output:

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

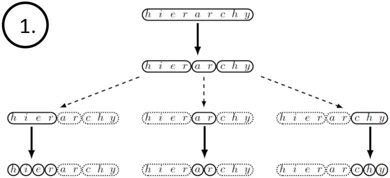
- If the algorithm has already been applied to part X, return the locally-optimal partition recorded in $\mathfrak{L}$.
- Initialization: $\mathcal{X}^* \leftarrow \{\{X\}\}$ and $\mathfrak{D}' \leftarrow \mathfrak{D}$.
- For each $\mathcal{Y} \in \mathfrak{C}(\{X\})$ such that $\overline{\mathcal{X}} \cup \mathcal{Y}$ does not refine any partition in $\mathfrak{D}$, do the following:
  - For each part $Y \in \mathcal{Y}$, call the algorithm with local inputs $X \leftarrow Y$, $\overline{\mathcal{X}} \leftarrow \overline{\mathcal{X}} \cup \mathcal{Y} \backslash \{Y\}$, 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 $\mathfrak{L}$.

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

1.



2.

Data Structure

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

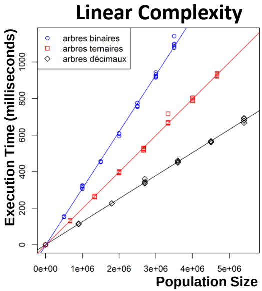
3.

Algorithm 1 for the HSPP

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

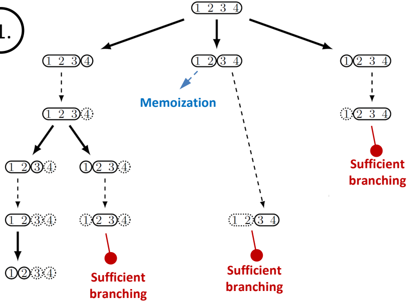
Ensure: Each node of the tree has a Boolean label *optimalCut* representing an optimal partition (see above).

procedure SOLVEHSPP($node$)
  if $node$ has no child **then**
    $node.optimalCost \leftarrow node.cost$
    $node.optimalCut \leftarrow true$
  else
    $MCost \leftarrow node.cost$
    $\mu Cost \leftarrow 0$
    for each $child$ of $node$ **do**
      SOLVEHSPP($child$)
      $\mu Cost \leftarrow \mu Cost + child.optimalCost$
    $node.optimalCost \leftarrow \max(\mu Cost, MCost)$
    $node.optimalCut \leftarrow (\mu Cost < MCost)$



Generic Algorithmic Framework

1.



2.

Data Structure

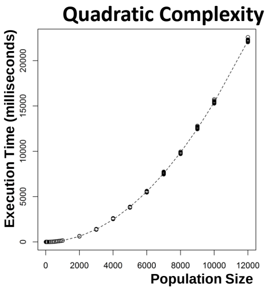
- Set of parts: triangular matrix
- Optimal partition: array of cuts
- Algorithm: dynamic programming

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).

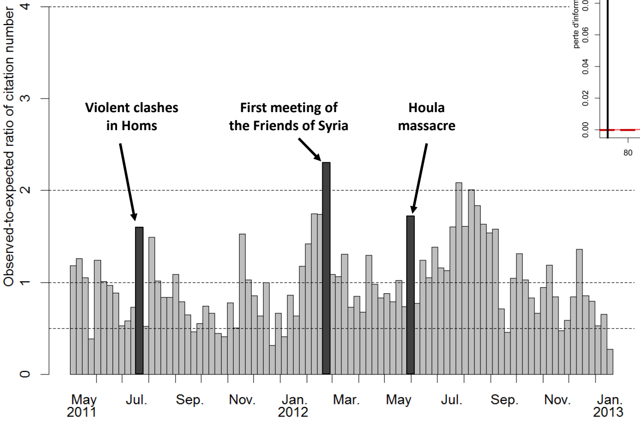
for $j \in [\![1, n]\!]$ **do**
  $optimalCost[j] \leftarrow cost[1, j]$
  $optimalCut[j] \leftarrow 1$
  for $cut \in [\![2, j]\!]$ **do**
    $\mu Cost \leftarrow optimalCost[cut-1] + cost[cut, j]$
    if $\mu Cost > optimalCost[j]$ **then**
      $optimalCost[j] \leftarrow \mu Cost$
      $optimalCut[j] \leftarrow cut$

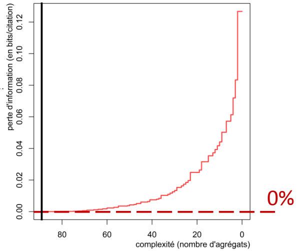


Part IV

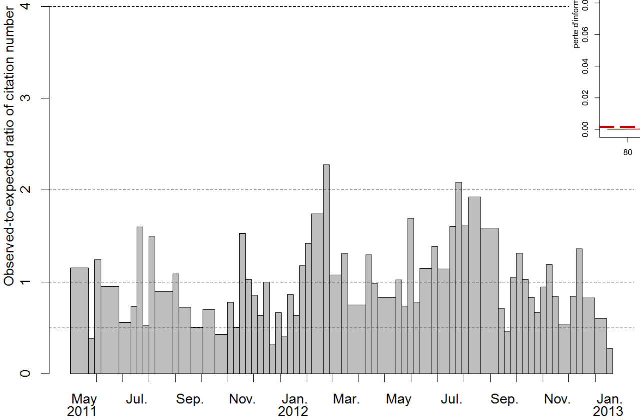
Results and Experiments

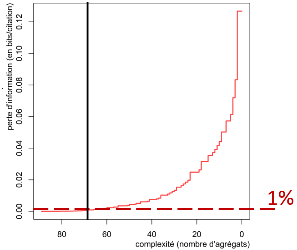
The Syrian Civil War according to *Le Monde* [TCCI 2014]



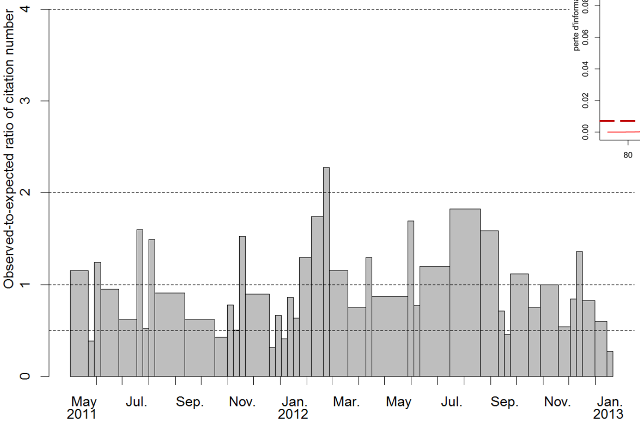


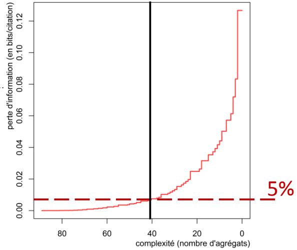
The Syrian Civil War according to *Le Monde* [TCCI 2014]

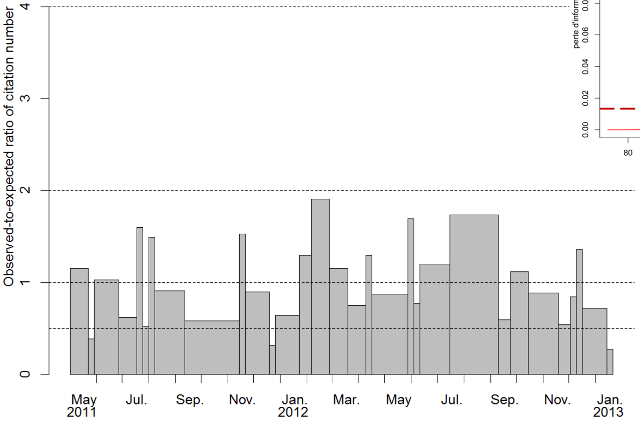


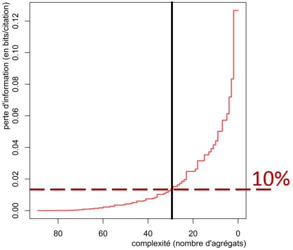


The Syrian Civil War according to *Le Monde* [TCCI 2014]

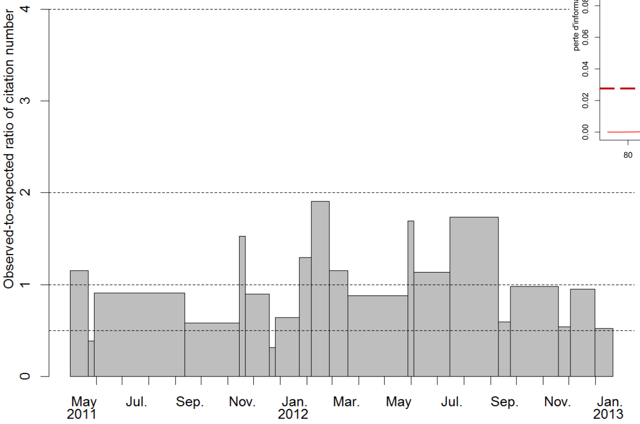


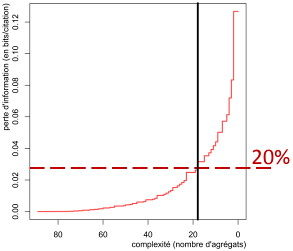



Observed-to-expected ratio of citation number
0
1
2
3
4
May 2011
Jul.
Sep.
Nov.
Jan. 2012
Mar.
May
Jul.
Sep.
Nov.
Jan. 2013


perte d'information (en bits/citation)
0.00
0.02
0.04
0.06
0.08
0.10
0.12
80
60
40
20
0
complexité (nombre d'agrégats)
10%

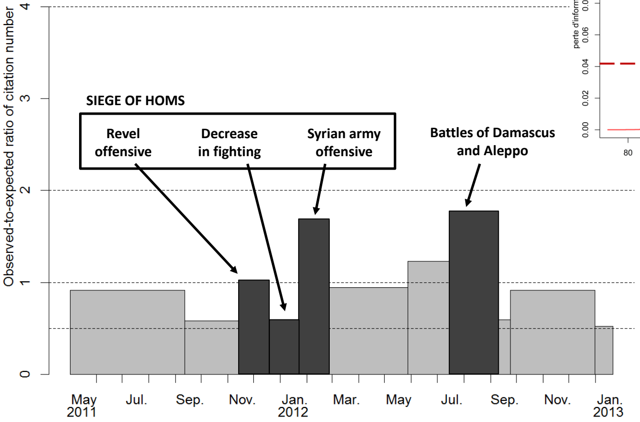
The Syrian Civil War according to *Le Monde* [TCCI 2014]

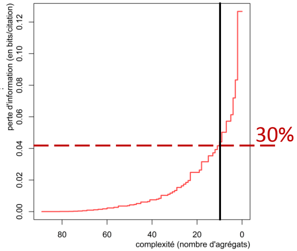




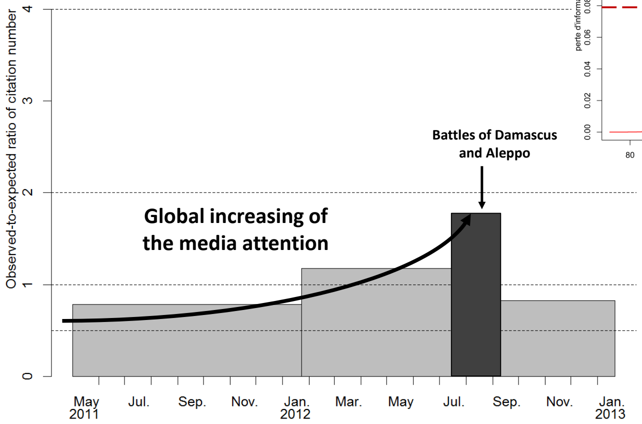
The Syrian Civil War according to *Le Monde* [TCCI 2014]

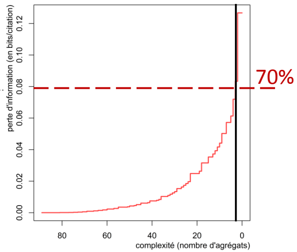
Source: Wikipedia
Timeline of the Syrian civil war
Siege of Homs





The Syrian Civil War according to *Le Monde* [TCCI 2014]

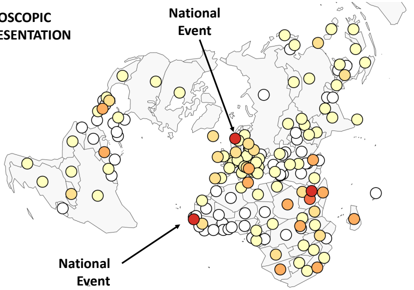




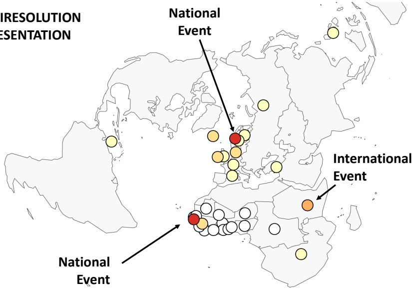
Geographical Aggregation

[TCCI 2014] News coverage of *Le Monde* in July 2011

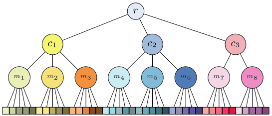
MICROSCOPIC REPRESENTATION

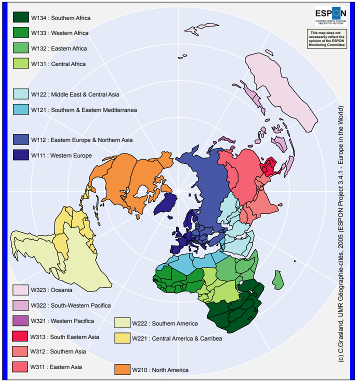


MULTIRESOLUTION REPRESENTATION



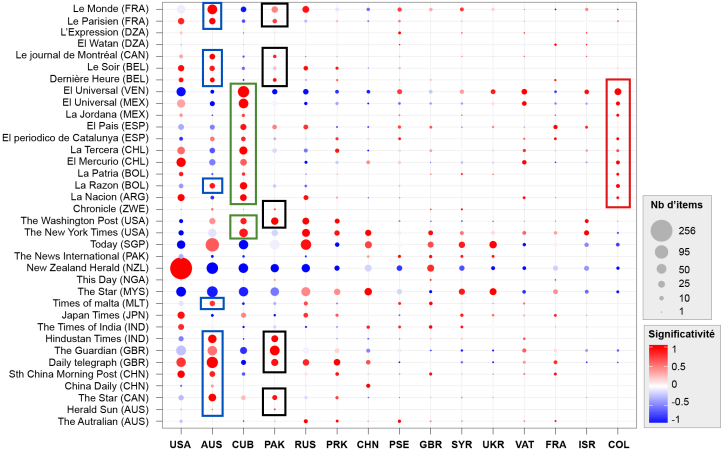
Territorial Hierarchy (WUTS)





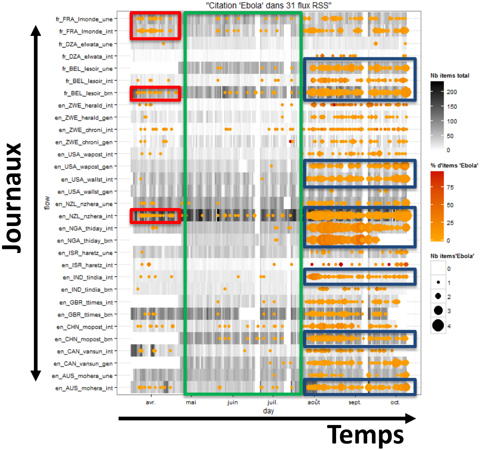
Media Aggregation

[Espace Géo 2016] News coverage the week of December 15, 2014

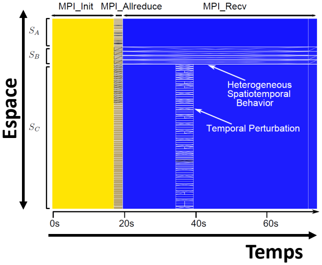


Multidimensional Aggregation

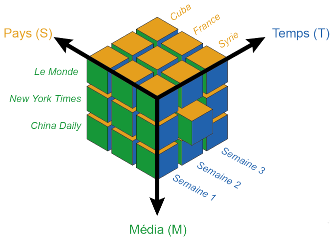
Media-temporal Aggregation



Spatio-temporal Aggregation



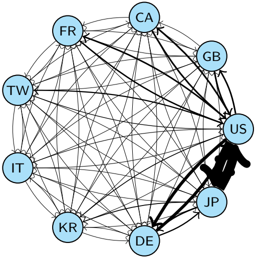
Geomedia Cube

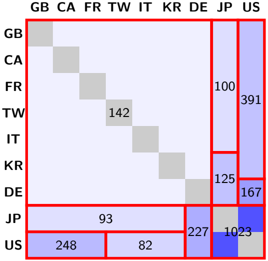


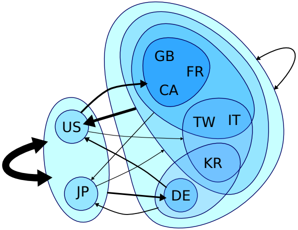
Aggregation of Social Relations

National Patent Citations. Unit: 100 patents; Period: 1990–1999;
Source: NBER U.S. Patent Citations Data File (http://www.nber.org/patents/)

	GB	CA	FR	TW	IT	KR	DE	JP	US
GB		3	5	1	2	0	11	23	82
CA	3		3	2	1	0	6	15	89
FR	5	3		1	3	1	14	28	83
TW	2	3	2		1	3	4	22	62
IT	2	1	3	1		0	7	12	31
KR	2	1	2	2	1		3	47	44
DE	11	6	12	2	6	1		78	167
JP	24	14	23	9	9	14	66		504
US	86	87	75	37	29	16	161	519	



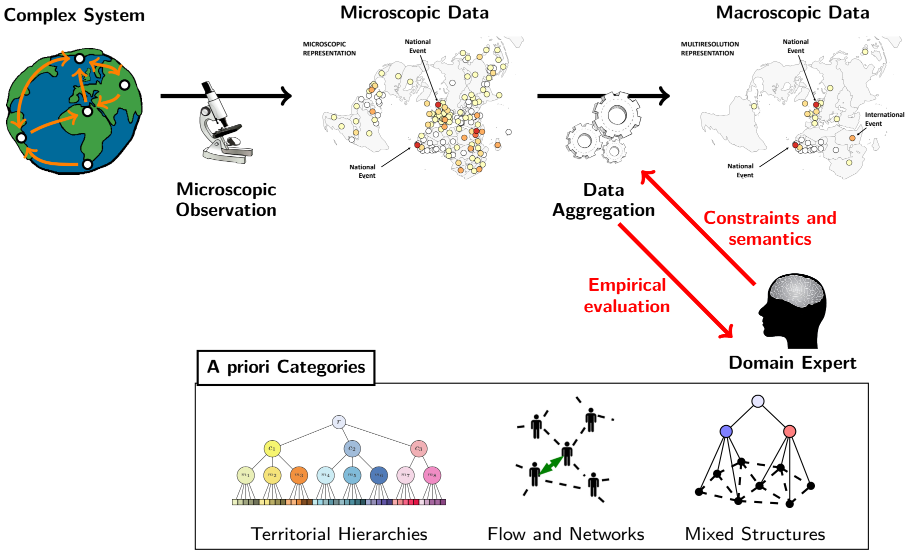




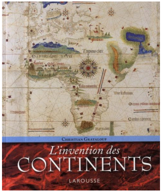
Part V

Final Thoughts

Conclusion

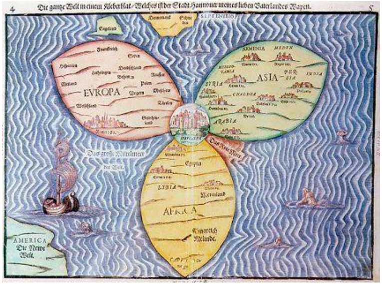


Complex System
Microscopic Data
Macroscopic Data
MICROSCOPIC REPRESENTATION
National Event
National Event
MULTIRESOLUTION REPRESENTATION
National Event
International Event
National Event
Microscopic Observation
Data Aggregation
Constraints and semantics
Empirical evaluation
Domain Expert
A priori Categories
Territorial Hierarchies
Flow and Networks
Mixed Structures

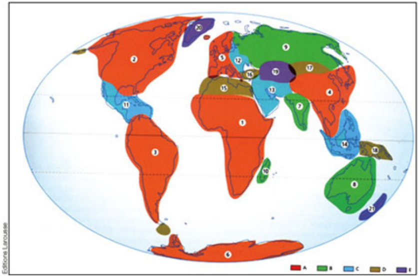


Christian Grataloup. 2009. *L'invention des continents : comment l'Europe a découpé le monde.* Paris, Larousse.

Outdated categories...



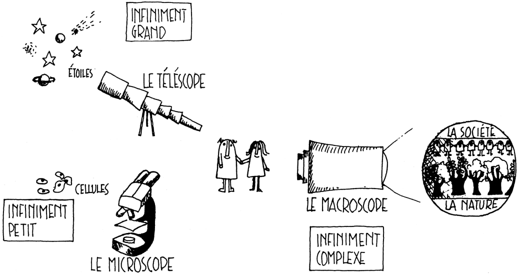
Building new ones?



Thank you for your attention

Mail: Robin.Lamarche-Perrin@lip6.fr

Web: www-complexnetworks.lip6.fr/~lamarche/



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

Joël de Rosnay. 1975. *Le Macroscope.*