

**Retraite MAGMA**

**Grenoble, 27 juin 2012**

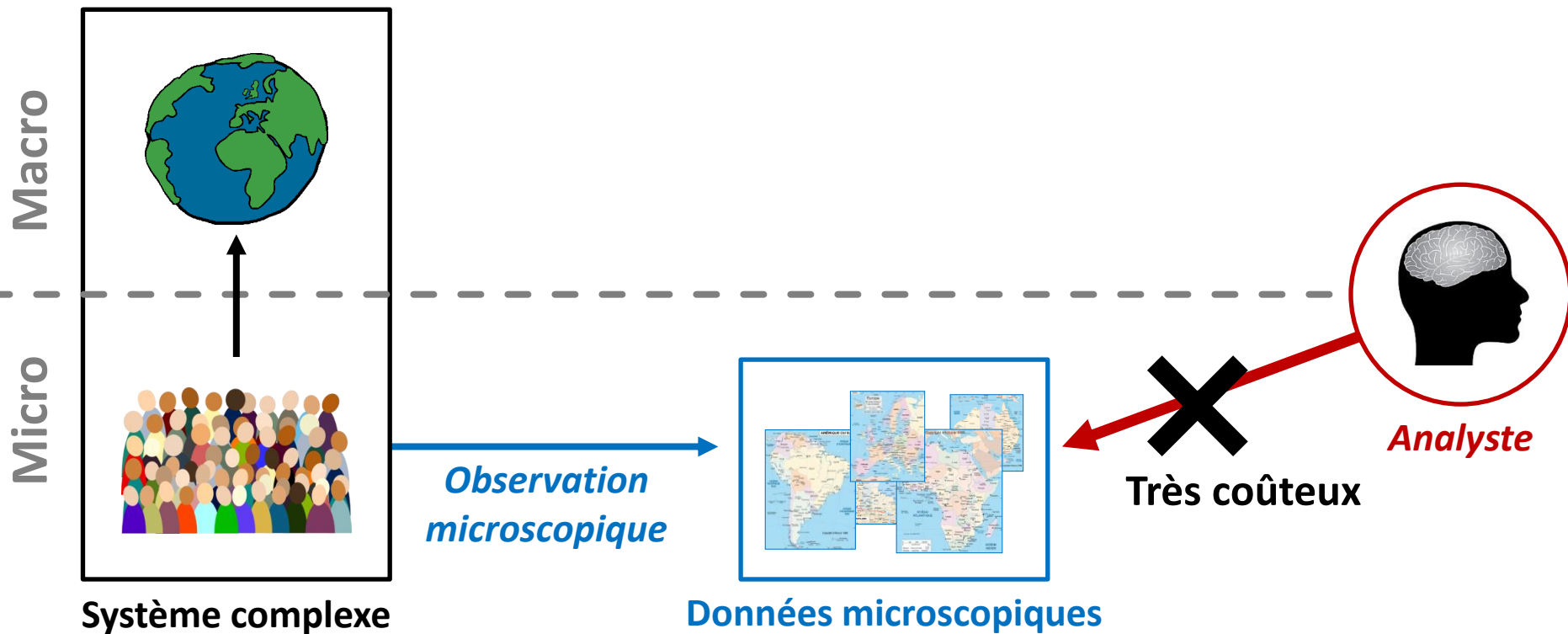
# **Agrégation de données pour l'analyse de systèmes complexes**

**Robin Lamarche-Perrin**

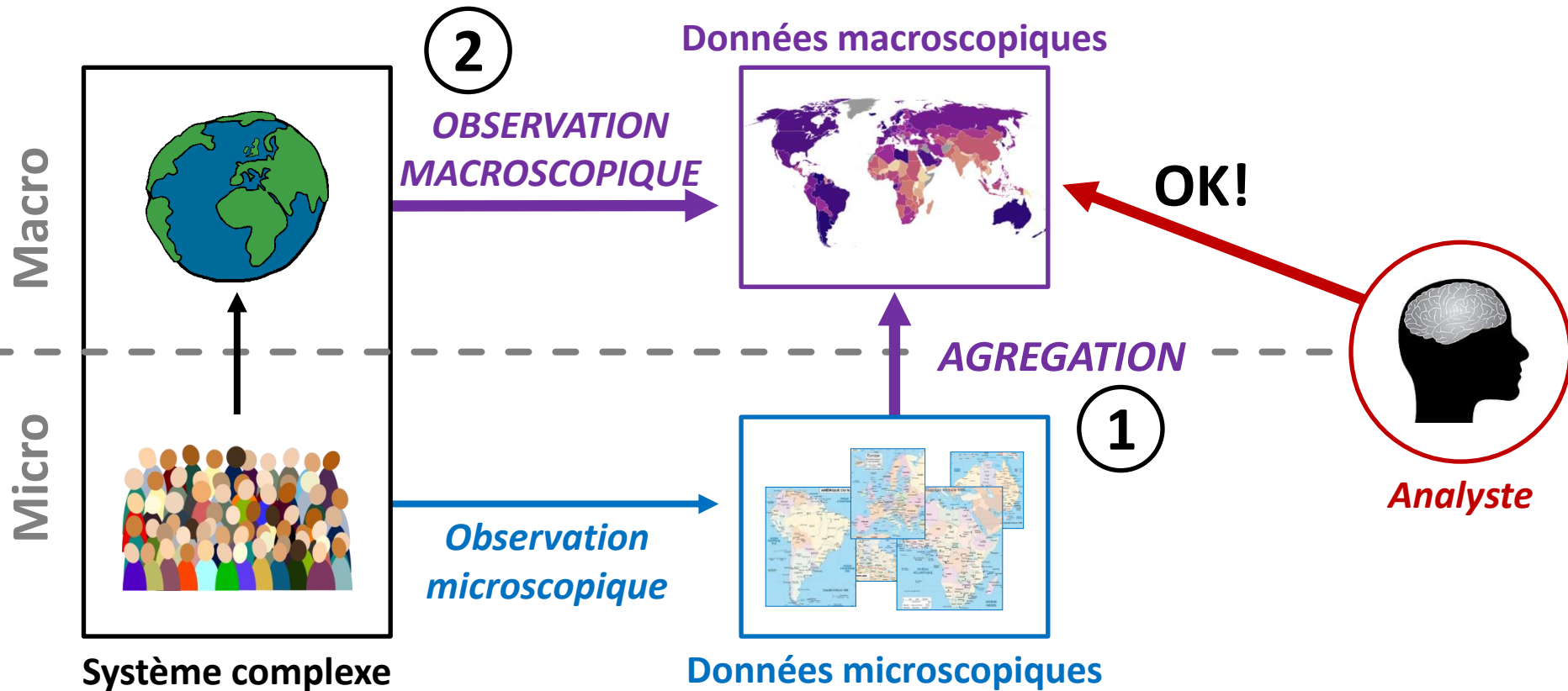
Yves Demazeau (MAGMA)

Jean-Marc Vincent (MESCAL)

# Changer le niveau de l'analyse



# Changer le niveau de l'analyse



**QU'EST-CE QU'UNE AGRÉGATION ?**

# Description unidimensionnelle

*Exécution  
d'un système*

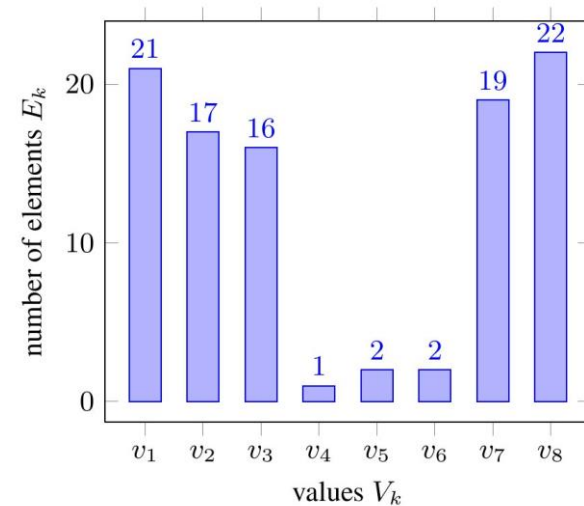


$E = 100$  entités



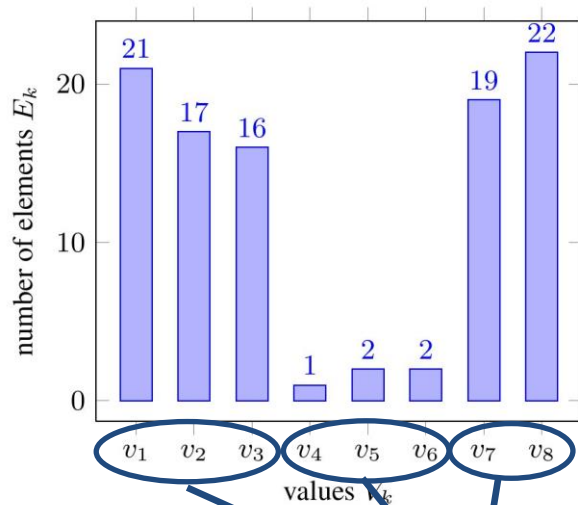
$V = 8$  valeurs

## Description

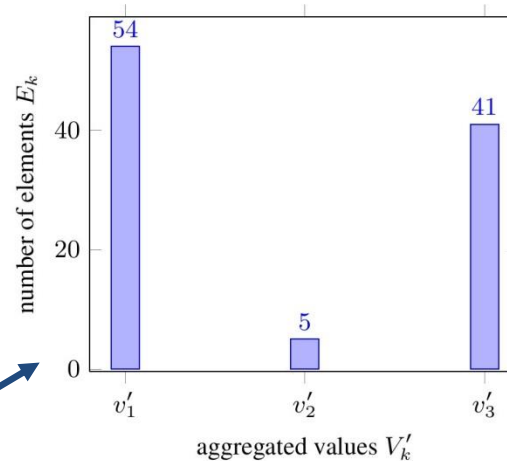


# Agrégation de descriptions

## Description source

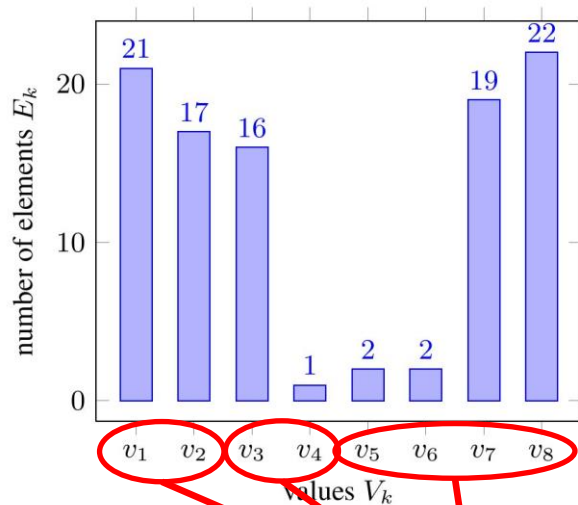


## Descriptions agrégées

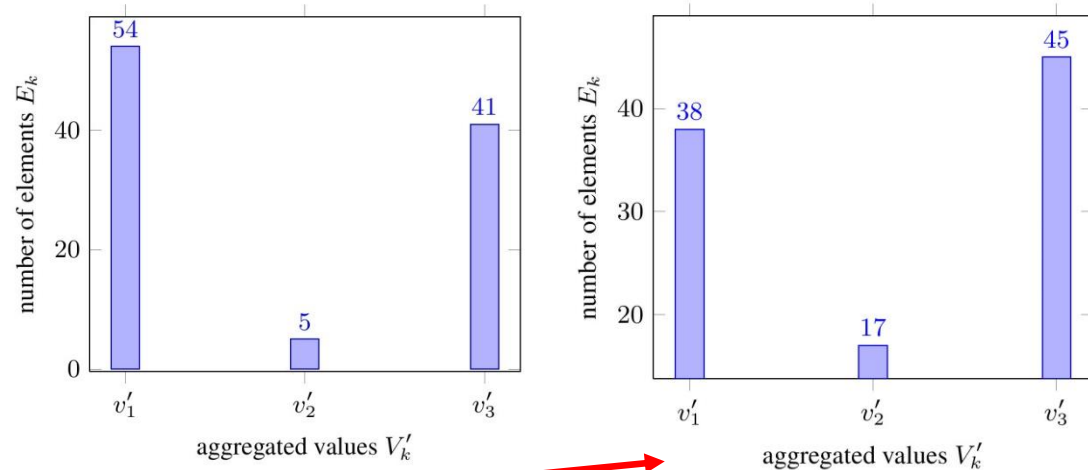


# Agrégation de descriptions

## Description source



## Descriptions agrégées

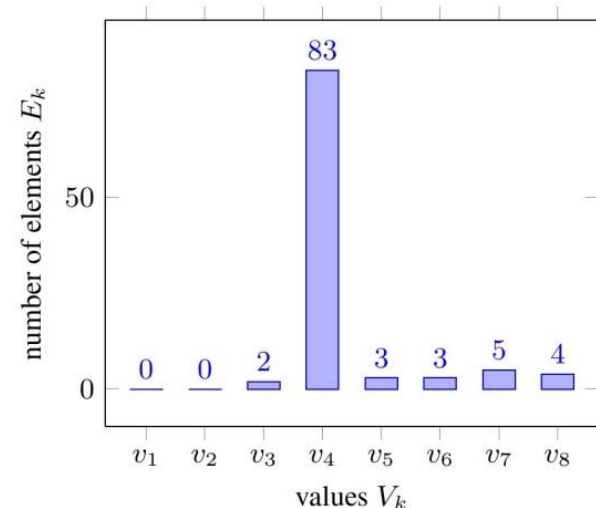
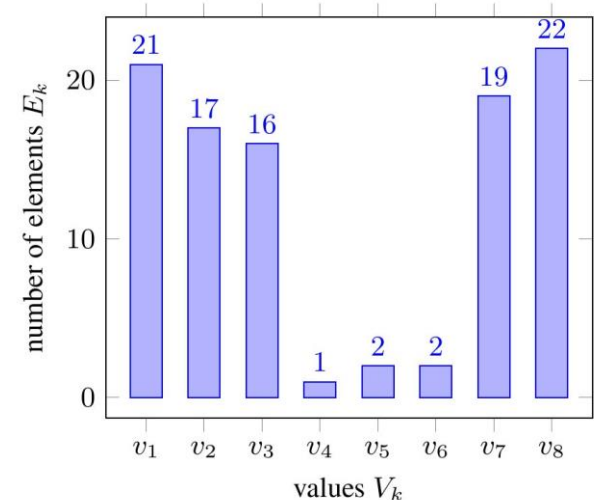


# **COMMENT ÉVALUER ET COMPARER LES AGRÉGATIONS ?**



# Mesures de complexité

- Mesures simples
  - Taille du système  $E = 100$
  - Variété  $V = 8$
- Théorie de l'information
  - Complexité de Kolmogorov
  - Profondeur logique de Bennett
  - Entropie de Shannon

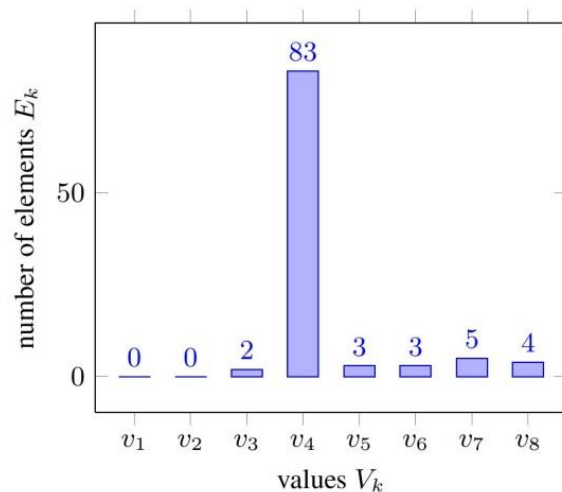


# Entropie (description)

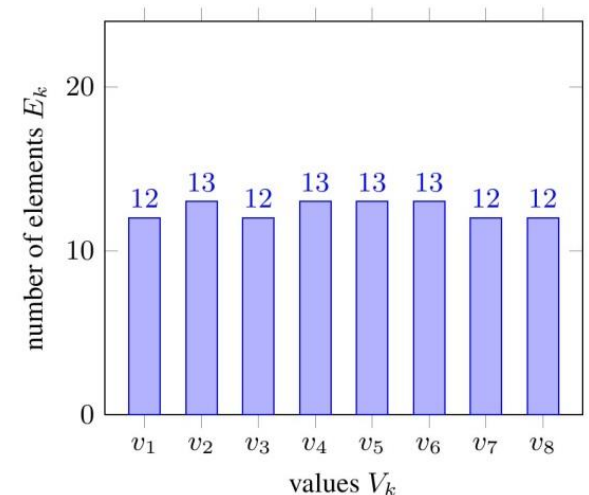
- Mesure la quantité d'information nécessaire pour coder une description donnée
- Mesure l'ordre de répartition des éléments

## Formule de Shannon

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



**Entropie minimale**

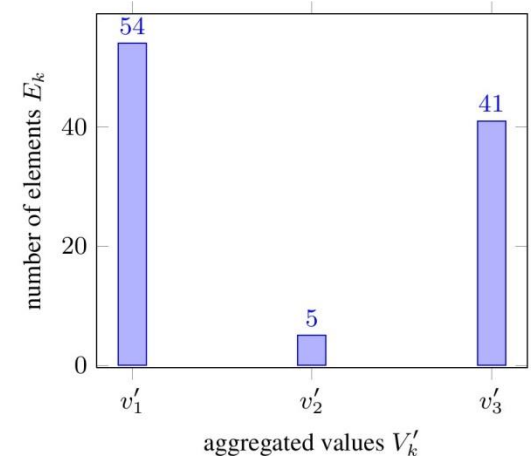
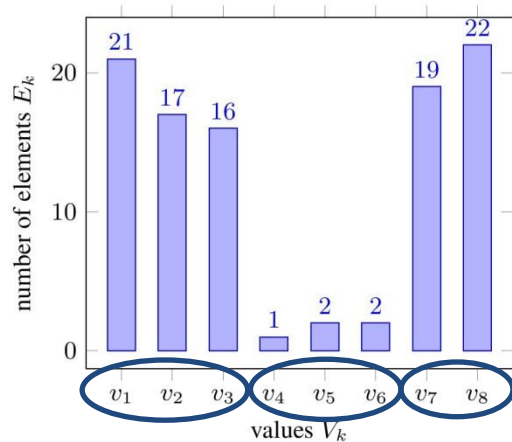


**Entropie maximale**

# Gain d'entropie (agrégation)

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

$$G = H_{source} - H_{agrégée}$$

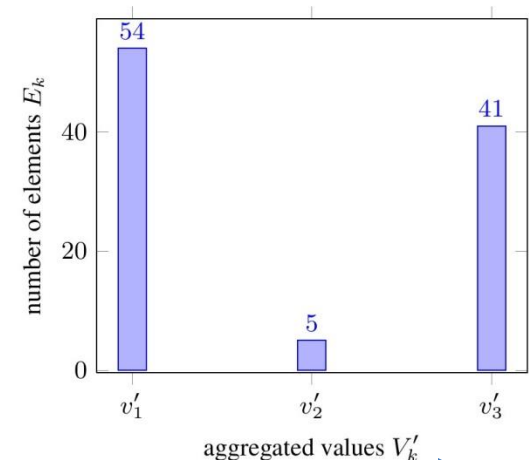
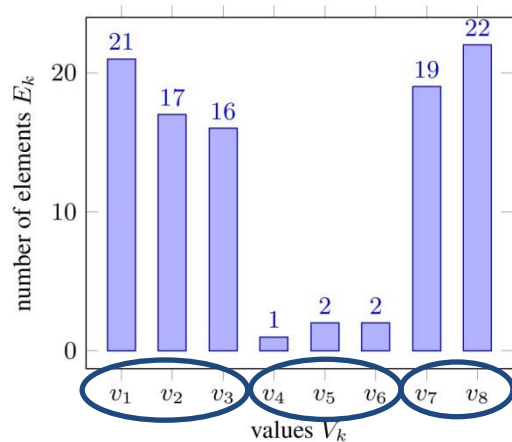


**Agrégation → diminution de l'entropie**

# Perte d'information (agrégation)

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

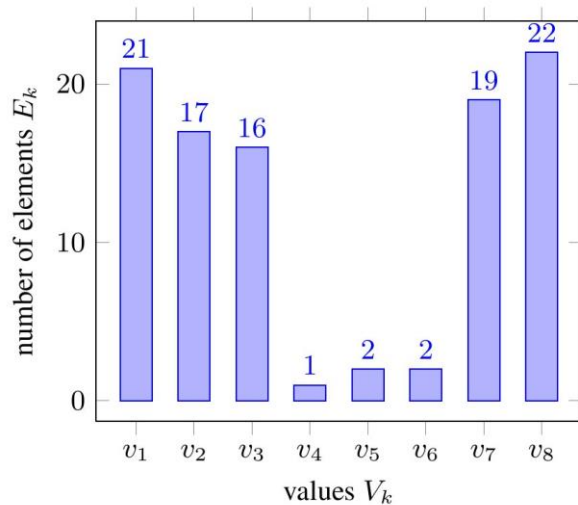
$$L = - \sum_{k \in V'} \frac{E'_k}{E} \log_2 \left( \frac{1}{V_k} \right)$$



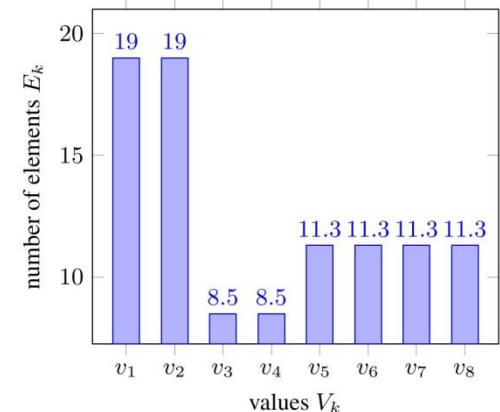
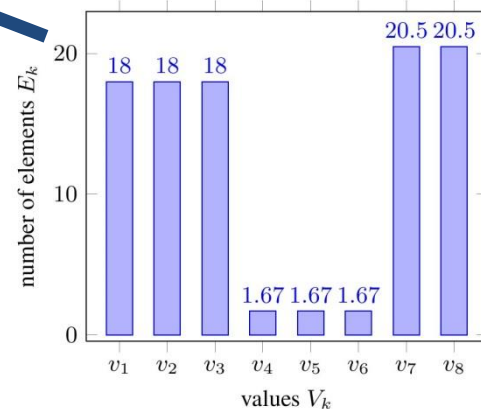
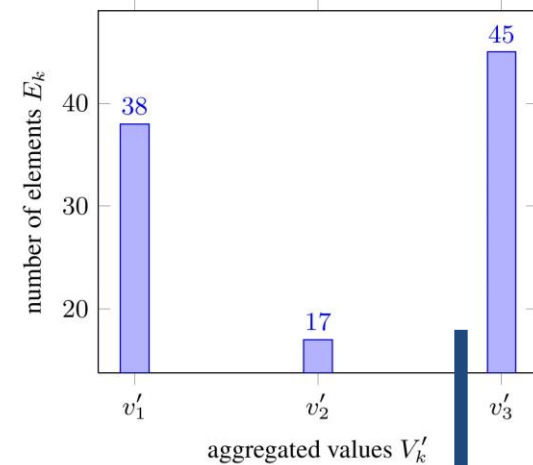
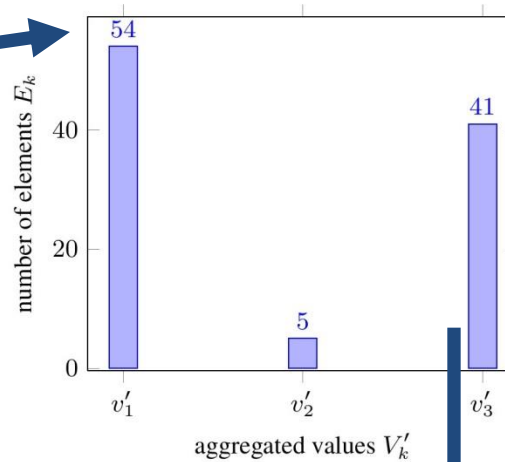
**Agrégation → perte d'information**

# Agrégation de descriptions

Description source

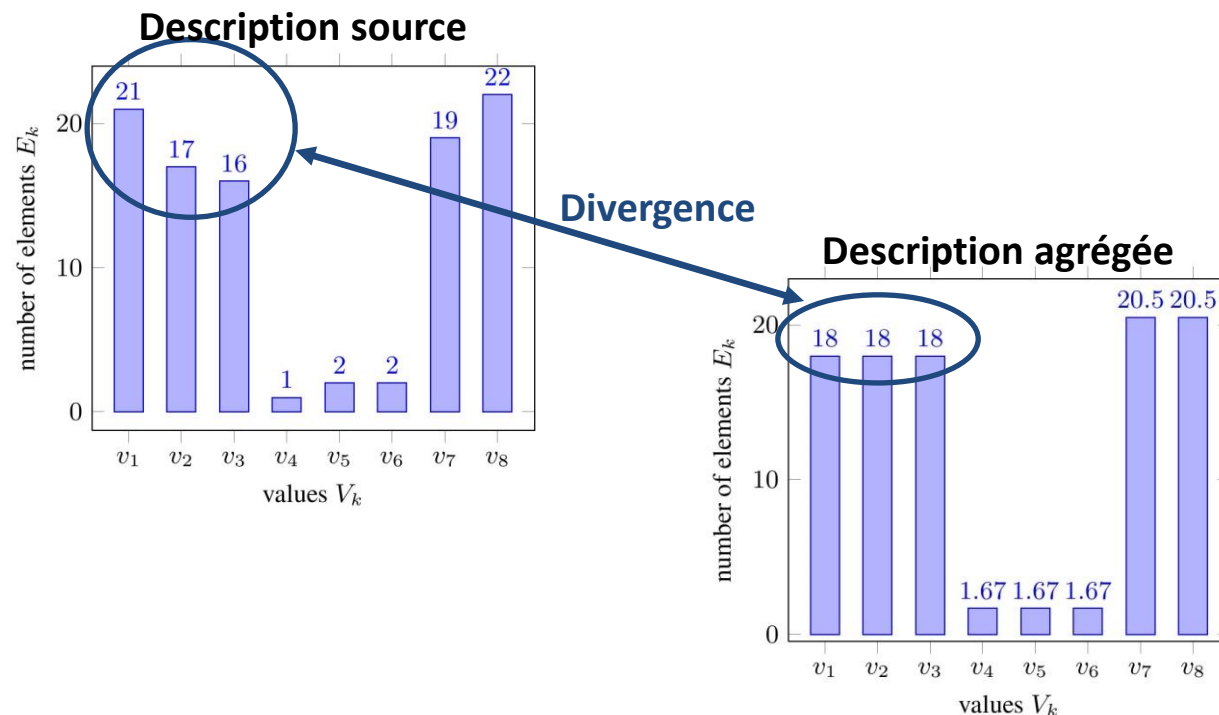


Descriptions agrégées



# Divergence (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**

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

$$D = L - G$$

# Relative Information Criterion

- *Aikaike Information Criterion* :  $AIC = V - D$
- *Our Information Criterion* :  $RIC = G - D$

- On représente un compromis :

Gain d'entropie vs. Divergence

$$pRIC = p \times G - (1 - p) \times D$$

Quelle agrégation  
maximise le pRIC ?

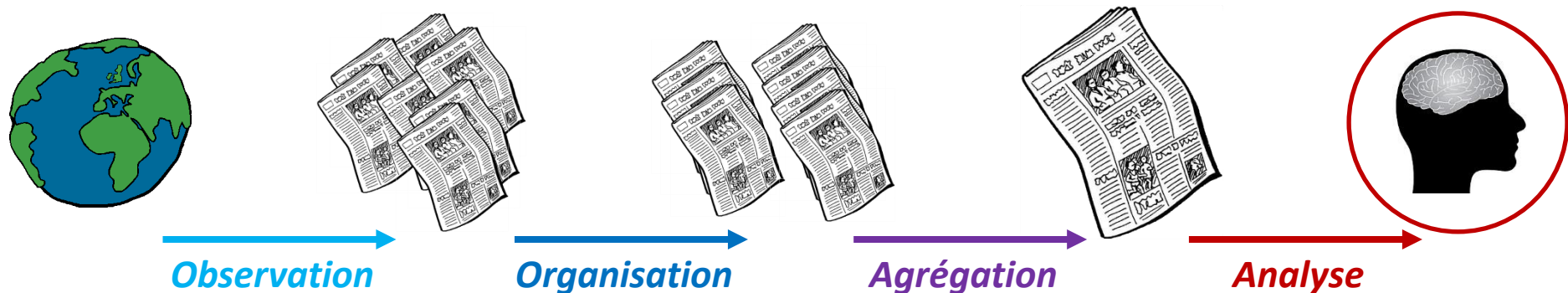
- $p = 0$   $\rightarrow$  Pas d'agrégation, information conservée
- $p = 1$   $\rightarrow$  Agrégation totale

**ANR CORPUS GEOMEDIA**  
**ANALYSE D'INFORMATIONS MÉDIATIQUES**  
**(M. Severo, T. Giraud, C. Grasland)**



# Projet GEOMEDIA

## Agrégations spatiales de données médiatiques



Quelles agrégations sont les plus pertinentes ?

	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



UNEP

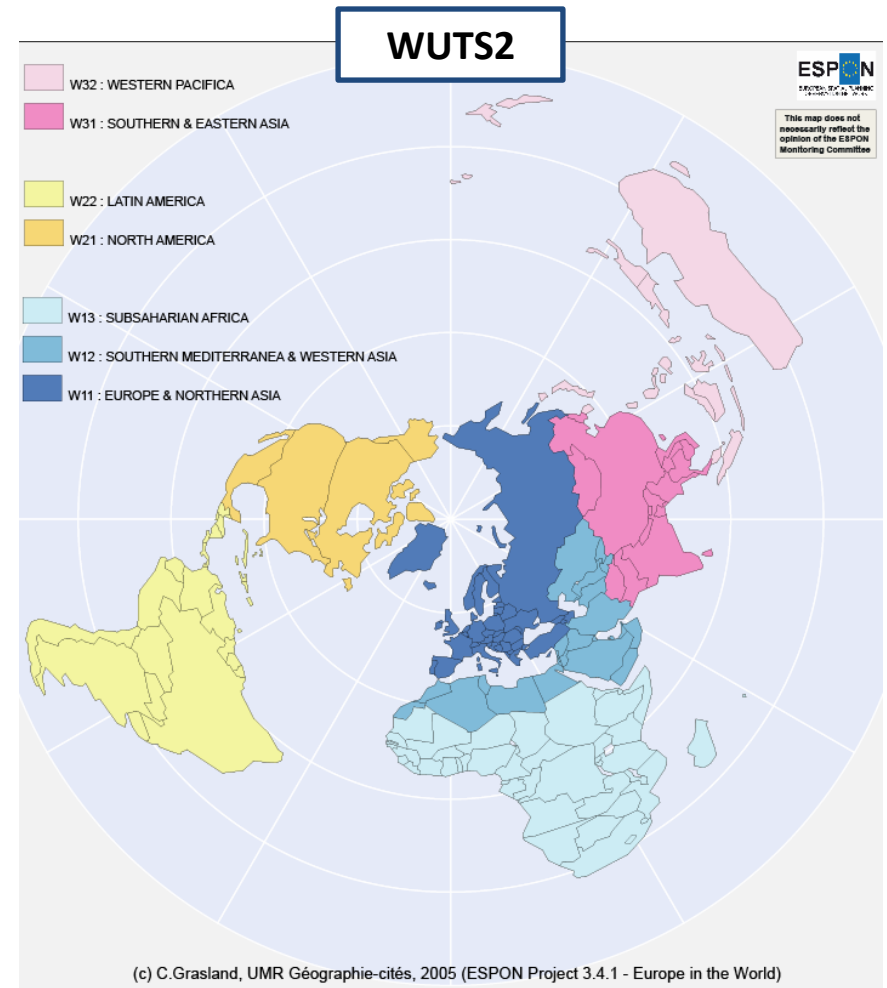


WUTS2

27 juin 2012



# Comparaison des agrégats



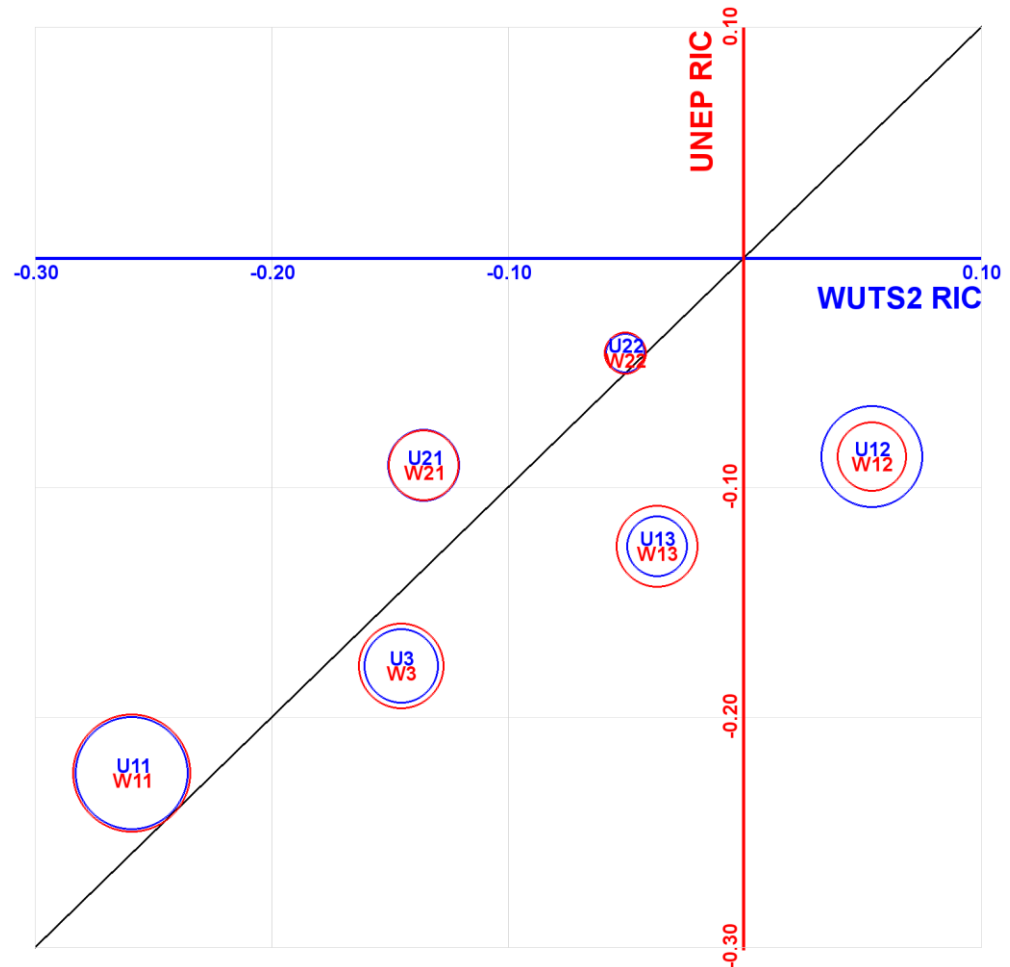
# Comparaison des agrégats



UNEP



WUTS2



Retraite MAGMA 2012

**TRIVA**

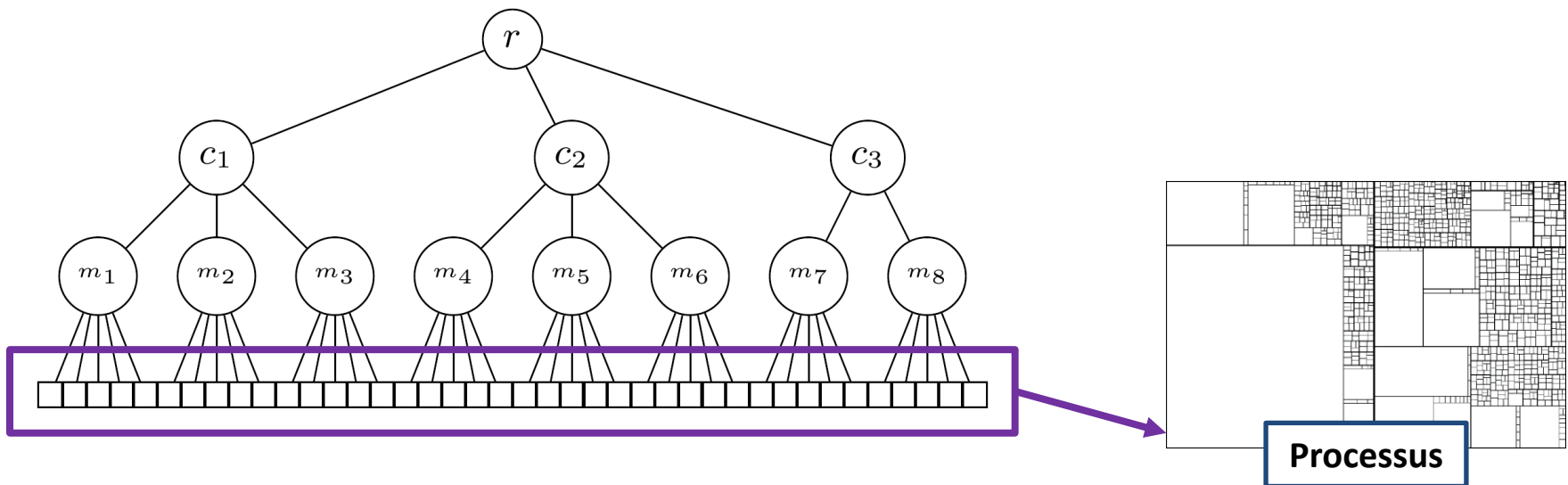
**ANALYSE DE SYSTÈMES DISTRIBUÉS  
(ANR SONGS, L. Schnorr, A. Legrand)**

# Projet TRIVA

## Agrégation et visualisation de systèmes distribués

**Quel niveau d'agrégation  
doit-on considérer ?**

**Quelle partie de la hiérarchie  
doit-on afficher ?**

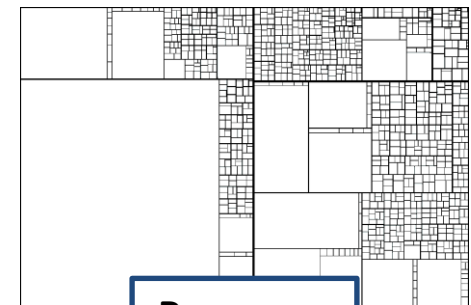
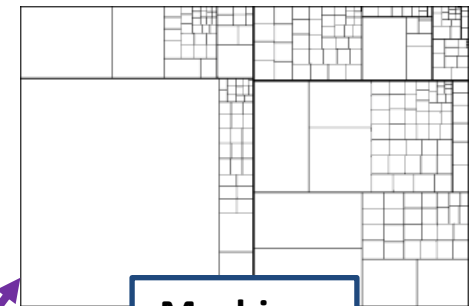
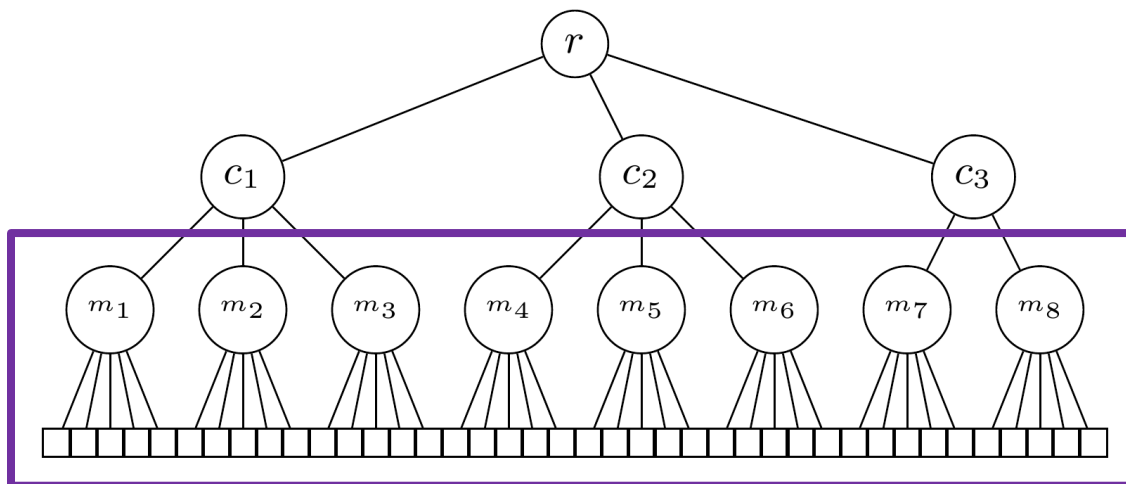


# Projet TRIVA

## Agrégation et visualisation de systèmes distribués

**Quel niveau d'agrégation  
doit-on considérer ?**

**Quelle partie de la hiérarchie  
doit-on afficher ?**

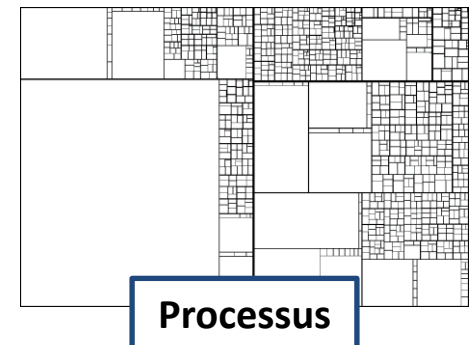
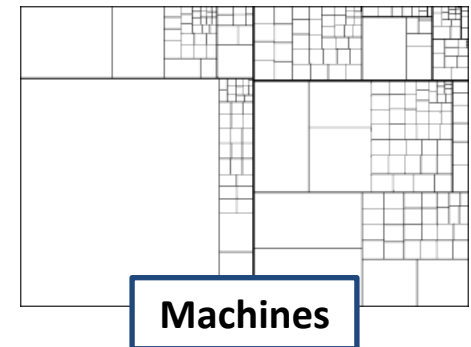
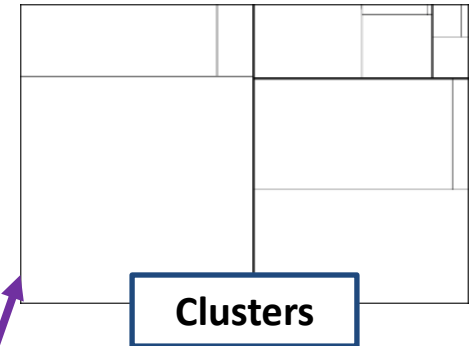
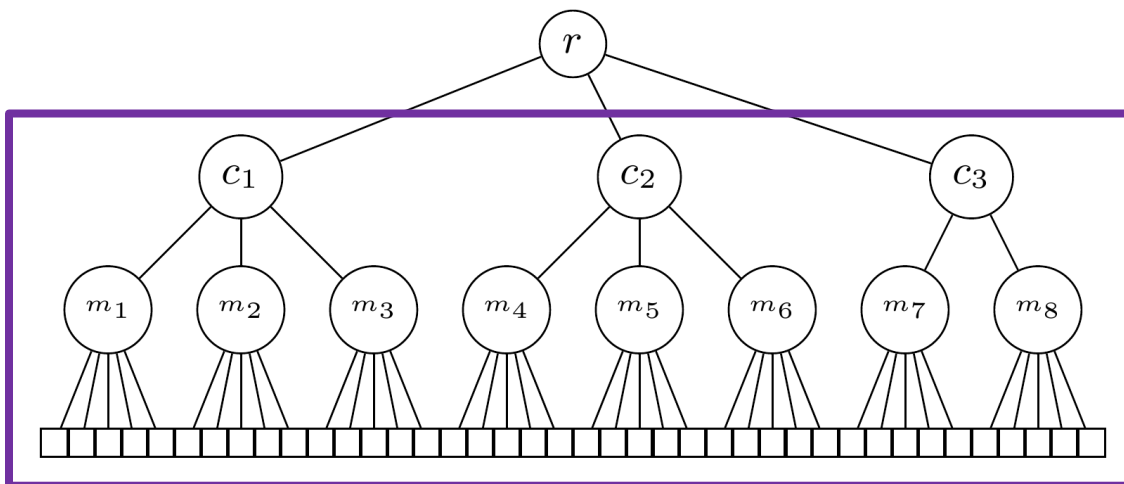


# Projet TRIVA

## Agrégation et visualisation de systèmes distribués

**Quel niveau d'agrégation  
doit-on considérer ?**

**Quelle partie de la hiérarchie  
doit-on afficher ?**



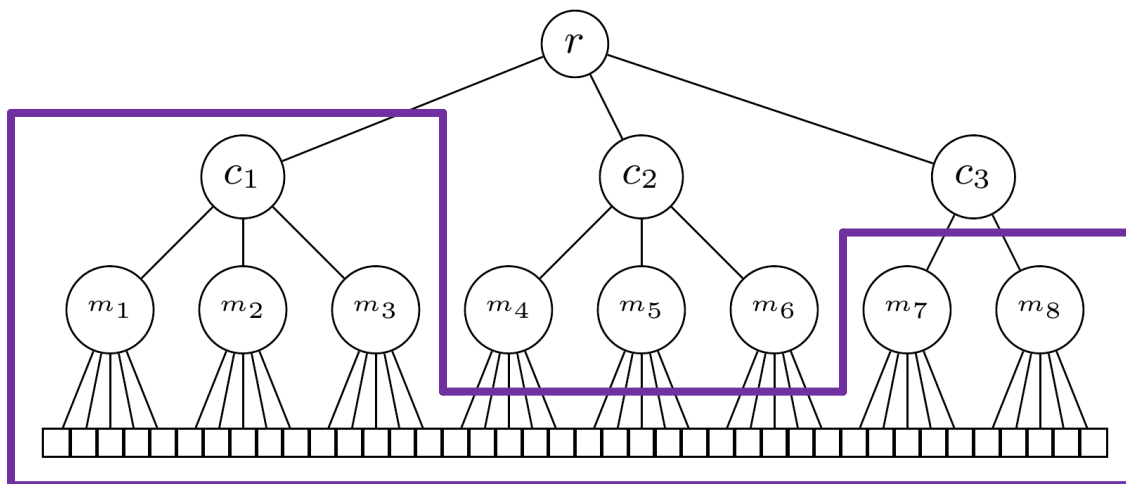


# Projet TRIVA

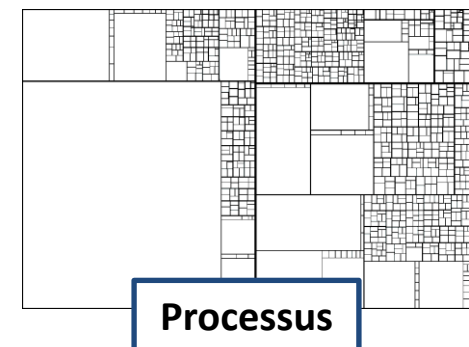
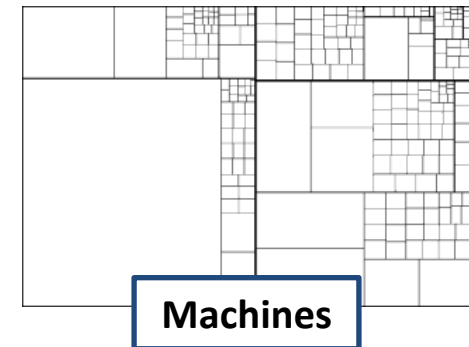
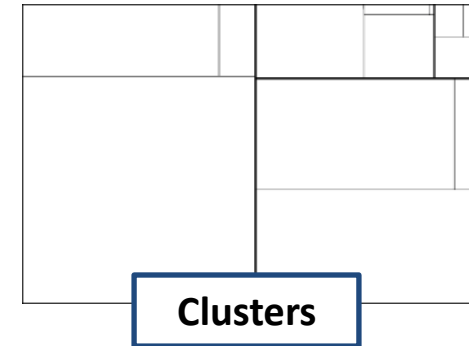
## Agrégation et visualisation de systèmes distribués

**Quel niveau d'agrégation  
doit-on considérer ?**

**Quelle partie de la hiérarchie  
doit-on afficher ?**



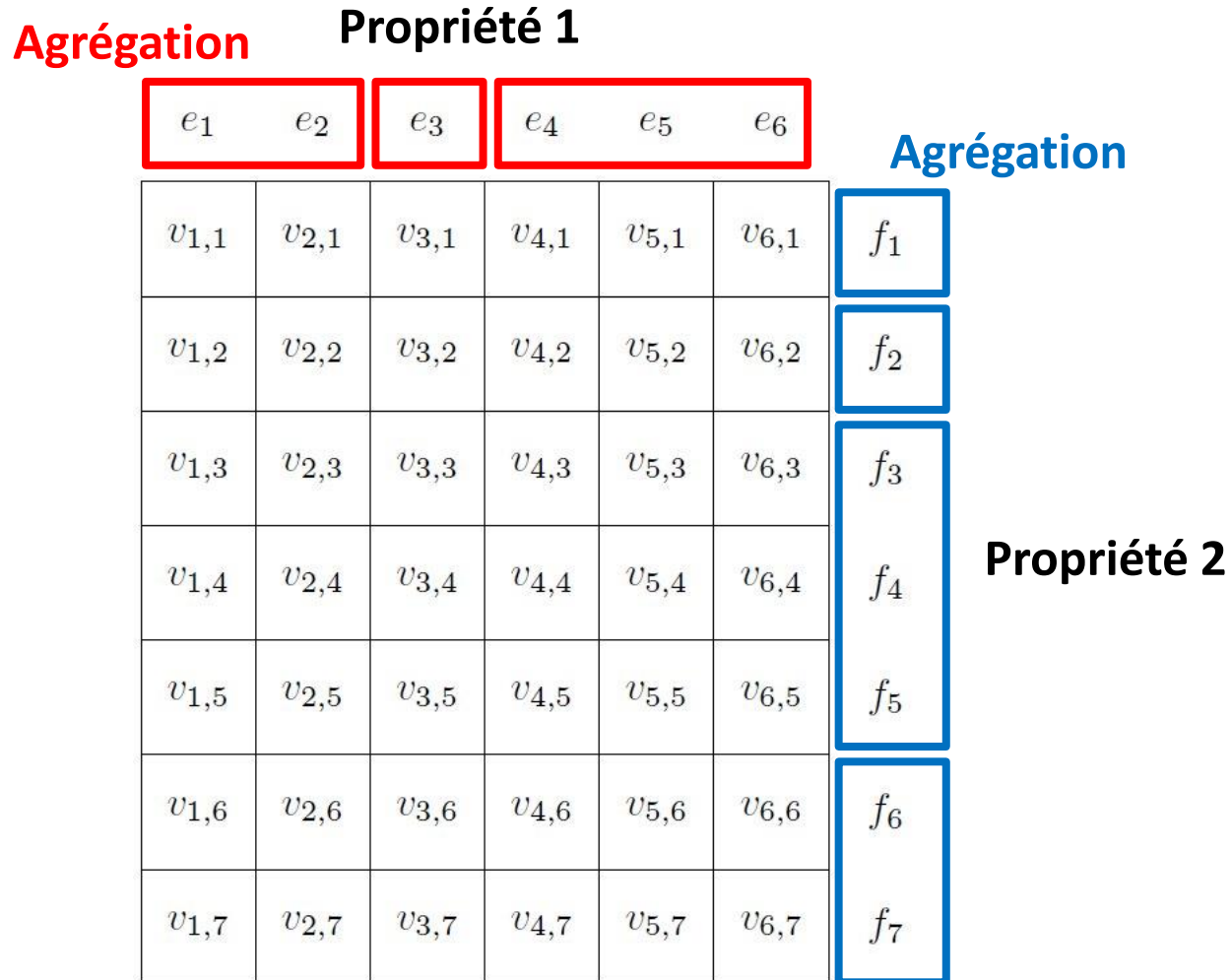
?



# Démonstration TRIVA

# Perspectives

## Agrégations multidimensionnelles



# Perspectives

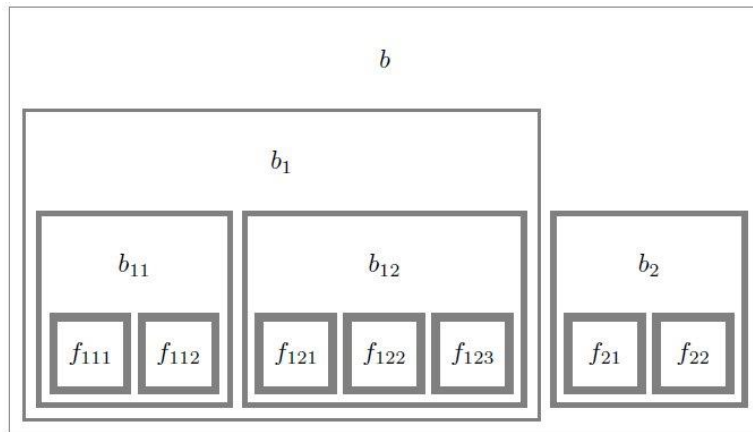
## Agrégations multidimensionnelles

Agrégation			Propriété 1			Agrégation		
$e_{11}$	$e_{12}$	$e_2$	$e_{31}$	$e_{32}$	$e_{33}$			
$v_{1,111}$	$v_{2,111}$	$v_{3,111}$				$f_{111}$		
$v_{1,112}$	$v_{2,112}$	$v_{3,112}$				$f_{112}$		
$v_{1,12}$	$v_{2,12}$	$v_{3,12}$				$f_{121}$		
						$f_{122}$		
						$f_{123}$		
$v_{1,2}$	$v_{2,2}$	$v_{3,2}$				$f_{21}$		
						$f_{22}$		

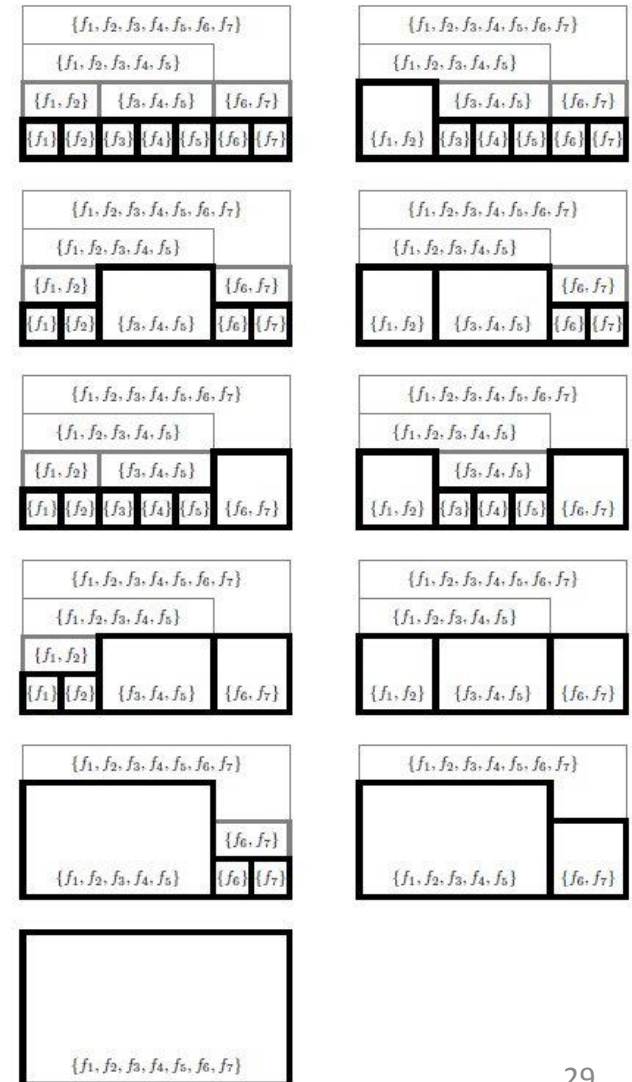
# Perspectives

## Topologies de partitionnements

### Hiérarchie



Agrégations  
possibles



D'autres topologies  
intéressantes ?

# Retraite MAGMA

Grenoble, 27 juin 2012

## Merci pour votre attention

