Information-Theoretic Measures of Aggregation for the Analysis of Complex Systems

Robin Lamarche-Perrin[1], Yves Demazeau[2], and Jean-Marc Vincent[3]

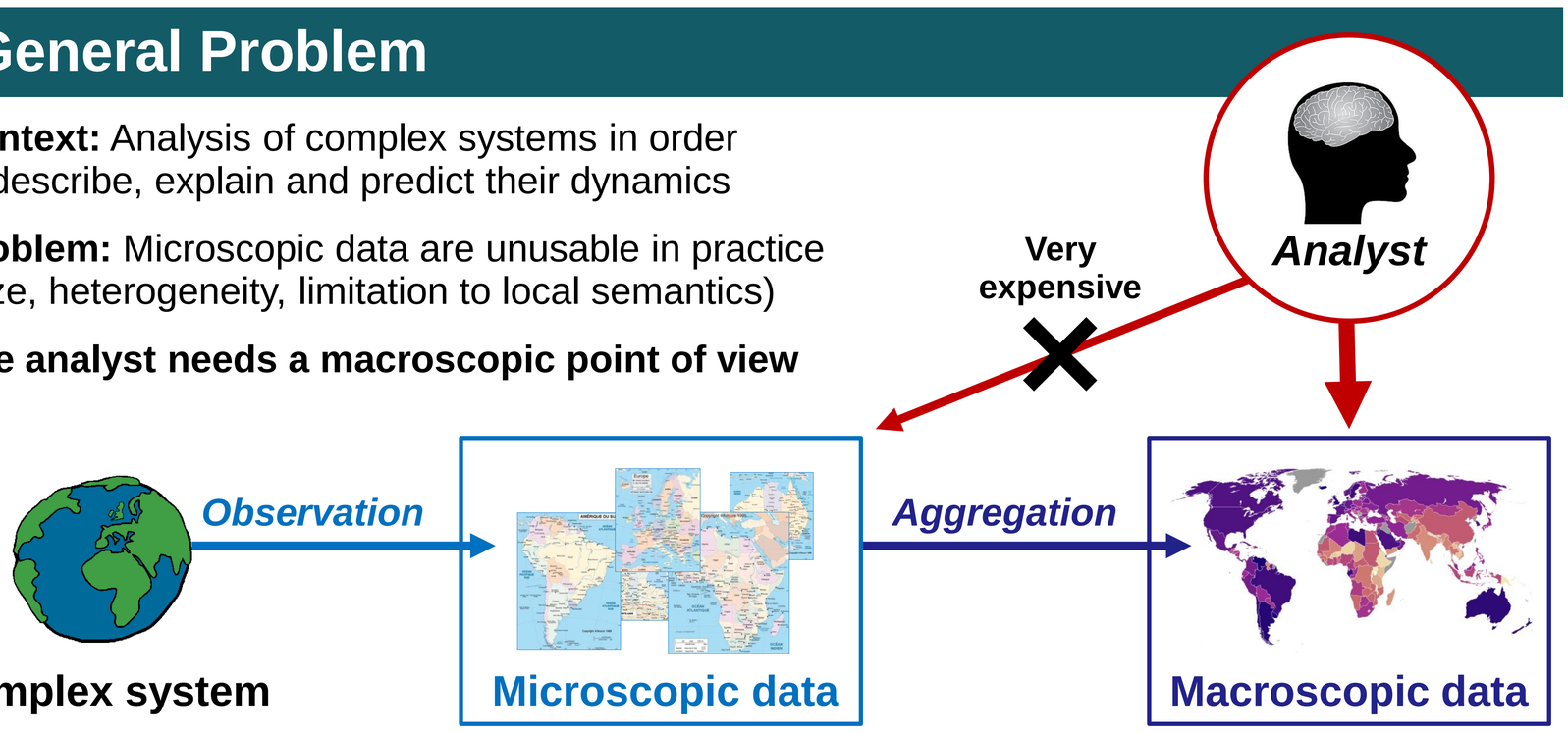
Laboratoire d'Informatique de Grenoble
[1]Université de Grenoble, [2]CNRS, [3]Université Joseph Fourier

General Problem

Context: Analysis of complex systems in order to describe, explain and predict their dynamics

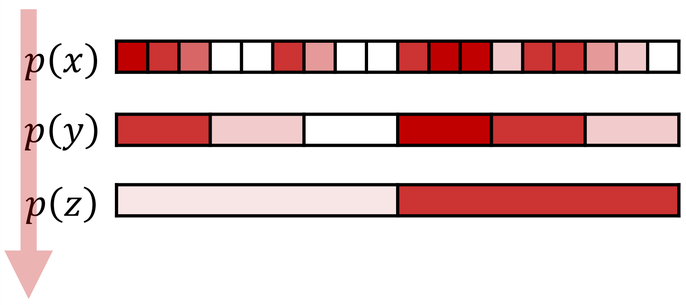
Problem: Microscopic data are unusable in practice (size, heterogeneity, limitation to local semantics)

The analyst needs a macroscopic point of view



Notion of Aggregation

Aggregation (or generalization) consists in loosing information in order to generate macroscopic points of view



How to measure and compare aggregations?

What do we Gain?

Shannon Entropy

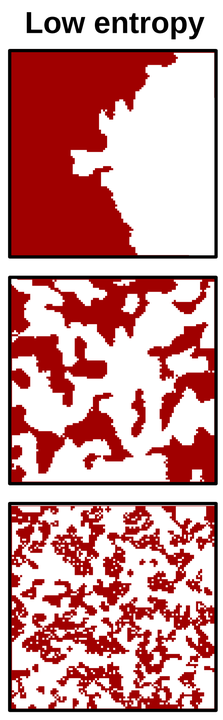
Measures the quantity of information needed to encode a set of data

$$H = -\sum_x p(x)\log_2\big(p(x)\big)$$

Entropy Gain

Measures the quantity of information reduced by the aggregation

$$G = H_{micro} - H_{macro}$$



What do we Lose?

Information Loss

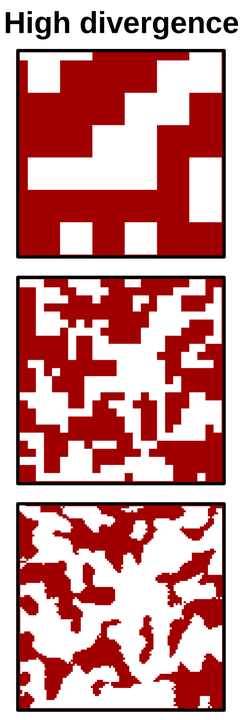
Measures the quantity of information needed to disaggregate a set of data

$$L = \sum_y p(y)\log_2|y|$$

K.-L. Divergence

Measures the quantity of information that differs between two sets of data

$$D = -\sum_x p(x)\log_2\left(\frac{p(y)}{p(x)|y|}\right)$$



Which Trade-Off?

Information Criterion

Measures a trade-off between what we gain and what we lose

$$C = G - D$$

Parametric Criterion

$$C_p = pG - (1-p)D$$

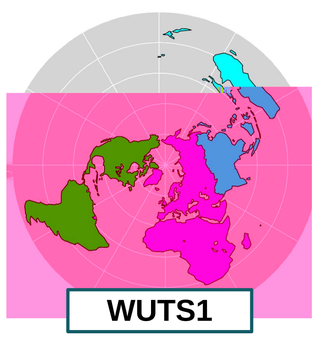
$p = 0 \rightarrow$ No aggregation at all
$p = \frac{1}{2} \rightarrow$ Aggregation iff $G > D$
$p = 1 \rightarrow$ Maximal aggregation

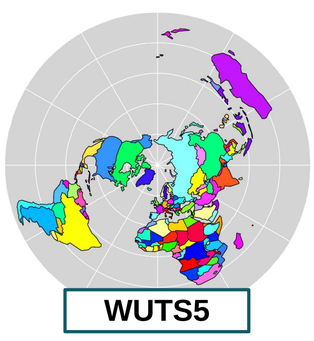
Territorial Aggregations

The GEOMEDIA Project (in collaboration with the CIST, Paris)

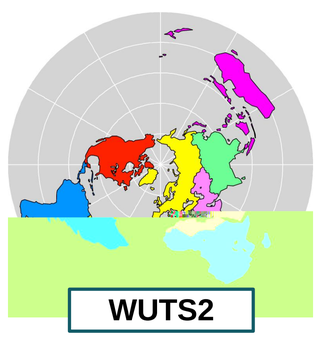
Project: building a platform for the global analysis of media information
Data: newspaper articles, demographic data, economic data
Aggregations: by states, by dates, by actors, by topics, *etc.*

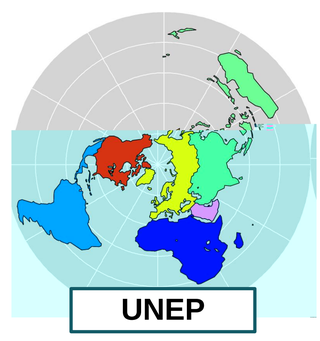
What is the best geographical level for a given analysis?





Which aggregates are the most meaningful for a given analysis?



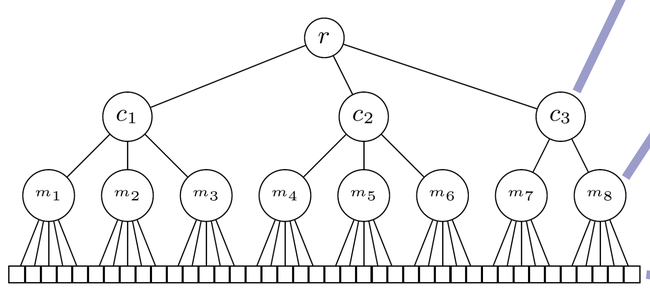


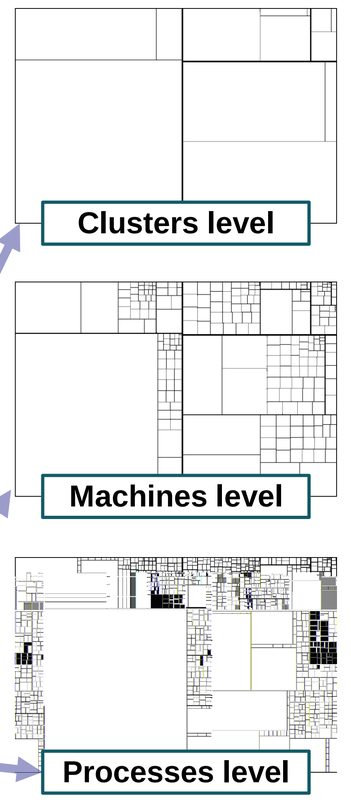
Processes Aggregations

The TRIVA Software

Project: building tools for the analysis of large-scale distributed systems
Data: data flows, communications, computing powers, internal states
Aggregations: in space and time

Which parts of the hierarchy should be aggregated?





Lamarche-Perrin, Vincent, Demazeau. Informational Measures of Aggregation for Complex Systems Analysis. Reseach Report RR-LIG-026, LIG, France, May 2012.