

# **Abstraction et observation macroscopique des systèmes**

Après-midi traces  
20 mars 2012

**Robin Lamarche-Perrin**

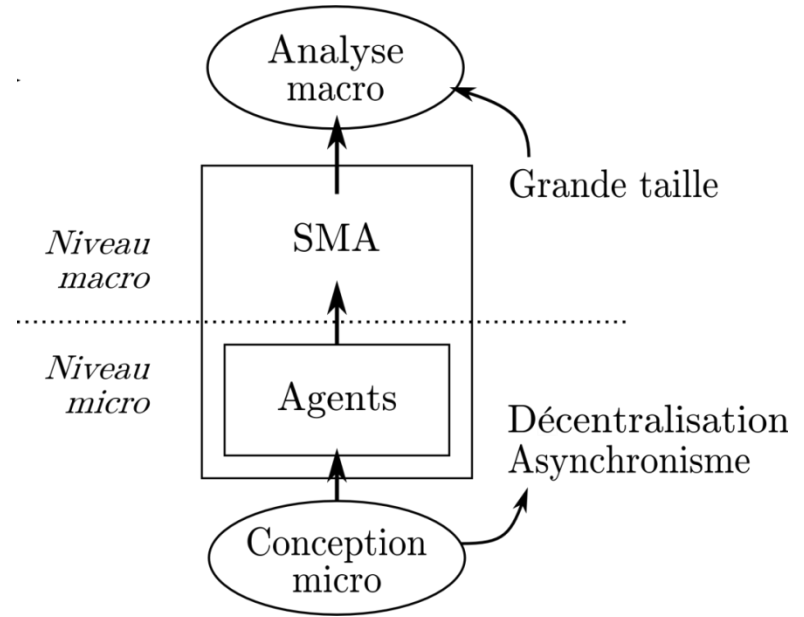
Jean-Marc Vincent (MESCAL)

Yves Demazeau (MAGMA)



# Problème général

- Systèmes
  - Décentralisés
  - Asynchrones
  - Complexes
  - De très grande taille



Problématique : comment avoir un point de vue macroscopique de processus microscopiques ?

Réponse : engendrer des abstractions !

# Qu'est-ce qu'une abstraction ?

Conception du système

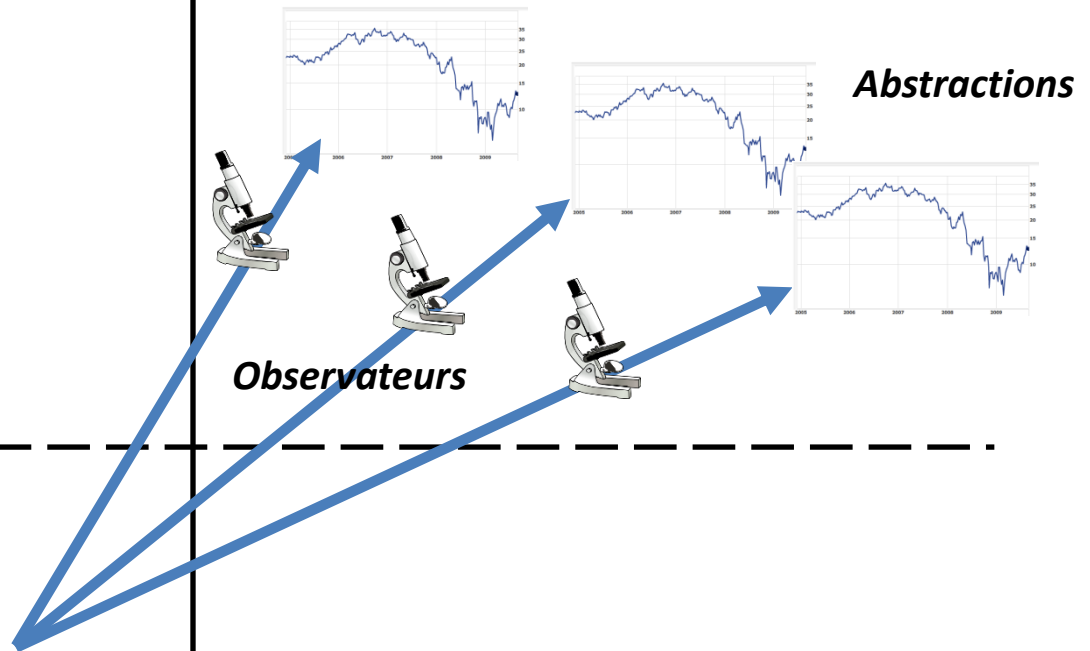
Description du système

Niveau macroscopique

Niveau microscopique

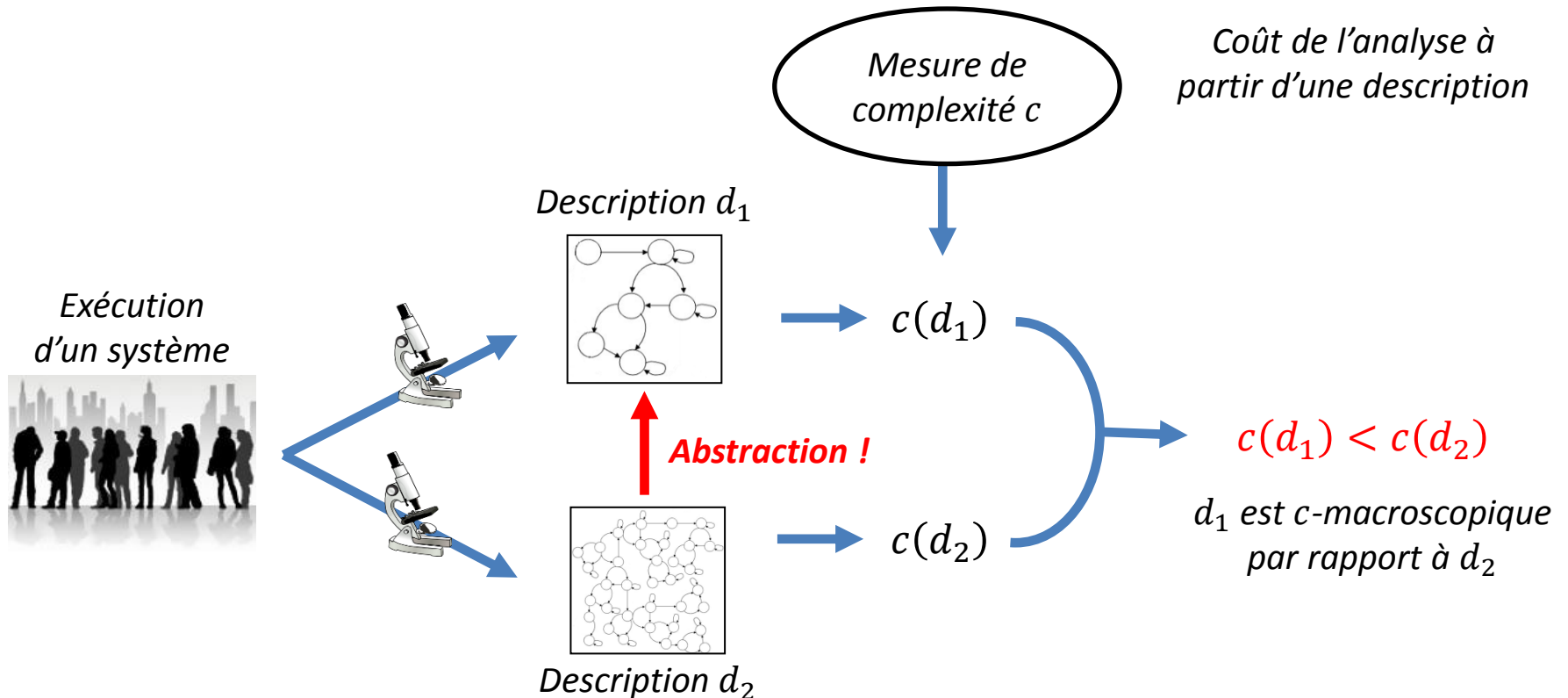


*Entités  
microscopiques*



# COMMENT MESURER ET ÉVALUER LES DESCRIPTIONS

# Descriptions et complexité

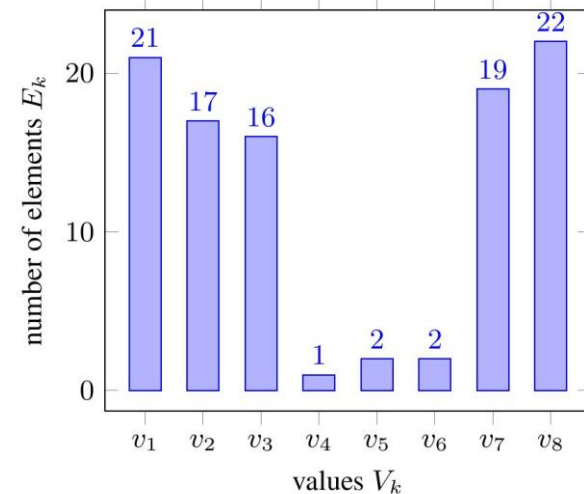


# Exemple minimal de descriptions



$V = 8$  valeurs

## Description

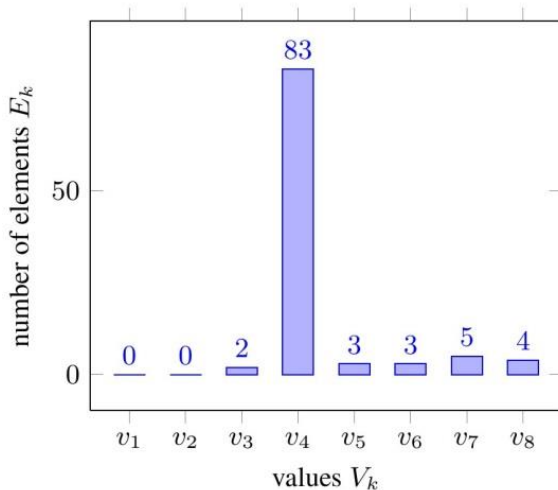


# Entropie des descriptions

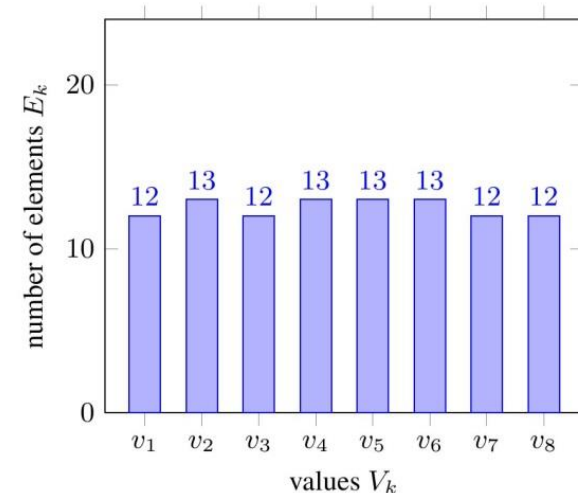
- Mesure la quantité d'information nécessaire pour coder une description donnée
- Mesure l'ordre des éléments selon les valeurs de la description

## Formule de Shannon

$$H = - \sum_{i \in V} \frac{E_k}{E} \log_2 \left( \frac{E_k}{E} \right)$$



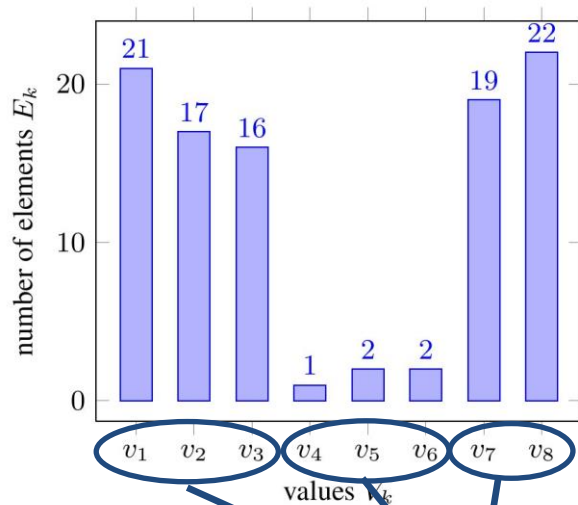
**Entropie minimale**



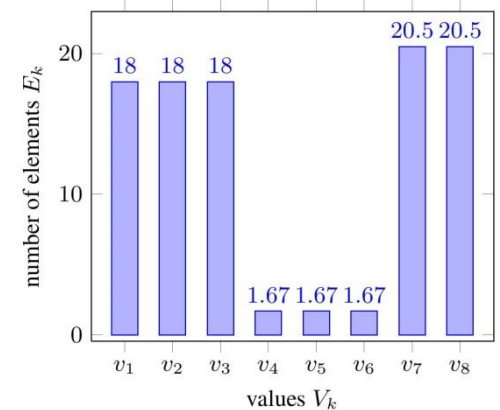
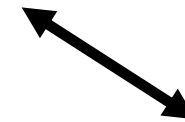
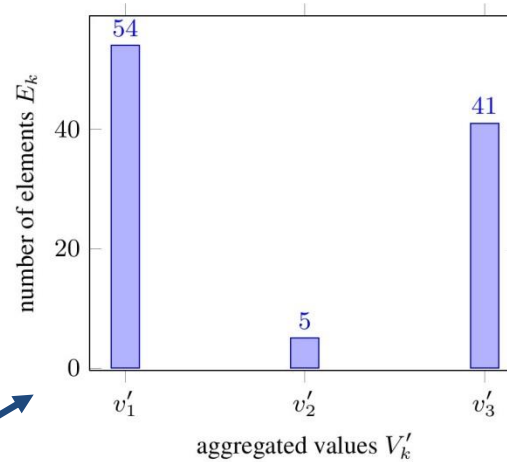
**Entropie maximale**

# Agrégation de descriptions

Description source



Description agrégée

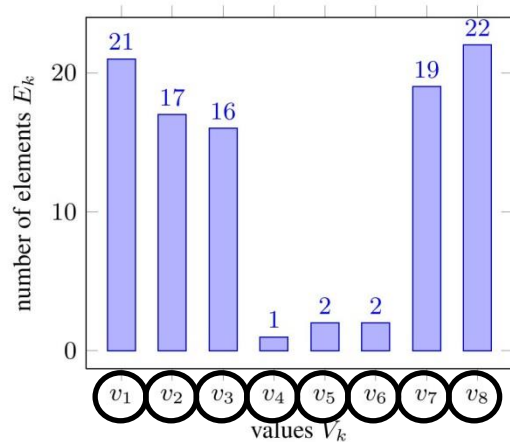




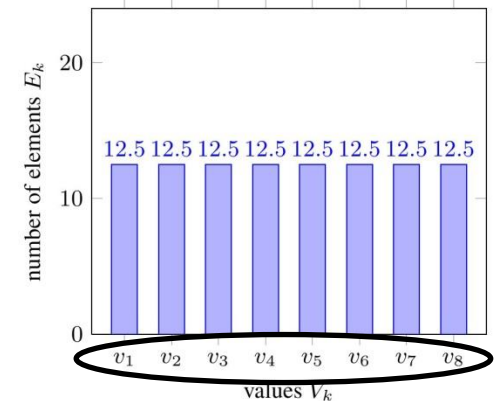
# Gain entropique d'une agrégation

- Mesure la quantité d'information économisée par une agrégation
- Dépend de la distribution initiale et de la fonction d'agrégation

$$G = H_{source} - H_{agrégée}$$



**Gain minimal**

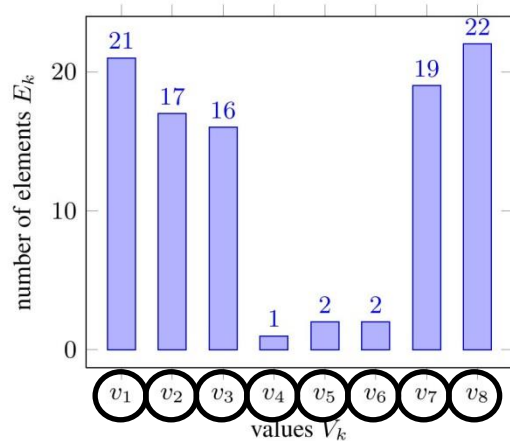


**Gain maximal**

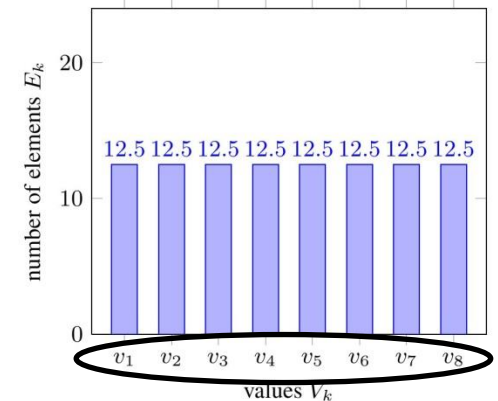
# Perte d'information d'une agrégation

- Mesure la quantité d'information nécessaire pour désagréger une description
- Dépend uniquement de la fonction d'agrégation

$$L = - \sum_{k \in V'} \frac{E'_k}{E} \log_2 \left( \frac{1}{V_k} \right)$$



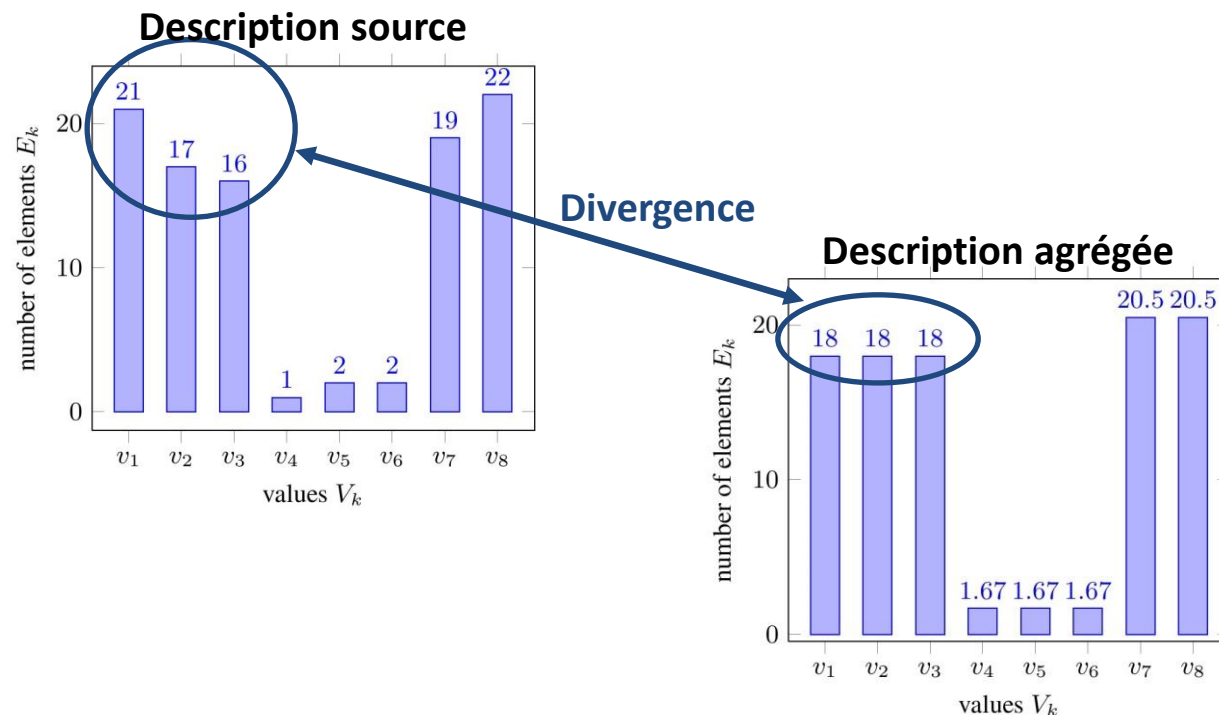
**Information conservée**



**Information perdue**

# Divergence d'une agrégation

- Mesure l'écart entre une description agrégée et sa description source
- Dépend de la distribution initiale et de la fonction d'agrégation



**Formule de Kullback-Leibler**

$$D = - \sum_{k \in V} \frac{E_k}{E} \log_2 \left( \frac{E'_k}{E \times V_k} \right)$$

$$D = L - G$$

# Mesure globale

- Représente un compromis

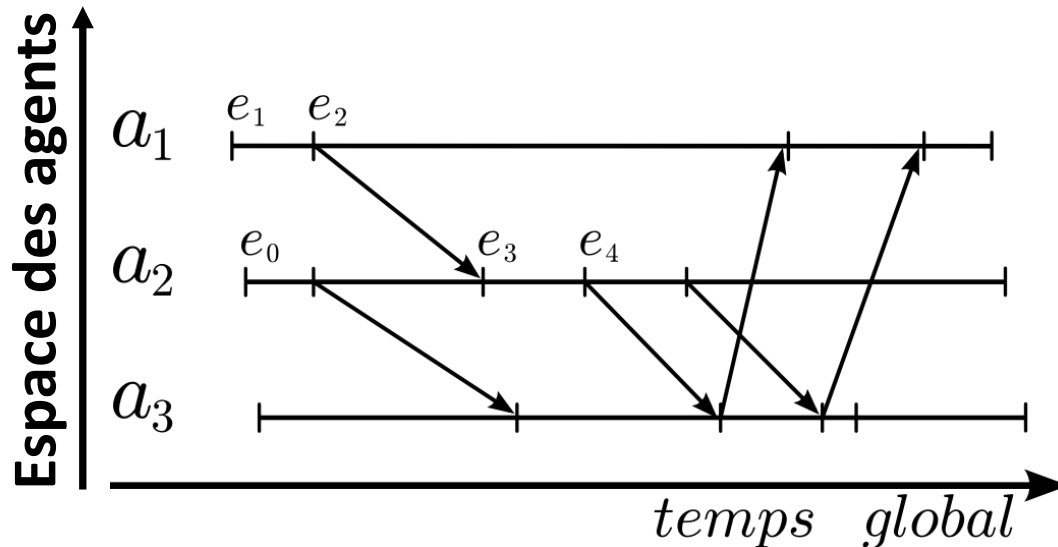
Perte d'information vs. Gain entropique

$$D_{\alpha} = (1 - \alpha) \times L - \alpha \times G$$

- $\alpha = 0 \rightarrow$  pas d'agrégation, information conservée
- $\alpha = 1 \rightarrow$  agrégation de toutes les valeurs
- Principe d'agrégation : trouver les descriptions qui minimisent la mesure globale

# DESCRIPTIONS CAUSALES

# Description de la causalité

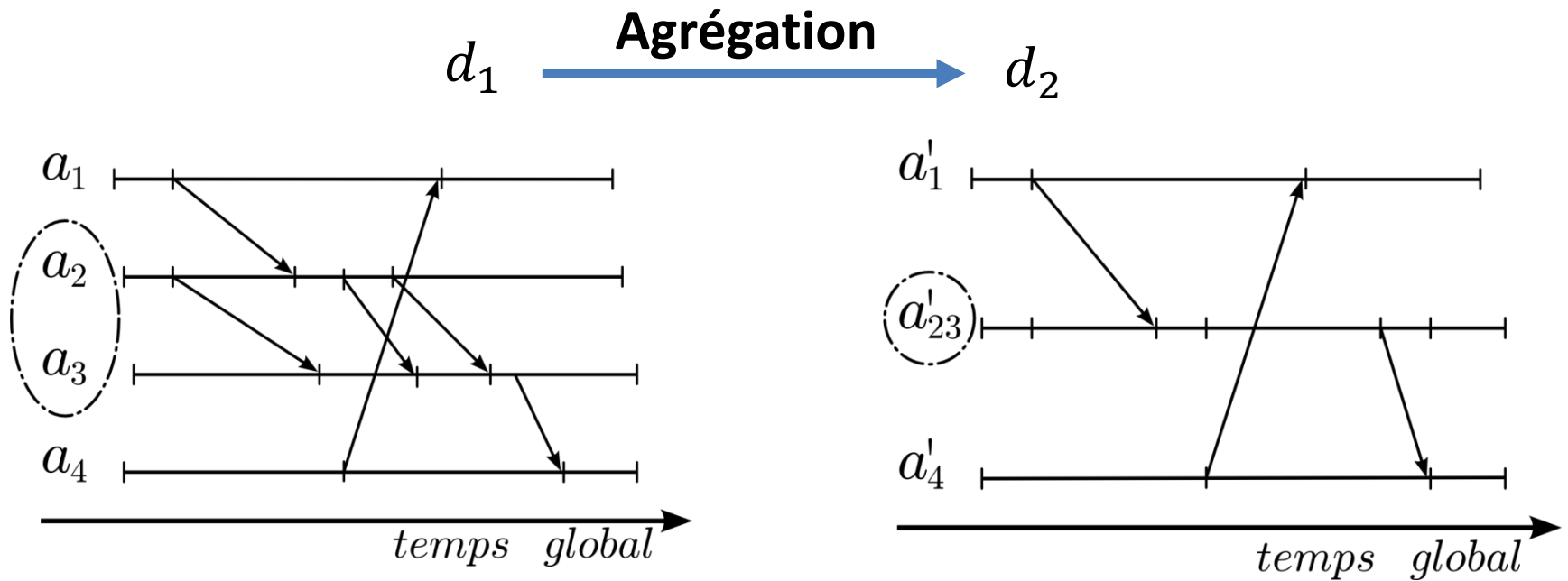


Temps logique global :  $e_1 < e_2 < e_3 < e_4$

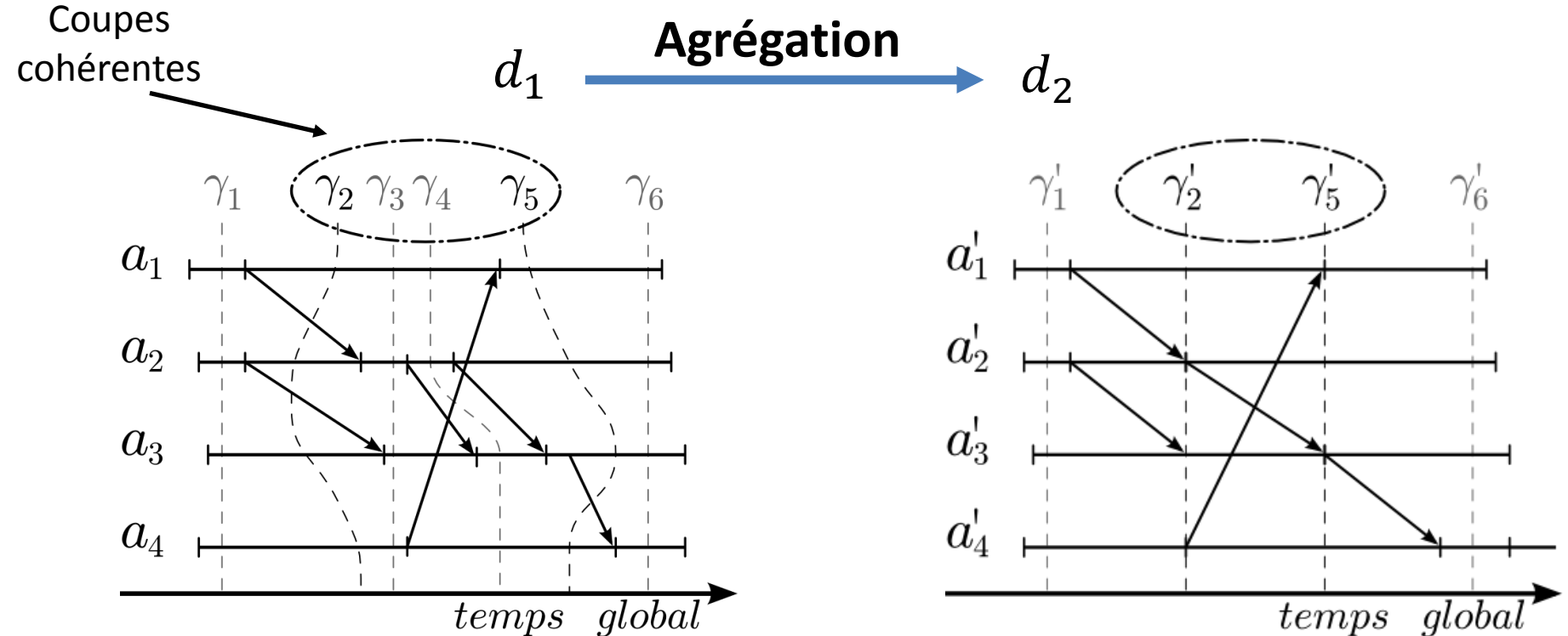
Indépendance causale :  $e_0 \parallel e_1$

F. Mattern, Virtual Time and Global States of Distributed Systems, *Parallel and Distributed Algorithms*, p. 215–226, North-Holland, Elsevier, 1989.

# Agrégations spatiales



# Agrégations temporelles



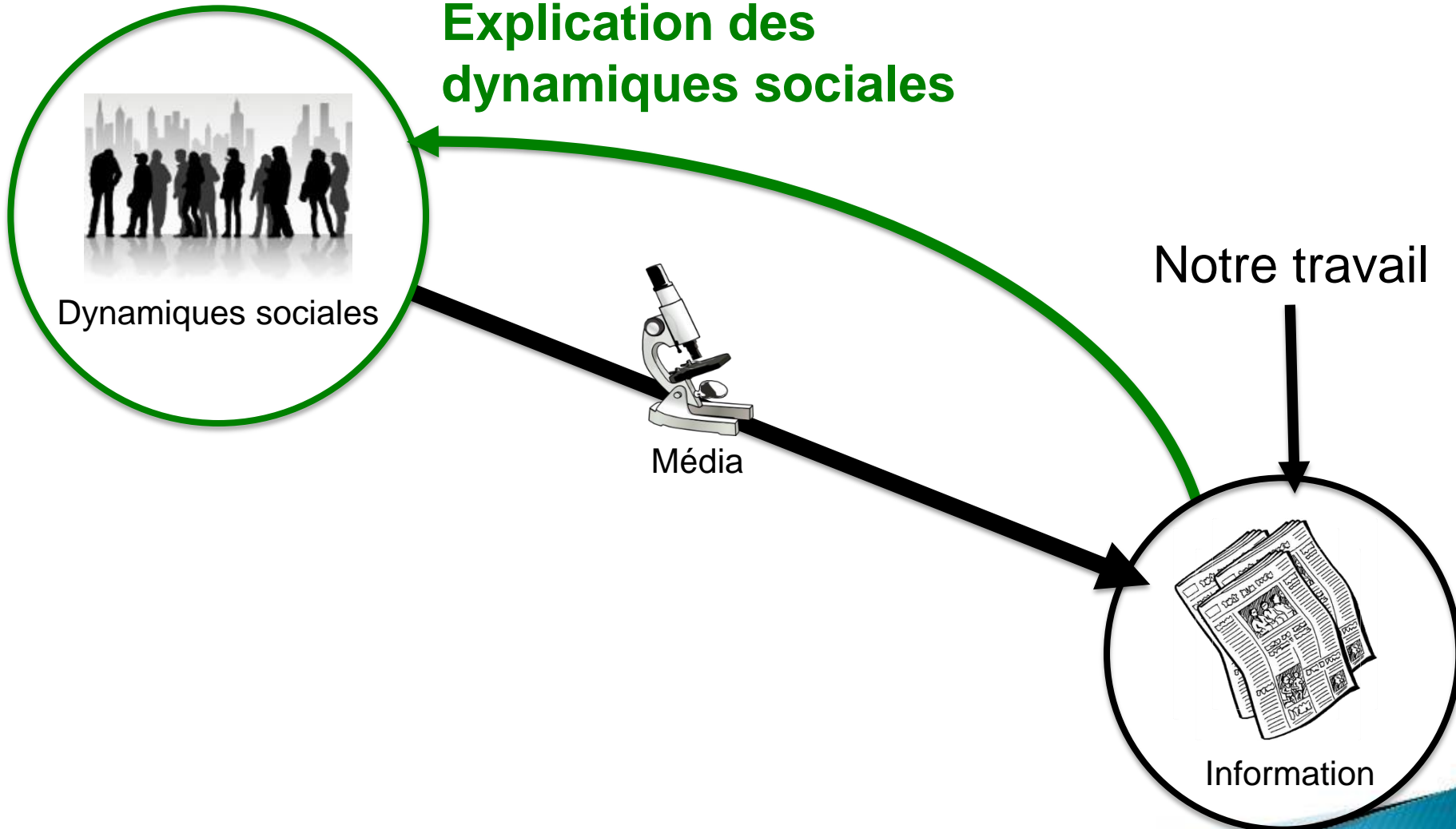
Robin Lamarche-Perrin, Yves Demazeau, Jean-Marc Vincent. Observation macroscopique et émergence dans les SMA de très grande taille. *JFSMA'11*:53-62, Valenciennes, France, oct. 2011.



# **AGRÉGATION D'INFORMATIONS MÉDIATIQUES**

# Objectif de l'analyse

Explication des dynamiques sociales



# Objectif de l'analyse



Dynamiques sociales

Évaluation des  
instruments  
d'observation



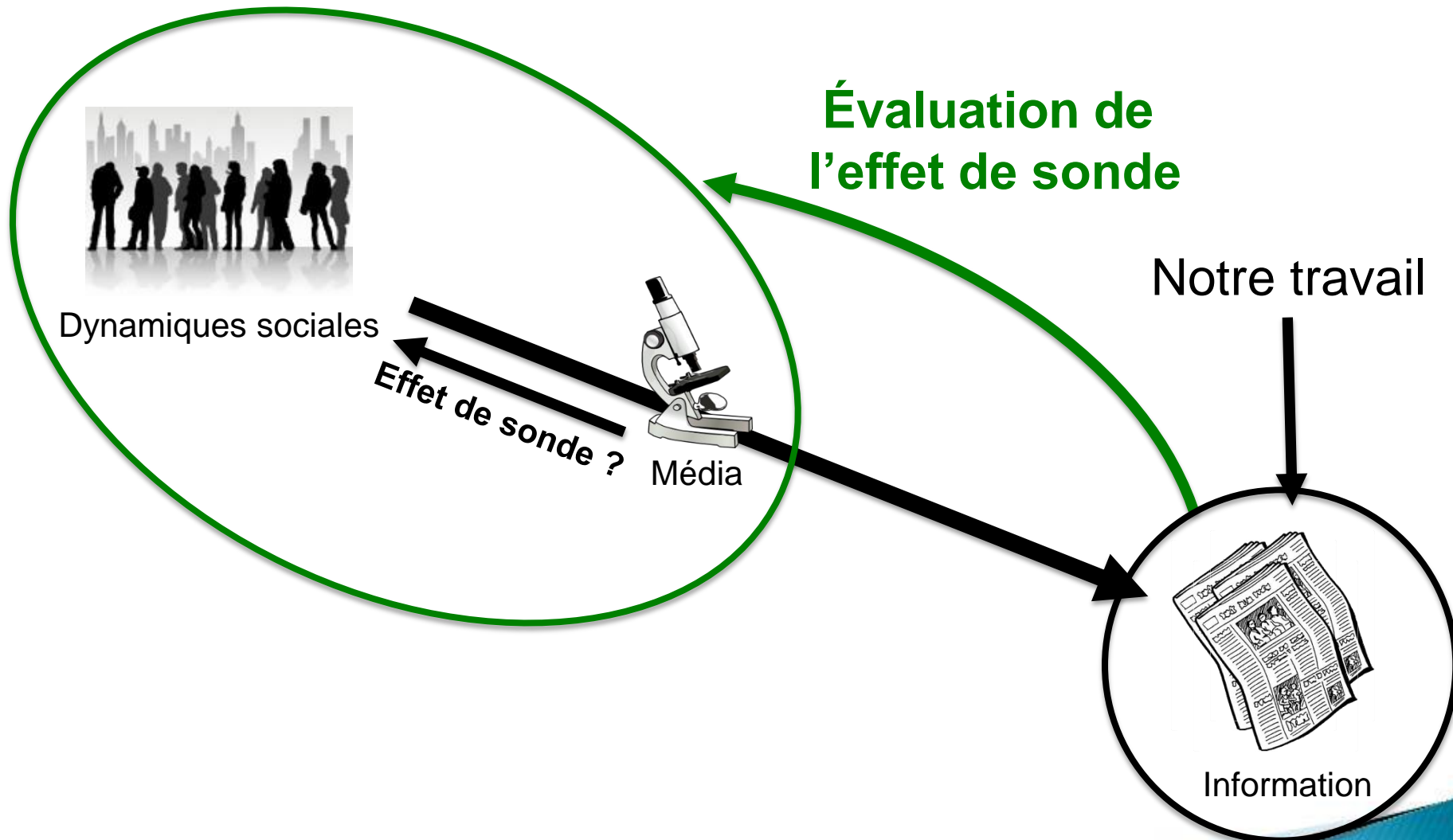
Média

Notre travail



Information

# Objectif de l'analyse



# Données et difficultés

- Projet GEOMEDIA
  - Collaboration avec le CIST (Paris 7)
  - Plateforme d'analyse des informations médiatiques
- 1. Données hétérogènes, non-structurées
- 2. Grandes quantités d'information
  - **70 flux RSS** pendant **10 mois** → **500 000 items**
    - The Gardian : **100 items/jours**
    - Les 10 plus actifs : **62 items/jours**
  - Exemple : analyse d'un phénomène sur **5 ans** selon **10 points de vue** → **1 130 000 articles** !

# Descriptions médiatiques

Dimensions		Références indirectes		Références directes
		Absolues	Relatives	
<b>Espace</b>	Où ?	« France » « Allemagne »	« Nos voisins »	Lieu de publication
<b>Temps</b>	Quand ?	« 12 octobre 2006 »	« En mai dernier »	Date de publication
<b>Thèmes</b>	Quoi ?	« Accords diplomatiques »	« Selon l'article X »	Ligne éditoriale

- **Descriptions multidimensionnels**

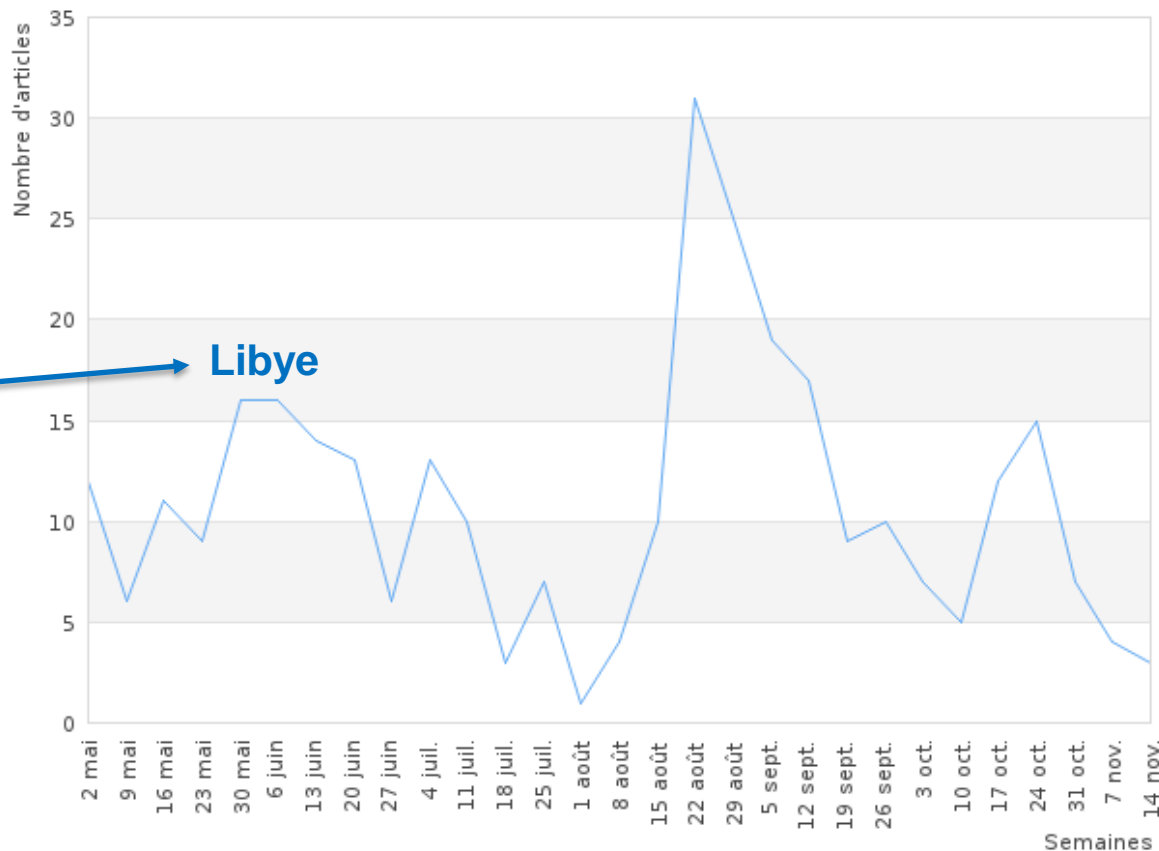
« Causalité » médiatique → Comment ?

Source : Le Monde international,  
3520 articles du 02 mai au 20 nov.

# Espace × Temps

Variation du poids médiatique des territoires

	USA	Libye	Syrie	France	Israël	...	Total
2 mai	25	12	11	10	4	...	142
9 mai	14	6	12	12	5	...	108
16 mai	20	11	12	6	9	...	142
23 mai	15	9	6	13	5	...	120
30 mai	10	16	17	9	4	...	137
6 juin	14	16	16	9	4	...	114
13 juin	15	14	17	9	6	...	119
20 juin	17	13	12	12	7	...	123
27 juin	7	6	7	20	2	...	103
4 juill.	12	13	8	10	6	...	129
11 juill.	21	10	10	14	3	...	107
18 juill.	7	3	8	4	5	...	61
25 juill.	16	7	6	13	4	...	128
1 août	21	1	9	7	4	...	88
8 août	16	4	10	11	4	...	112
15 août	12	10	7	11	9	...	108
22 août	13	31	7	5	3	...	106
...	...	...	...	...	...	...	...
<b>Total</b>	<b>423</b>	<b>308</b>	<b>260</b>	<b>248</b>	<b>153</b>	<b>...</b>	<b>3520</b>



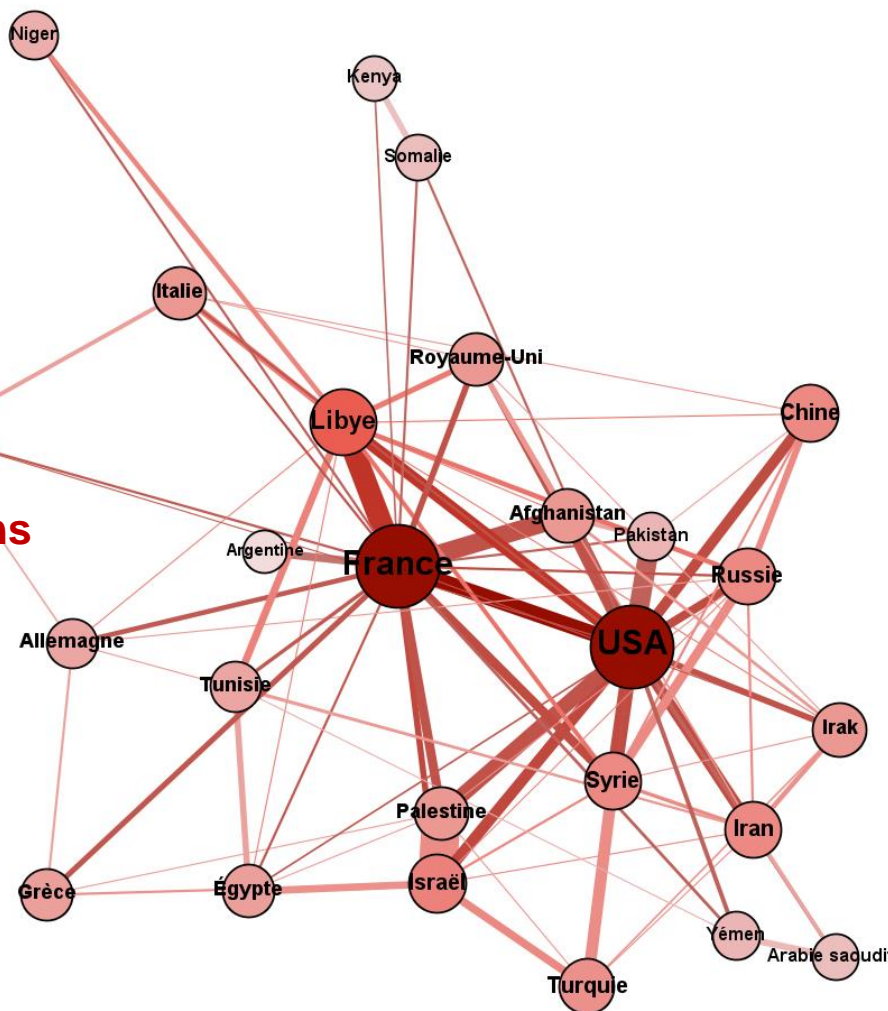
# Espace<sup>2</sup>

## Poids des relations médiatiques entre les territoires

Source : Le Monde international,  
3520 articles du 02 mai au 20 nov.

	USA	France	Libye	Israël	Syrie	Palestine	Afghanistan	...	Total
USA	x	25	19	18	24	24	24	...	423
France	25	x	36	10	15	15	30	...	248
Libye	19	36	x	0	7	0	2	...	308
Israël	18	10	0	x	4	62	0	...	153
Syrie	24	15	7	4	x	1	0	...	260
Palestine	24	15	0	62	1	x	0	...	126
Afghanistan	24	30	2	0	0	0	x	...	131
...	...	...	...	...	...	...	...	...	...
Total	423	248	308	153	260	126	131	...	3520

Relations

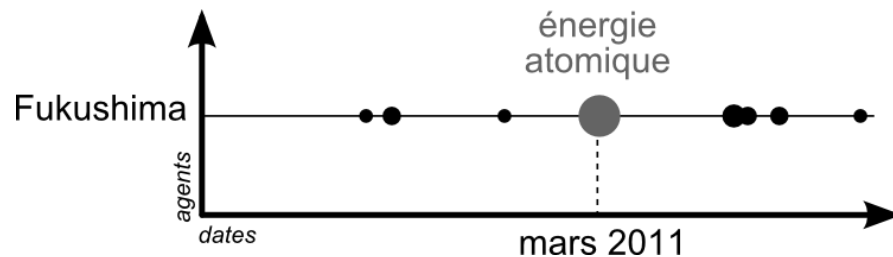




# Algèbre sur les dimensions

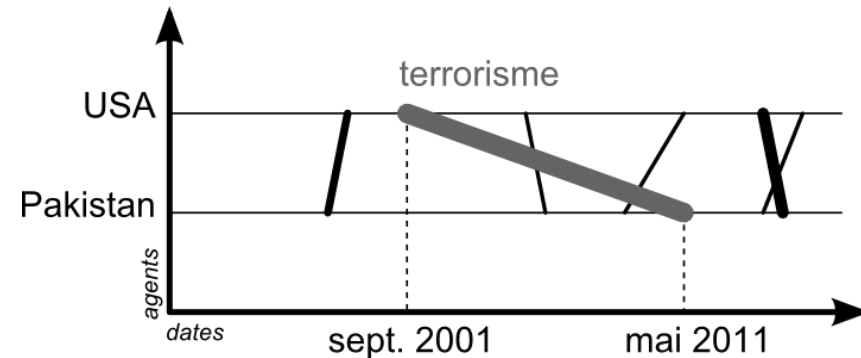
**Espace × Temps × Thème**

Faits médiatiques



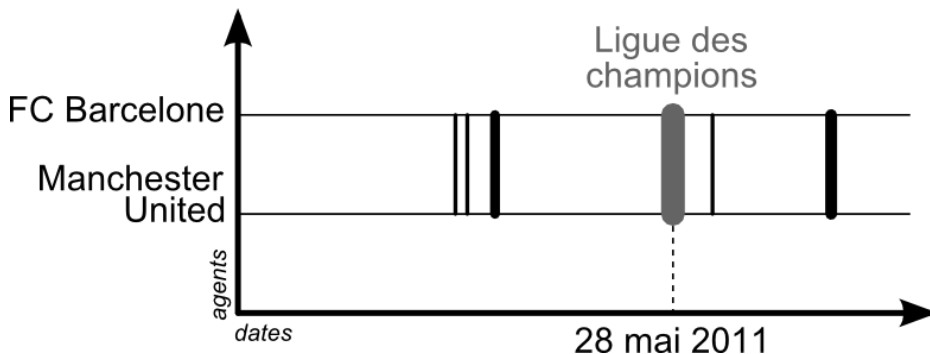
**Espace<sup>2</sup> × Temps<sup>2</sup> × Thème**

Interactions spatio-temporelles



**Espace<sup>2</sup> × Temps × Thème**

Interactions spatiales



**Quelle sémantique  
donner aux structures  
multidimensionnelles ?**

# Agrégation *a priori*

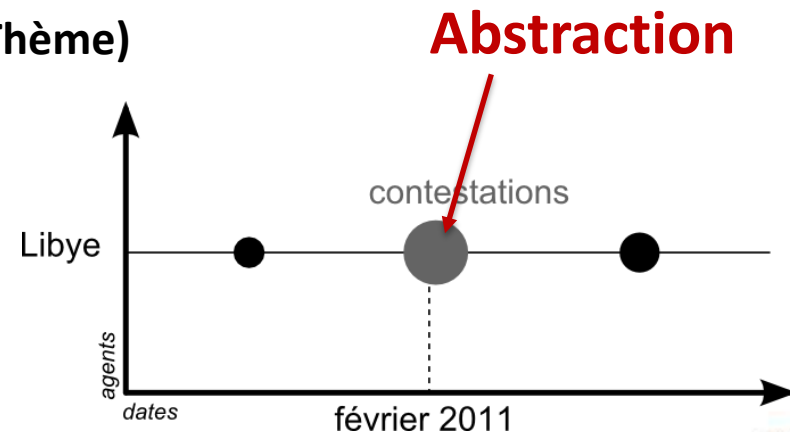
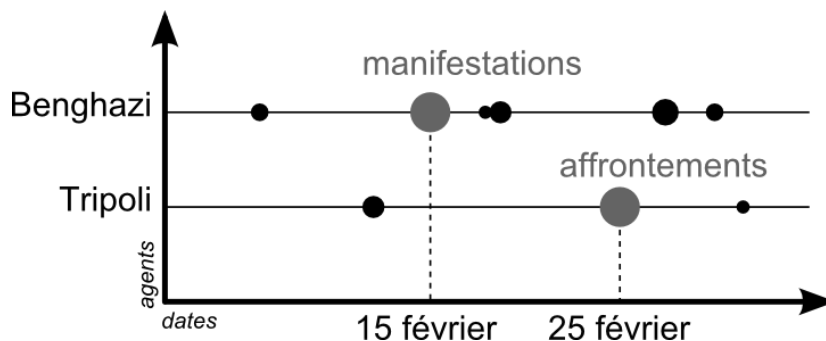
**Micro** → **Macro**

**Espace**      Communes → Pays

**Temps**        Jours → Mois

**Thèmes**      Mots-clés → Catégories

**Faits médiatiques** (Espace × Temps × Thème)



# Agrégation *a posteriori*

Espace × Espace								
	USA	France	Libye	Israël	Syrie	Palestine	Afghanistan	...
USA	×	25	19	18	24	24	24	...
France	25	×	36	10	15	15	30	...
Libye	19	36	×	0	7	0	2	...
Israël	18	10	0	×	4	62	0	...
Syrie	24	15	7	4	×	1	0	...
Palestine	24	15	0	62	1	×	0	...
Afghanistan	24	30	2	0	0	0	×	...
...	...	...	...	...	...	...	...	...
Total	423	248	308	153	260	126	131	...

- Principe d'agrégation

$$D_{\alpha} = (1 - \alpha) \times L - \alpha \times G$$

- Une valeur de  $\alpha$  pour chaque dimension...

# Agrégation *a posteriori*

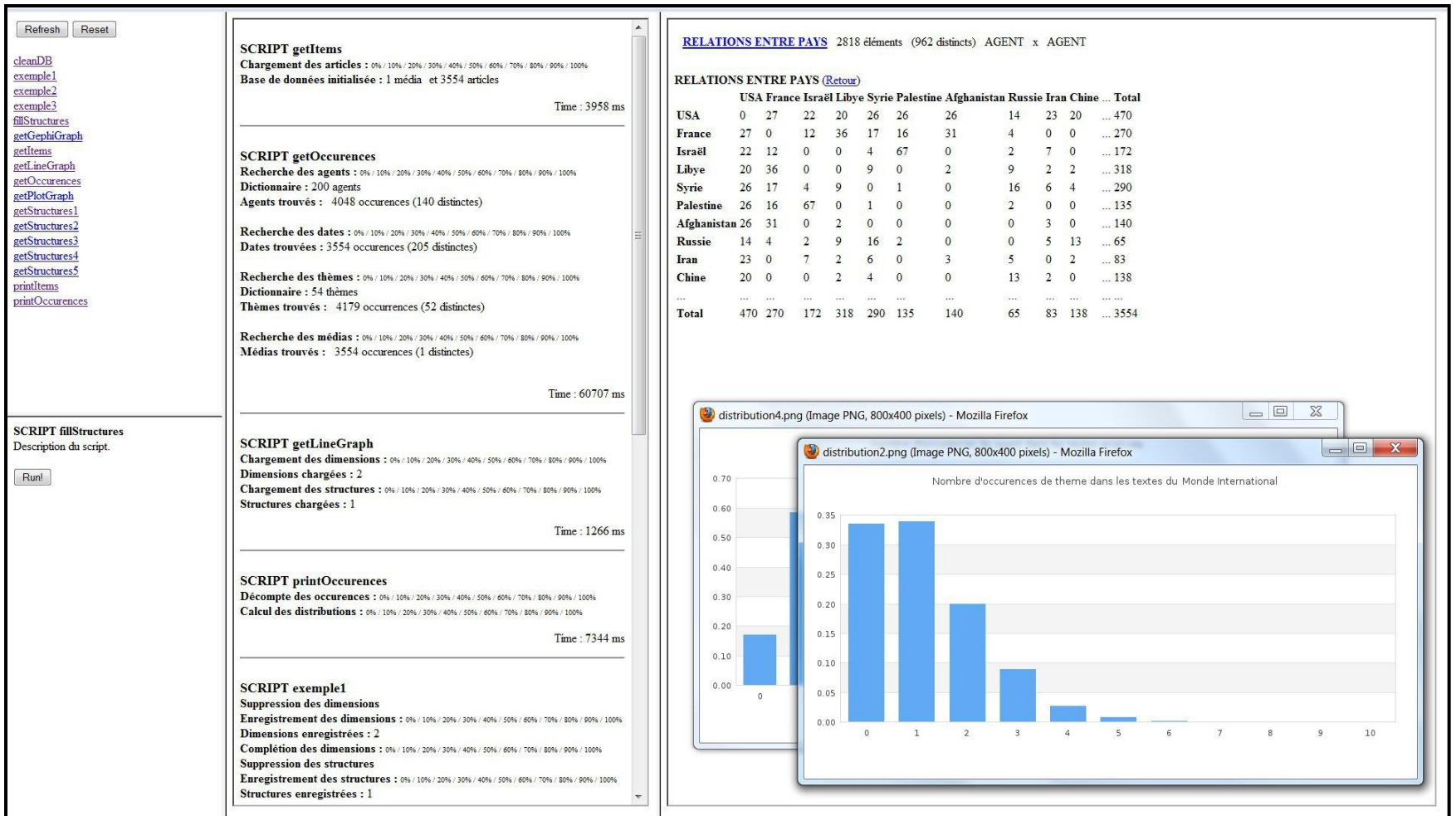
Espace × Espace								
	USA	France	Libye	Israël	Syrie	Palestine	Afghanistan	...
USA	×	25	19	18	24	24	24	...
France	25	×	36	10	15	15	30	...
Libye	19	36	×	0	7	0	2	...
Israël	18	10	0	×	4	62	0	...
Syrie	24	15	7	4	×	1	0	...
Palestine	24	15	0	62	1	×	0	...
Afghanistan	24	30	2	0	0	0	×	...
...	...	...	...	...	...	...	...	...
Total	423	248	308	153	260	126	131	...

- Principe d'agrégation

$$D_{\alpha} = (1 - \alpha) \times L - \alpha \times G$$

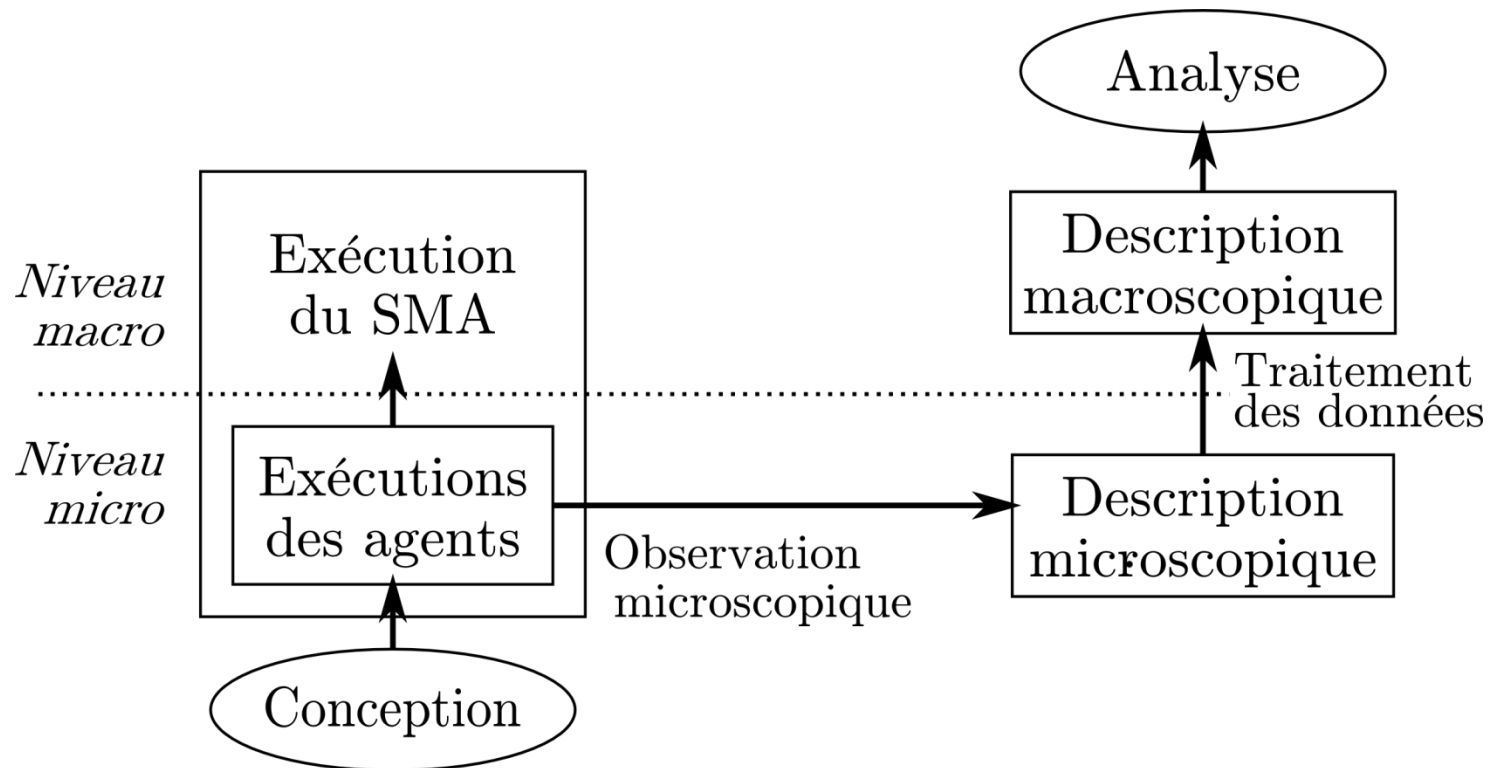
- Une valeur de  $\alpha$  pour chaque dimension...

- Agréger les entités qui ont « le même comportement médiatique », *i.e.* qui sont **fortement corrélées**

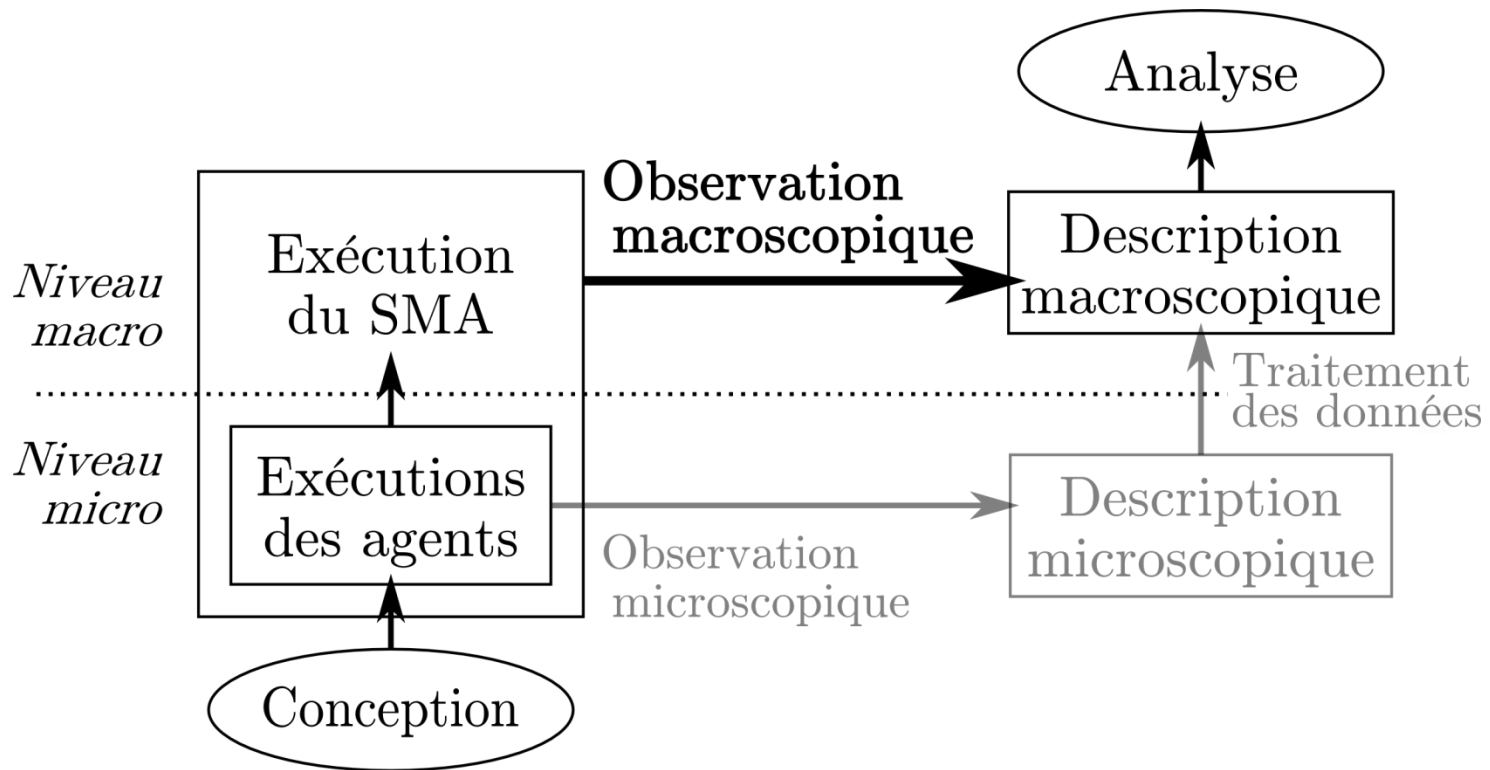


# **CALCUL DES AGRÉGATIONS**

# Observation macroscopique

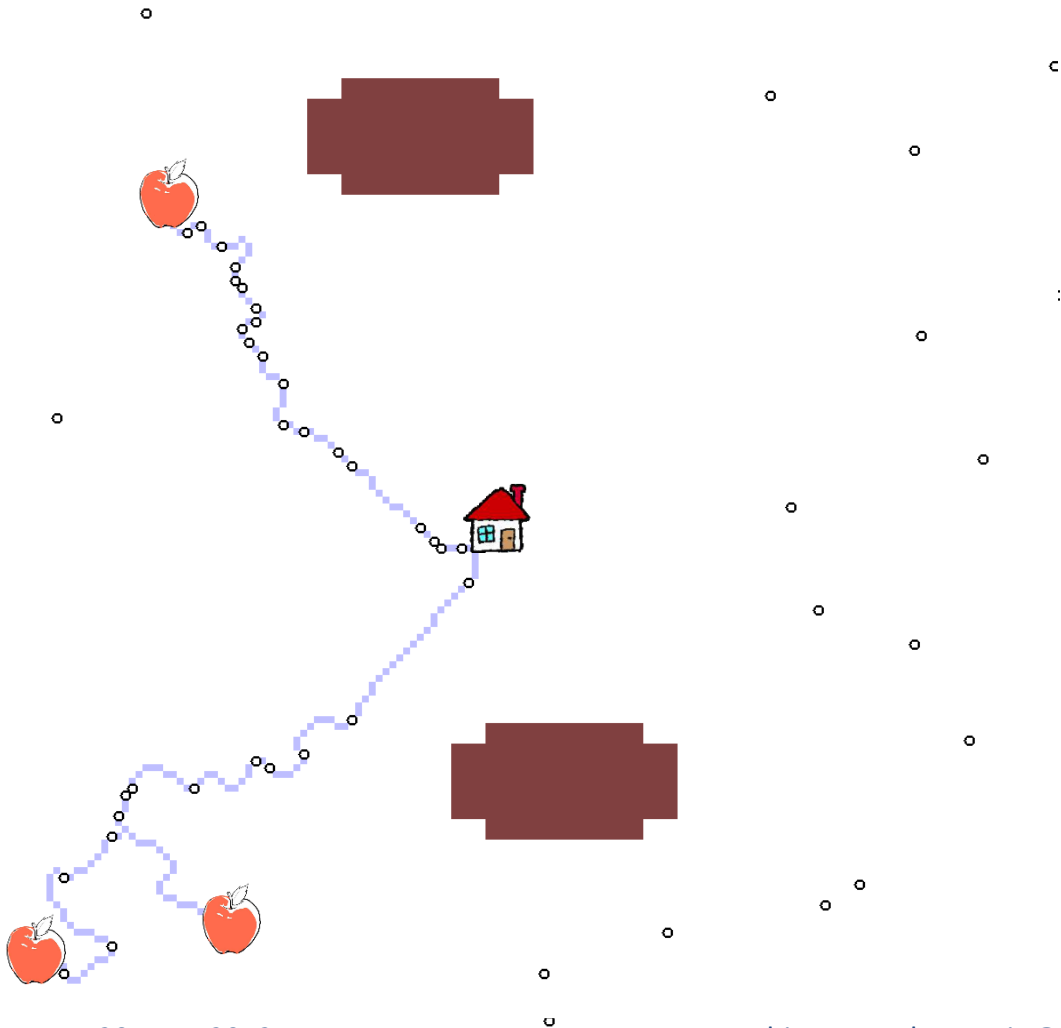


# Observation macroscopique

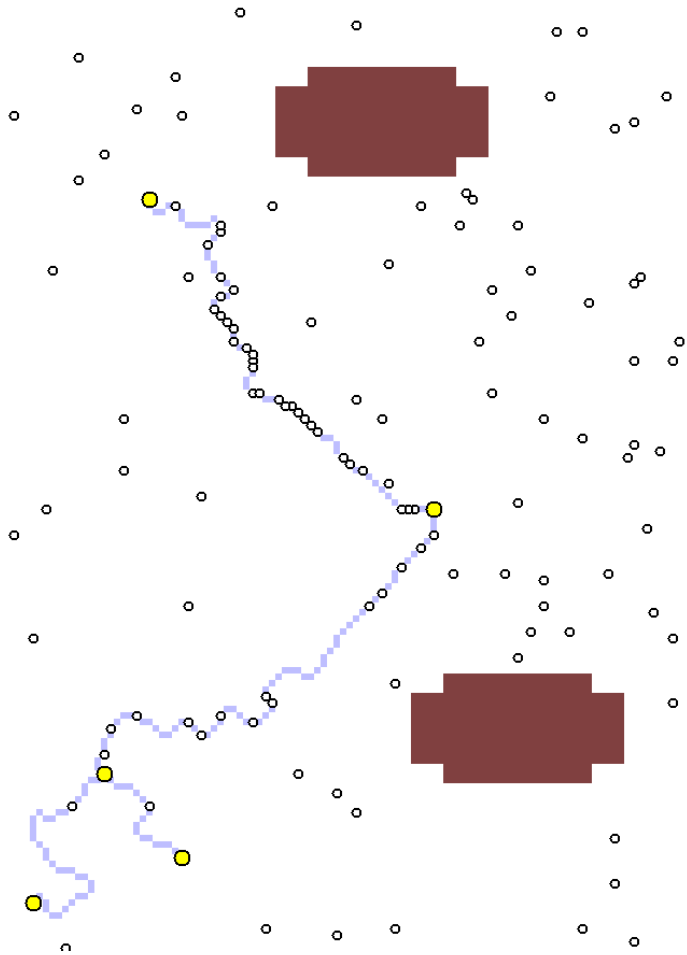




# Expérimentations

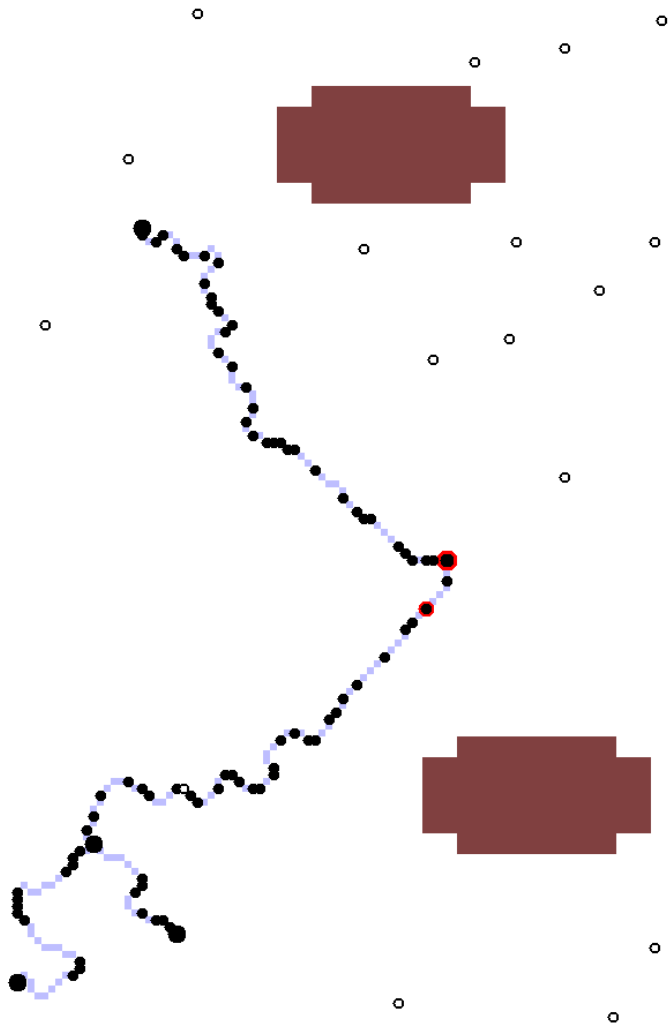


# Sondes macroscopiques et agrégation spatiale



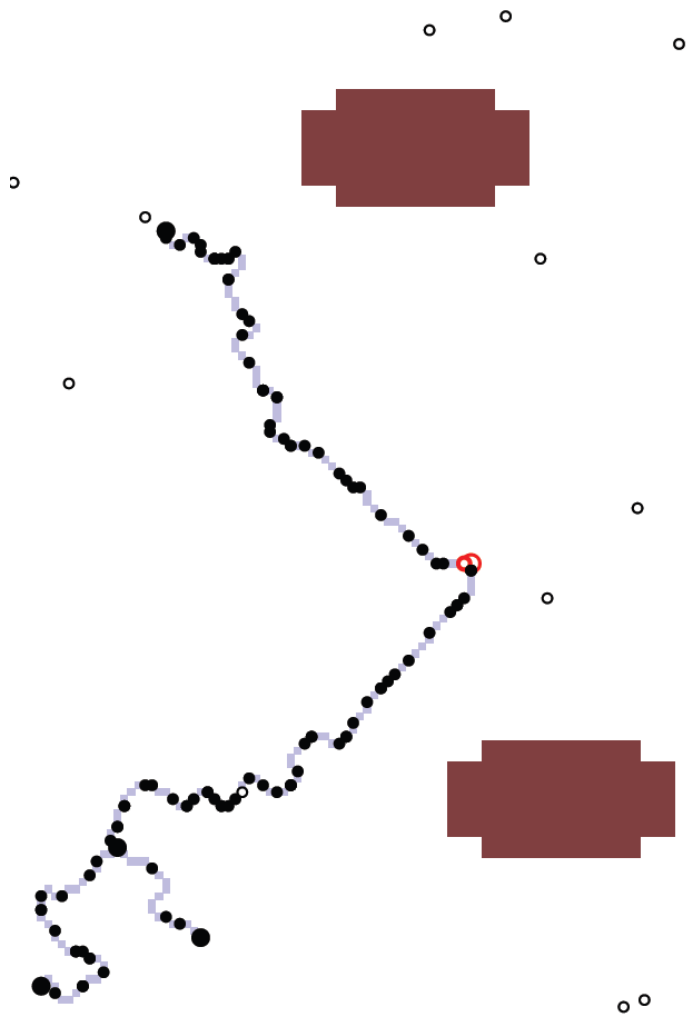
- Centralisation de l'information
- **Agents agrégés** : fourmis partant d'une sonde
- **Interactions entre sondes** : transfert d'une fourmi d'une piste à l'autre

# Temps distribué et agrégation temporelle



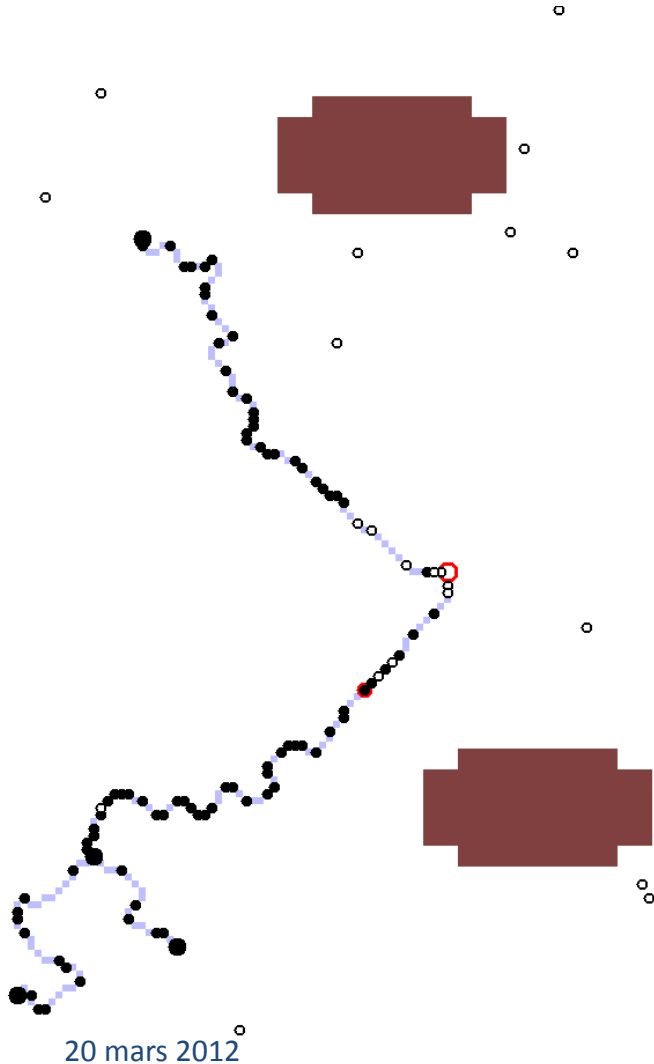
- Synchronisation de l'information  
Adapté du *Snapshot algorithm*  
[Chandy et Lamport, 1985]
- **Intervalles agrégés** : correspondant à l'aller-retour d'une fourmi
- **Interaction entre sonde** : flux d'activité pendant un aller-retour

# Temps distribué et agrégation temporelle



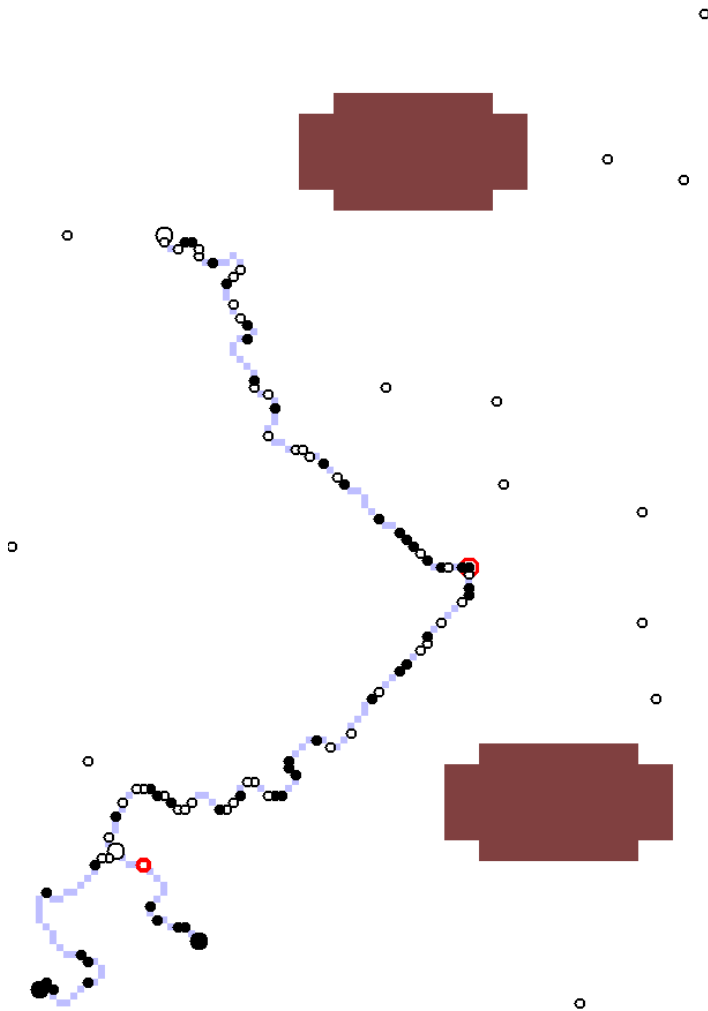
- Synchronisation de l'information  
Adapté du *Snapshot algorithm*  
[Chandy et Lamport, 1985]
- **Intervalles agrégés** : correspondant à l'aller-retour d'une fourmi
- **Interaction entre sonde** : flux d'activité pendant un aller-retour

# Temps distribué et agrégation temporelle



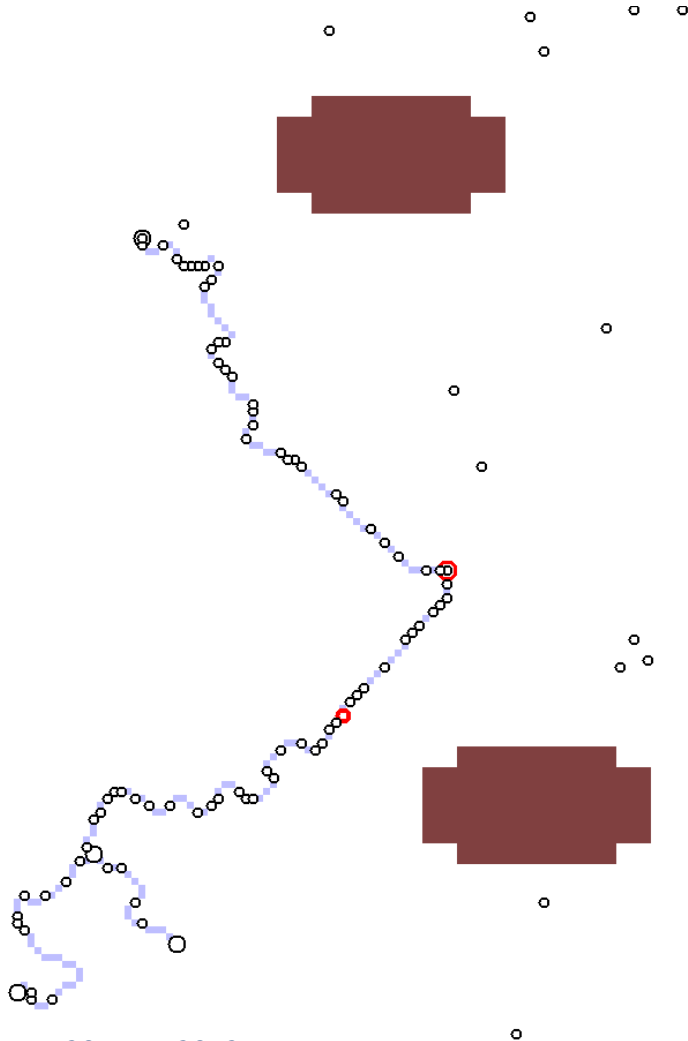
- Synchronisation de l'information  
Adapté du *Snapshot algorithm*  
[Chandy et Lamport, 1985]
- **Intervalles agrégés** : correspondant à l'aller-retour d'une fourmi
- **Interaction entre sonde** : flux d'activité pendant un aller-retour

# Temps distribué et agrégation temporelle



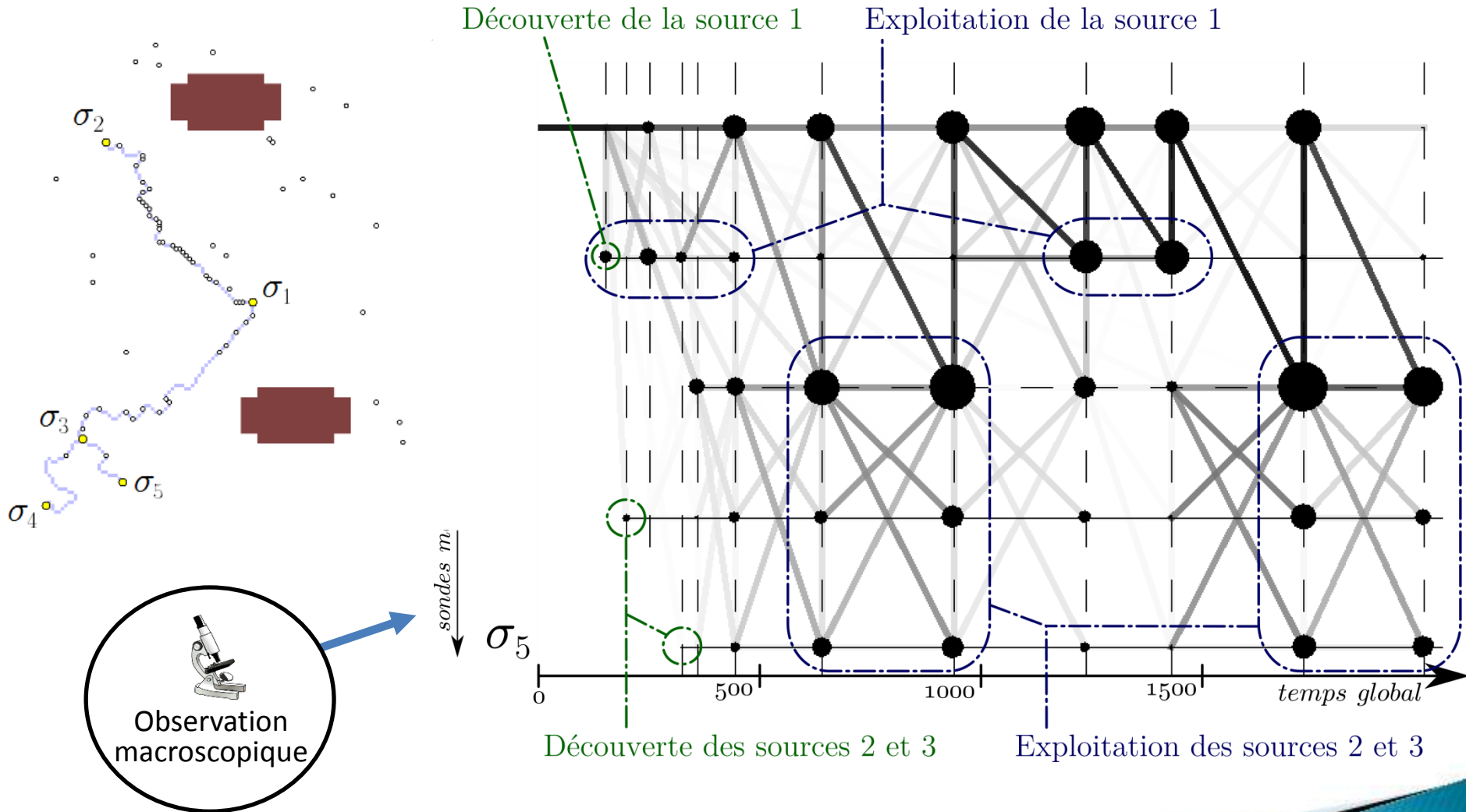
- Synchronisation de l'information  
Adapté du *Snapshot algorithm*  
[Chandy et Lamport, 1985]
- **Intervalles agrégés** : correspondant à l'aller-retour d'une fourmi
- **Interaction entre sonde** : flux d'activité pendant un aller-retour

# Temps distribué et agrégation temporelle



- Synchronisation de l'information  
Adapté du *Snapshot algorithm*  
[Chandy et Lamport, 1985]
- **Intervalles agrégés** : correspondant à l'aller-retour d'une fourmi
- **Interaction entre sonde** : flux d'activité pendant un aller-retour

# Description macroscopique





# **BILAN DES PROBLÈMES POSÉS**

# Problèmes posés

---

- Qu'est-ce qu'une abstraction ?
- Comment mesurer/évaluer une description ?
- Comment agréger des relations causales ?
- Comment agréger l'information médiatique ?
- Comment calculer efficacement les agrégations ?