SeCoGIS 2012, October 16, 2012, Florence, Italy

Towards a spatio-temporal form of entropy

Christophe Claramunt
Naval Academy Research Institute, France
claramunt@ecole-navale.fr

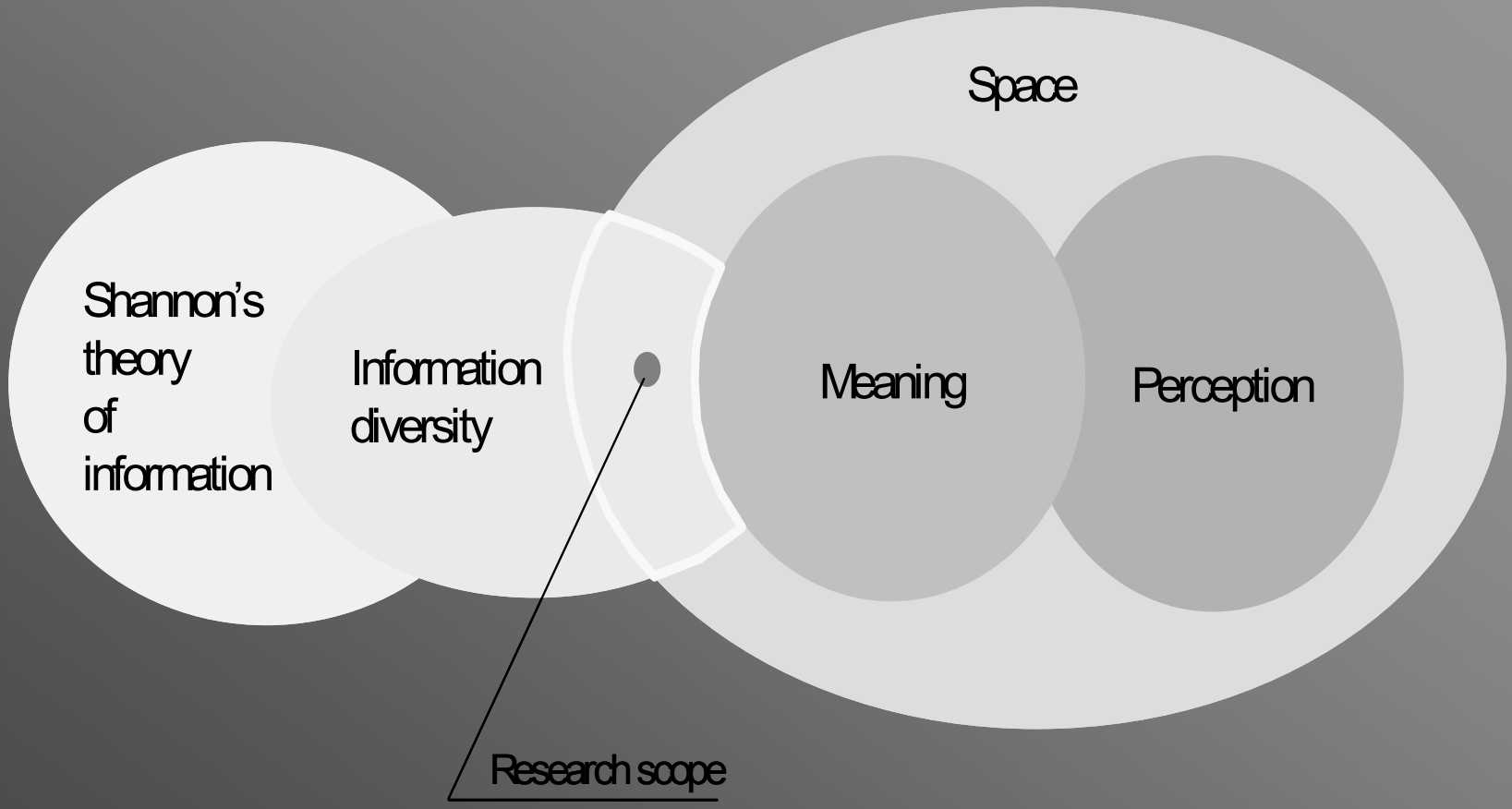
Talk and research outline

- Shannon's information theory has long provided a mathematical framework for statistically evaluating the diversity a given signal is able to produce.

- However, the initial concept of entropy cannot be directly applied to evolving geographical systems.

- We introduce a series of entropy measures that integrate the spatial and temporal dimensions, and evaluate the distribution of entities, events and categories in space and time.

From information to perception



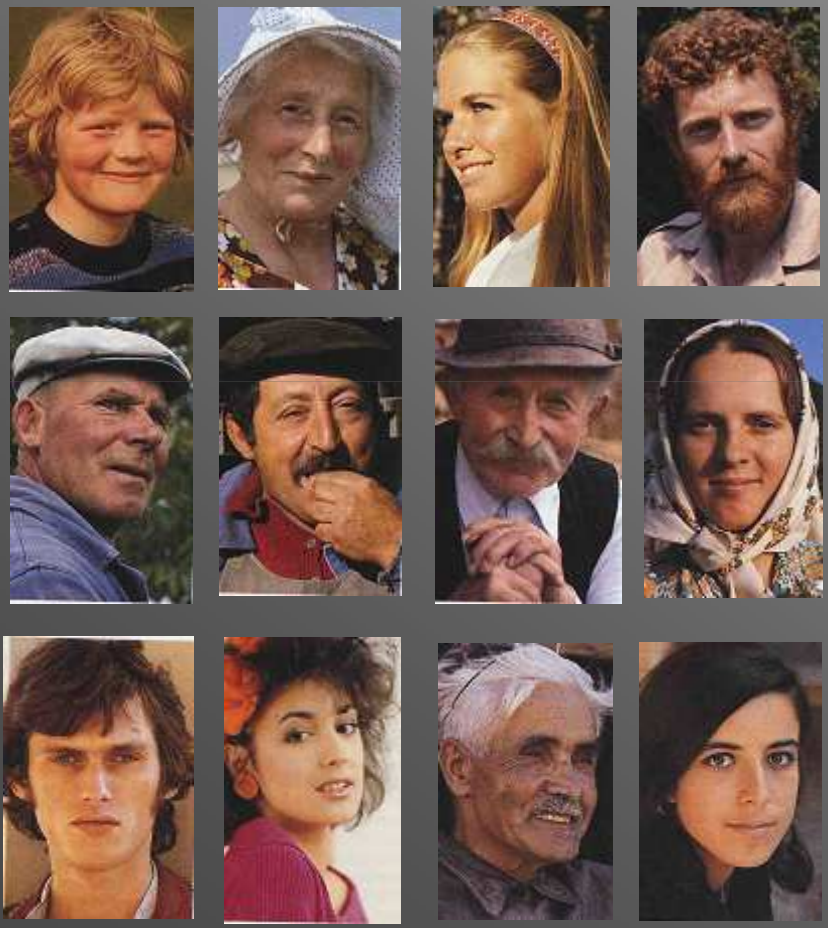
From the physical world to our minds, legibility, categories and structure

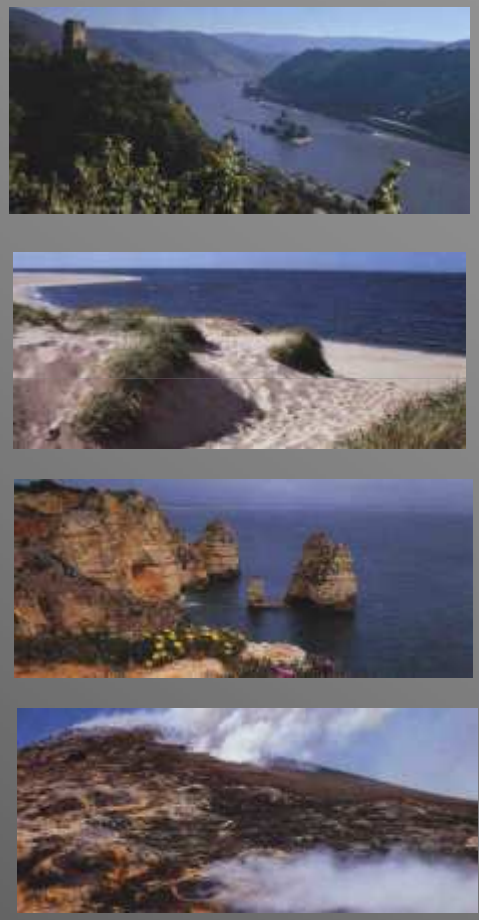
Diversity

Related to variety and absence of uniformity

Collection of different things, i.e. categories

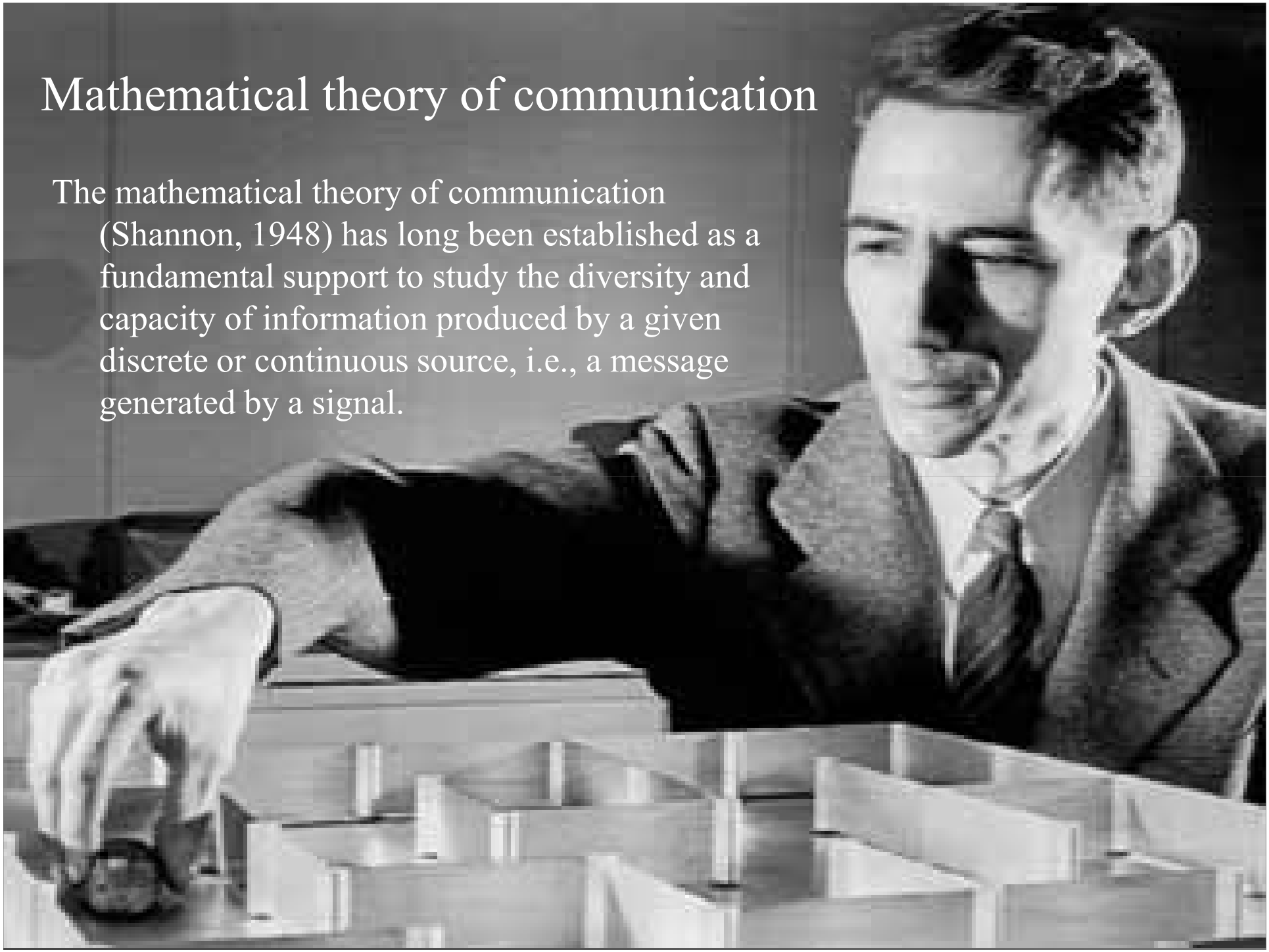
But there is a diversity … of diversities





Mathematical theory of communication

The mathematical theory of communication (Shannon, 1948) has long been established as a fundamental support to study the diversity and capacity of information produced by a given discrete or continuous source, i.e., a message generated by a signal.



Information theory basics

- Shannon's definition of entropy is given by

$$H = -K \sum_{i=1}^{n} pi \log_2 (pi)$$

where p_i is defined as the proportion of the total number of entities N_i of the class i over the total number N of entities, i.e

$$pi = \frac{Ni}{N}$$

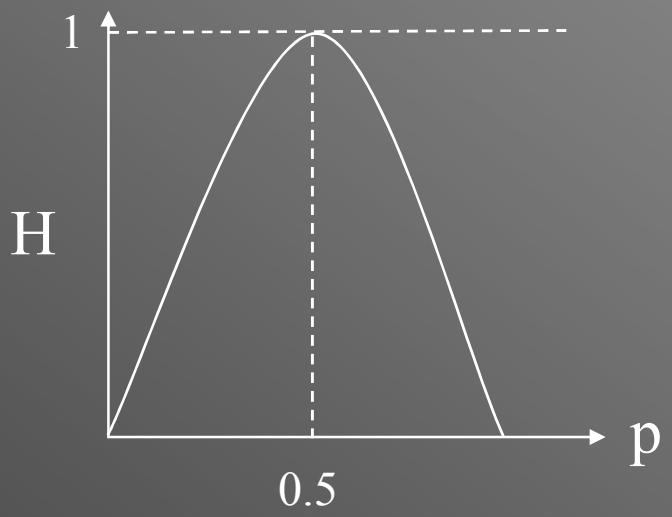
n is the number of classes, and K a positive constant

Information theory basics

- Shannon's definition of entropy is given by

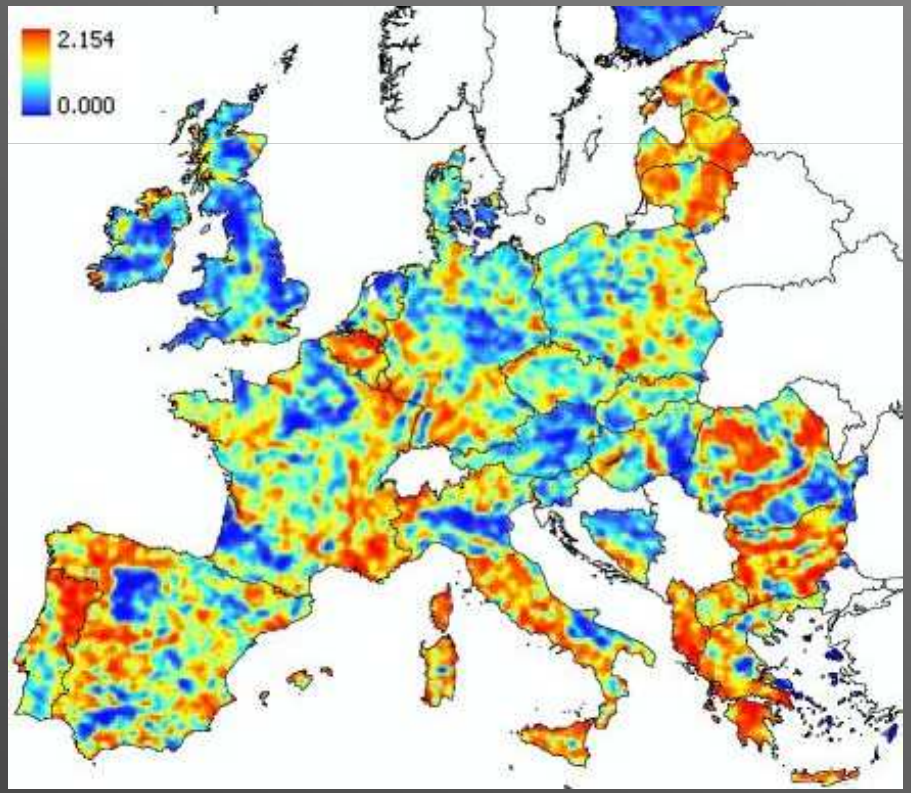
$$H = -K\sum_{i=1}^{n} pi \log_2 (pi)$$

For $p_1 = 1 - p_2$



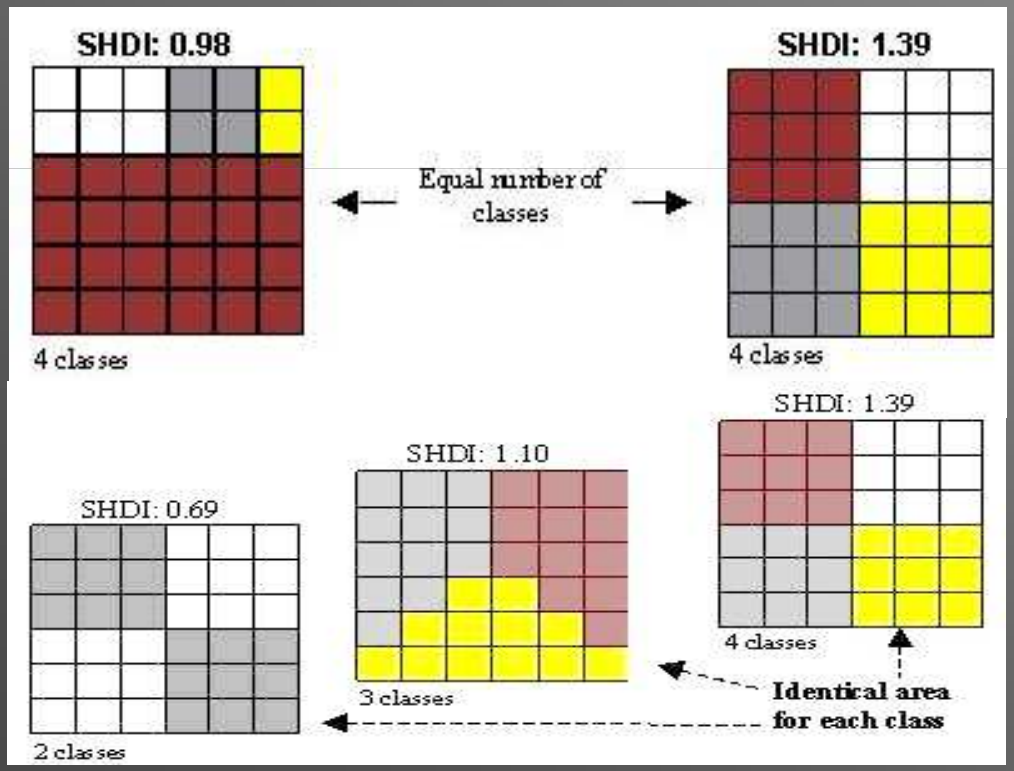
Entropy as applied to space

- Entropy can evaluate the distribution of category types in a given landscape.



Entropy as applied to space

- Entrlopy can measure the distribution of category types in a given image-based representation of space



From http://europa.eu.int/comm/agriculture/publi/landscape/ch1.htm

Entropy vs. space: related work

- The measure of *contagion* gives the degree to which patches of the same attribute class are clumped into patches of the same attribute class :

$$C = 1 + \sum_{i}^{n} \sum_{j}^{n} a_{ij} \frac{log_2(a_{ij})}{2 log_2(n)}$$

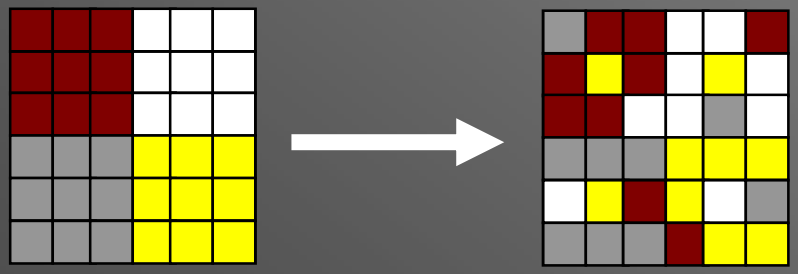
where n is the number of classes

a_{ij} evaluates to which extent regions of a given class i are adjacent to regions of another class j

Diversity vs. space: related work

- A representative example is given by the measure of *contagion* evaluates to which degree patches of the same attribute class are clumped into patches of the same attribute class :

$$C = 1 + \sum_{i}^{n} \sum_{j}^{n} a_{ij} \frac{log_2(a_{ij})}{2 log_2(n)}$$



Related measures include fragmentation, spatial heterogeneity and dominance indices

a_{ij} gives the proportion of patchs i adjacent to patchs j
Clumped (→ 1) vs. dispersed patches (→ 0)

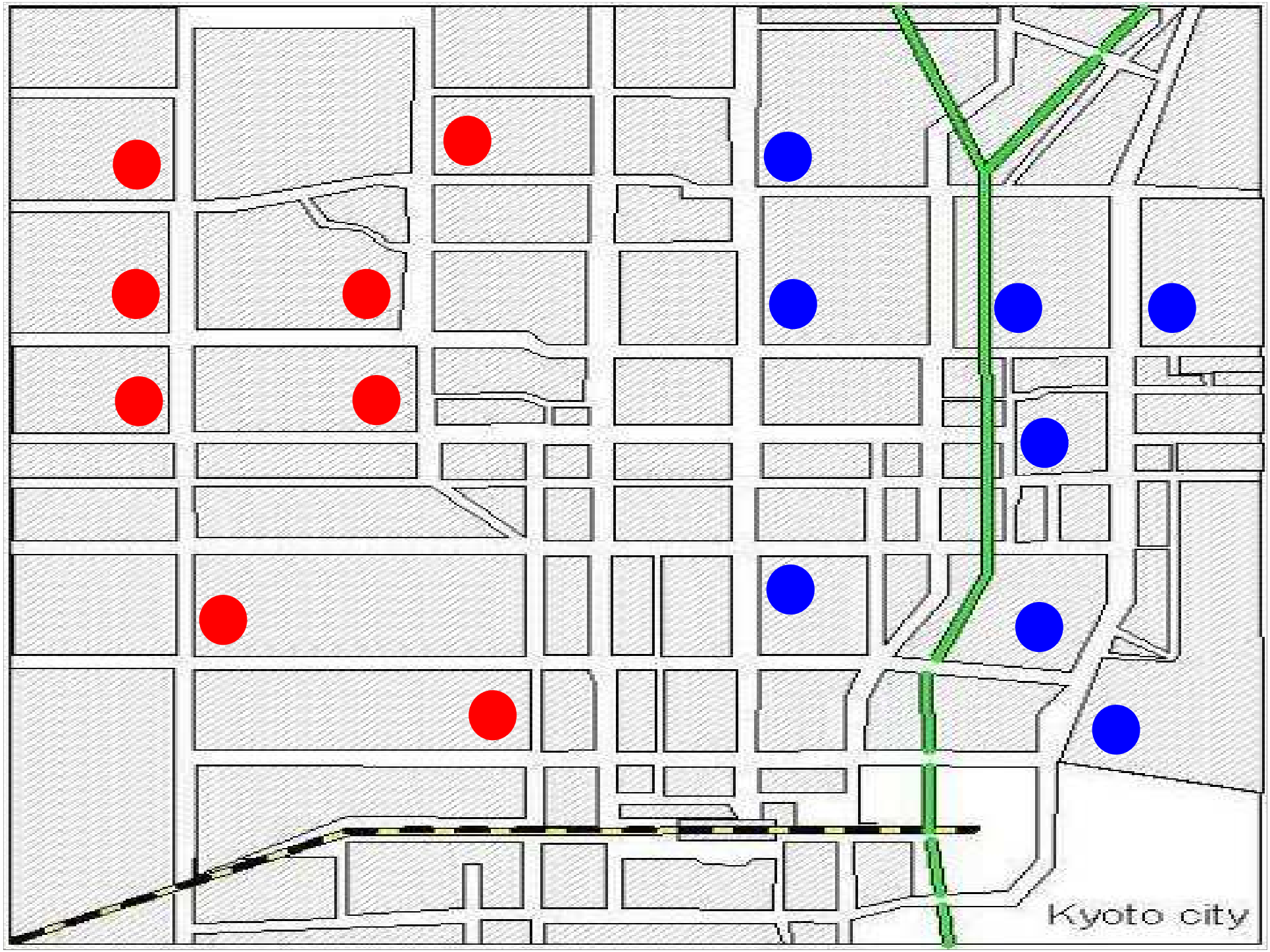
Limitations of current work

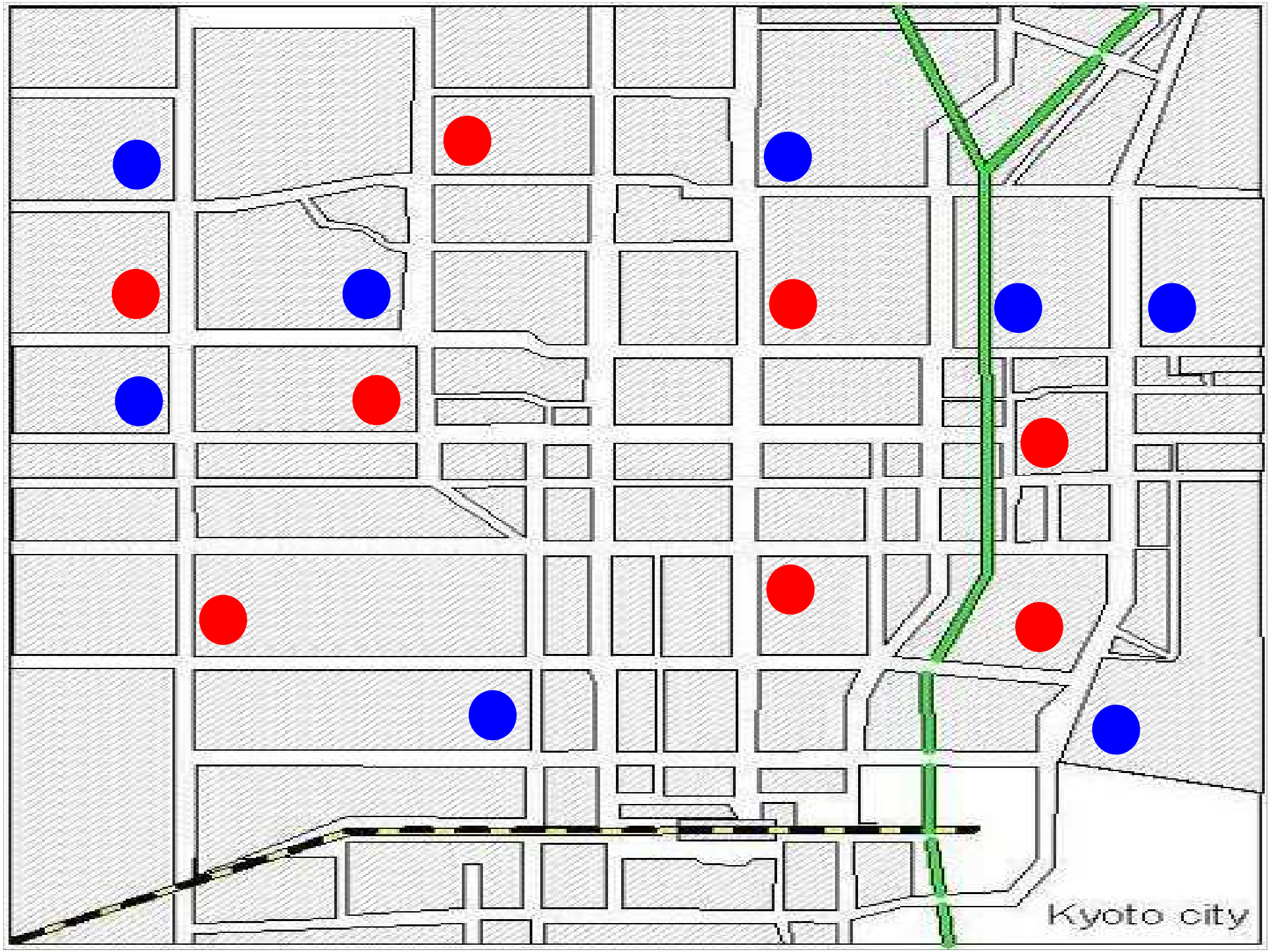
- Such measures of diversity are independent of the relative dispersion and distances between the regions that compose the different classes, this leading to a lack of complete consideration of the underlying spatial structure and arrangements.

- When applied, adjacency relationships introduce a quantitative bias as this measure does not take into account proximities of higher orders.

- Adjacency is not always the most predominant spatial relationship and property of interest.

Let us consider a discrete space, such as landmarks in an urban space or the distribution of a built environment


Kyoto city


Kyoto city

Theory of communication vs. space

- Surprisingly, and with the exception of the field of image processing, a few studies have explicitly and qualitatively evaluated to which extent Shannon's quantitative theory principles can be applied to space, and how space influences diversity.

- This is probably due to the fact that space is a specific form of multi-dimensional system where the different dimensions are intimately linked, while discrete systems studied by Shannon are made of messages decomposed into one-dimensional signals.

Diversity vs. space ?

- One intriguing question is whether the notion of diversity as it is defined in the theory of information is, or not, influenced by some of the fundamental properties space generates and conveys.

- From another point of view it has been already observed that information theory will be useful in clarifying some quantitative aspects of geographical information in the future (Tobler, 1997).

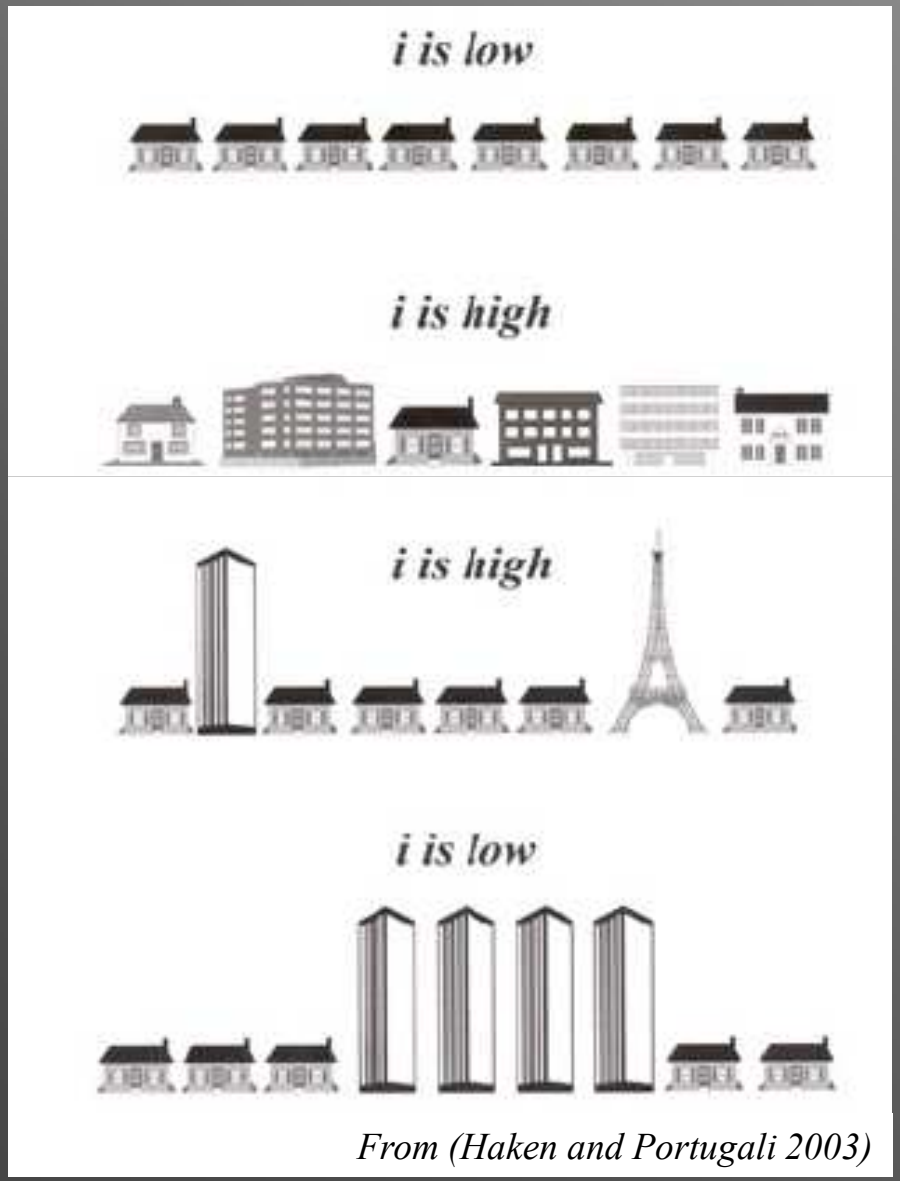
- This might be useful to evaluate how much information and diversity a geographical system can produce.

So still a need for a spatial and temporal form of diversity

But how to measure the roles played by space and time ?

Are there any primal relationships in space & time ?

Diversity vs. landmarks vs. space ?



From (Haken and Portugali 2003)

Related to the role of categories, perception and memorisation (Haken and Portugali 2003)

But also to the way landmarks are distributed in space

Diversity vs. semantic categorisation

- Haken and Portugali (2003) made a conceptual connection between the degree of information conveyed by a given spatial system, grouping and categorization processes, and the degree of perception and memorization by humans.

- They also intuit that the original Shannon's measure of diversity should be adapted to space, particularly by integrating a factor that evaluates the **probability of finding two different entities at the same location.**

- Our objective is to make an attempt to introduce an additional component to Haken and Portugali's systematic approach by specifically representing and measuring the influence of the spatial dimension on a measure of diversity, rather by evaluating it using probabilities

First law of geography (Tobler 1970)

Everything is related to everything else, but near things are more related than distant things

This leads to the derivation of two basic rules

rule 1:*when different entities are closer, diversity increases*

rule 2:*when similar entities are closer, diversity decreases*

Intra- vs. extra-distances

- The Intra-Distance evaluates the average distance between the entities of a same class:

$$d_j^{int} = \frac{1}{Nj \times (Nj - 1)} \sum_{\substack{i=1 \\ i \in Cj}}^{Nj} \sum_{\substack{k=1 \\ k \neq i \\ k \in Cj}}^{Nj} d_{i,k}$$

- While the Extra-Distance calculates the average distance between the entities of a given class and the entities of the other classes:

$$d_j^{ext} = \frac{1}{Nj \times (N - Nj)} \sum_{\substack{i=1 \\ i \in Cj}}^{Nj} \sum_{\substack{k=1 \\ k \notin Cj}}^{N-Nj} d_{i,k}$$

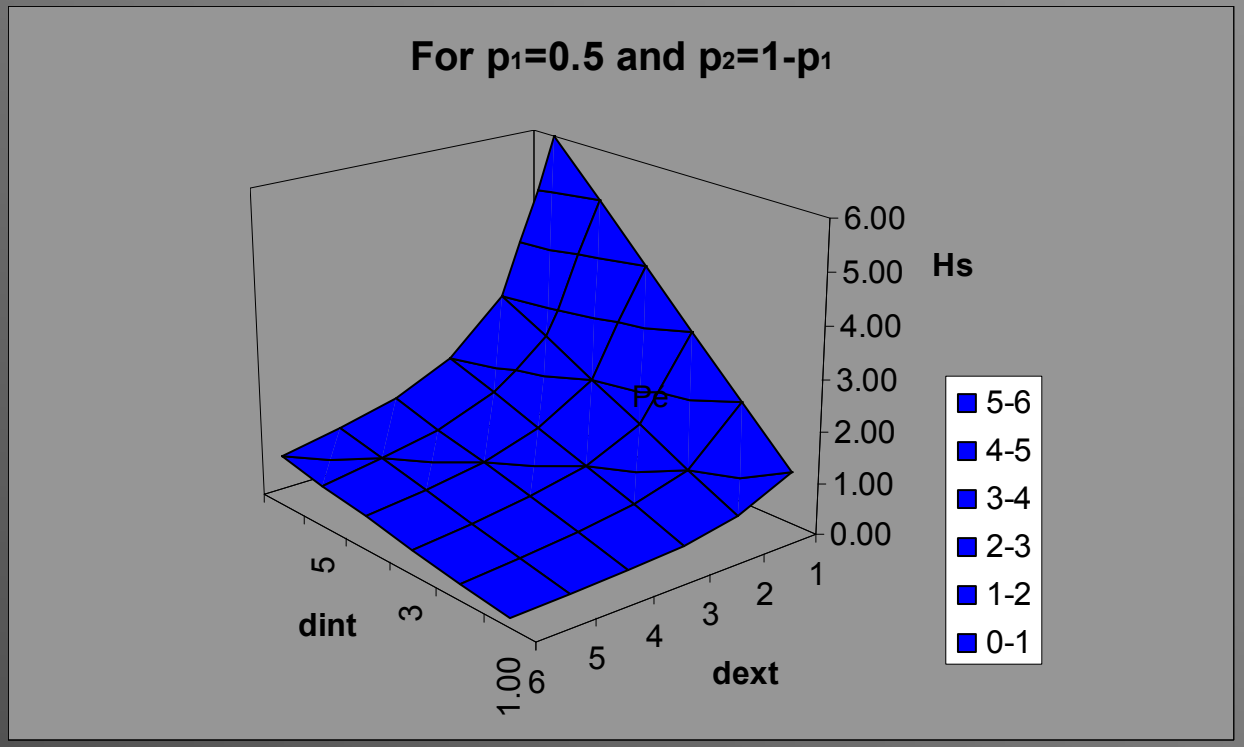
Spatial entropy

- We introduce a new measure of diversity, called *spatial entropy* :

$$Hs = -\sum_{i=1}^{n} \frac{d_i^{*\text{int}}}{d_i^{*ext}} p_i \log_2(p_i)$$

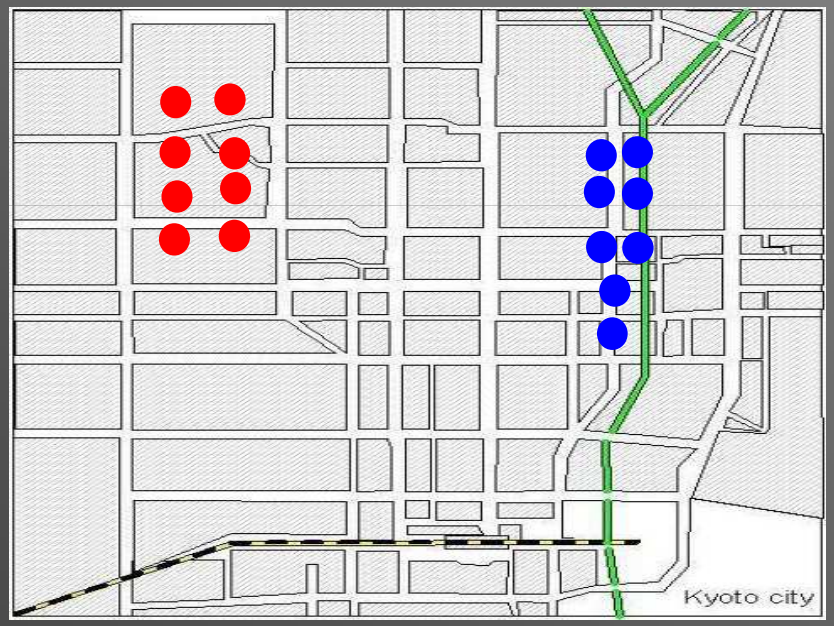
- The *spatial entropy Hs* is semi bounded by the real positive interval [0,+∞[.
- For some given *Intra-* and *Extra-Distance* values, *Hs* is maximum when the classes are evenly distributed.
- For a given distribution of classes, the *spatial entropy* increases when either the *Intra-Distance* augments, or the *Extra-Distance* decreases.

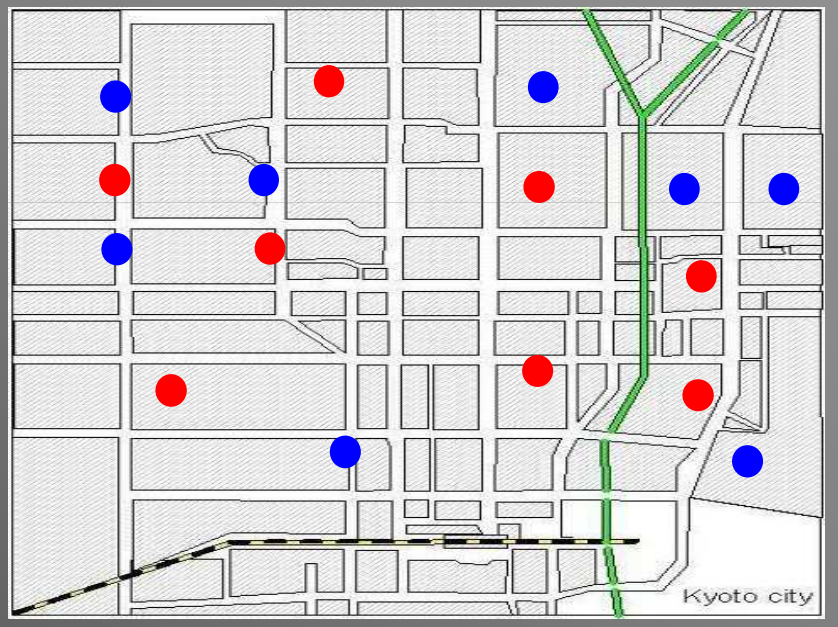
Spatial entropy: range of values



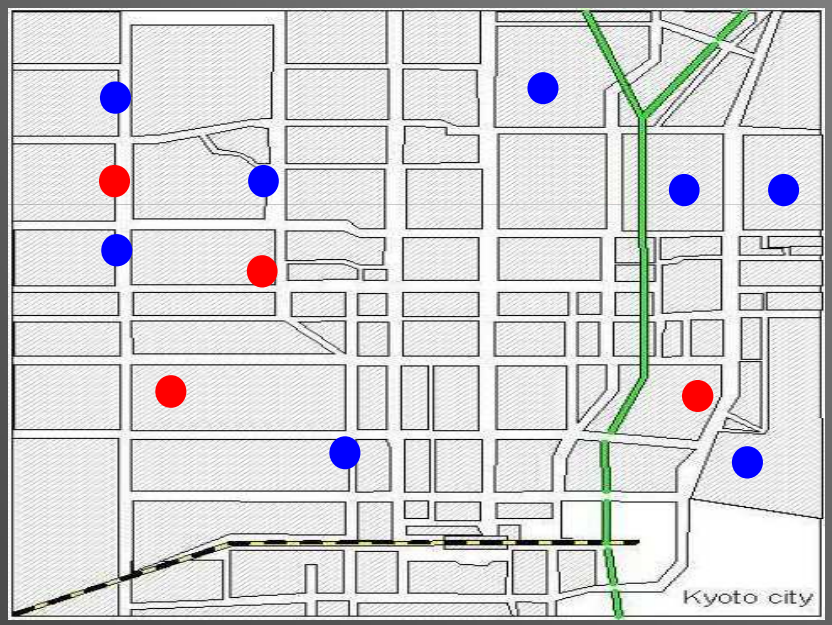
$$Hs = -\sum_{i=1}^{n} \frac{d_i^{int}}{d_i^{ext}} p_i \log_2(p_i)$$

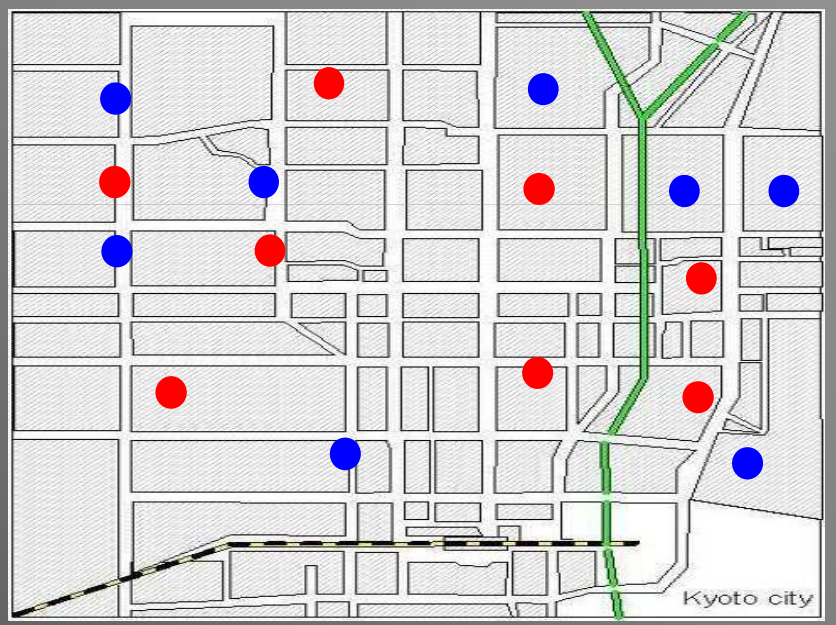
Spatial entropy: from low to high values



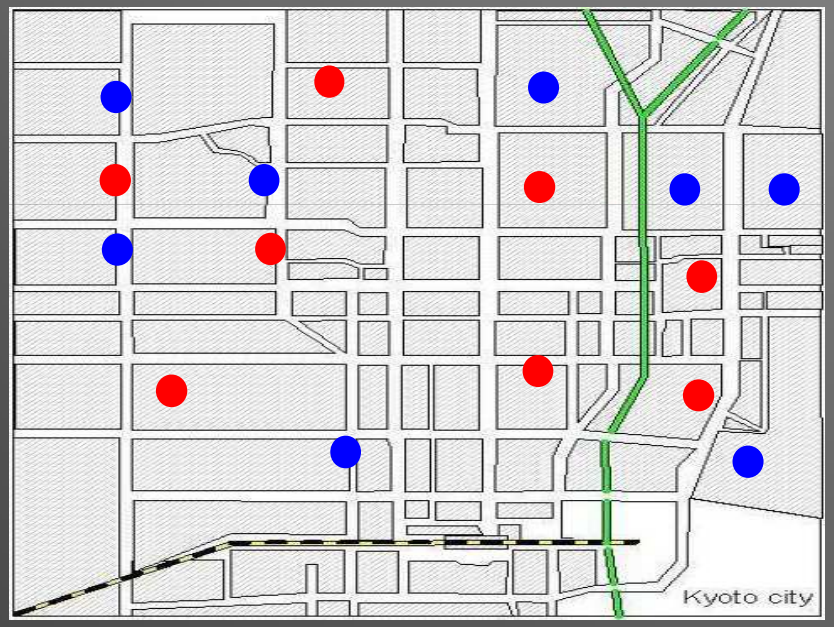


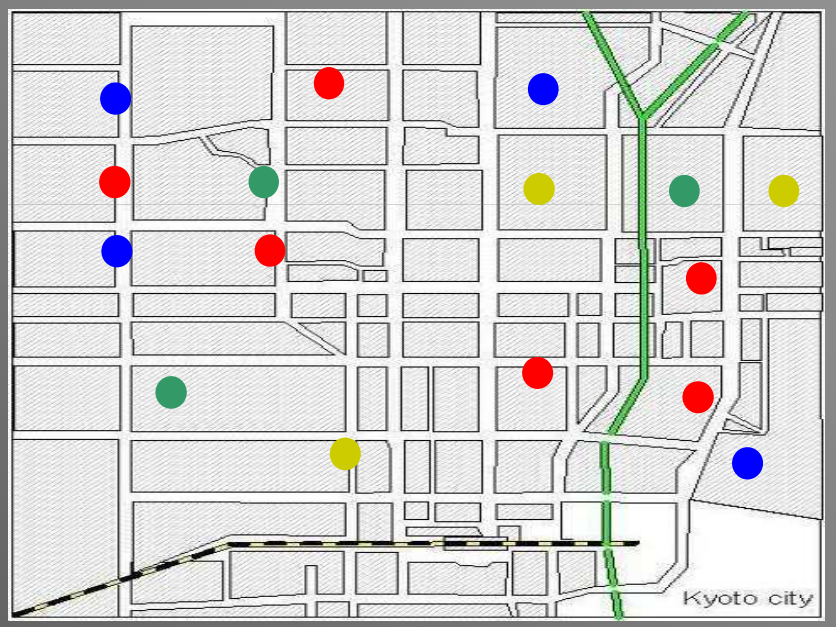
Spatial entropy: from low to high values





Spatial entropy: from low to high values





Spatial entropy: from low to high values

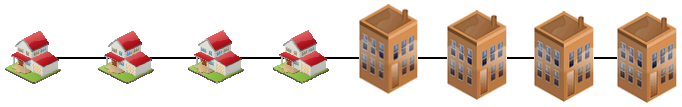


Figure 1.a *Hs* = 0.4

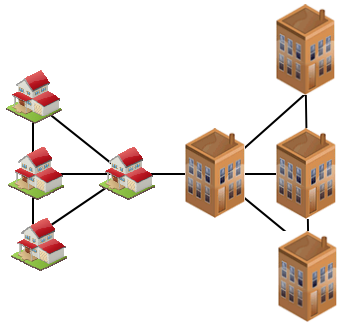


Figure 1.b *Hs* = 0.41

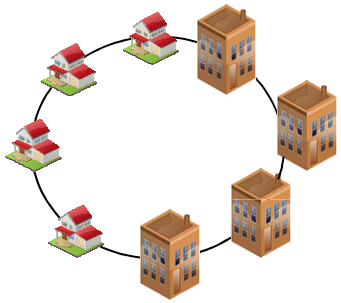


Figure 1.c *Hs* = 0.66

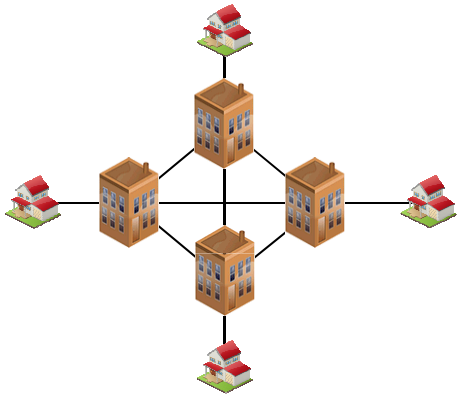


Figure 1.d *Hs* = 1.14

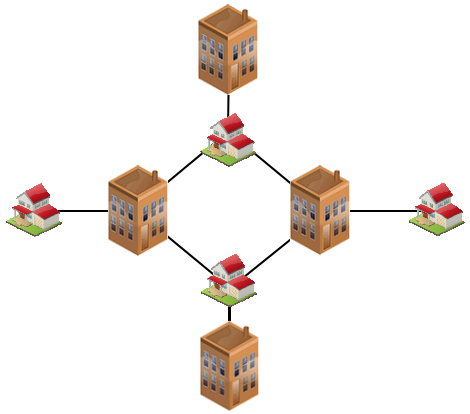


Figure 1.e *Hs* = 1.16

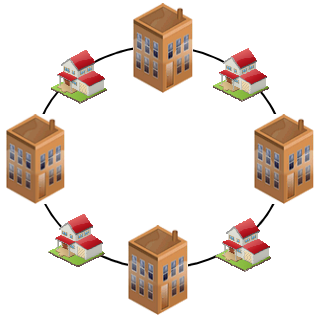


Figure 1.f *Hs* = 1.33

When considering time, and by extension

rule 1:*when different entities are closer in time, diversity increases*

rule 2:*when similar entities are closer in time, diversity decreases*

Intra- vs. Extra-Timedistances

- The Intra-TimeDistance evaluates the average time distance between the entities of a same class:

$$td_j^{\text{int}} = \frac{1}{Nj \times (N_j - 1)} \sum_{\substack{i=1 \\ i \in Cj}}^{Nj} \sum_{\substack{k=1 \\ k \neq i \\ k \in Cj}}^{Nj} td_{i,k}$$

- While the Extra-Time distance calculates the average time distance between the entities of a given class and the entities of the other classes:

$$td_j^{ext} = \frac{1}{Nj \times (N - Nj)} \sum_{\substack{i=1 \\ i \in Cj}}^{Nj} \sum_{\substack{k=1 \\ k \notin Cj}}^{N-Nj} td_{i,k}$$

Temporal entropy

- We introduce a new measure of diversity, called *temporal entropy* :

$$H_T = -\sum_{i=1}^{n} \frac{td_i^{*\text{int}}}{td_i^{*ext}} p_i \log_2(p_i)$$

- The *temporal entropy* H_T is semi bounded by the real positive interval $[0,+\infty[$.
- For some given Intra- and Extra-TimeDistance values, H_T is maximum when the classes are evenly distributed.
- For a given distribution of classes, the *temporal entropy* increases when either the *Intra-TimeDistance* augments, or the *Extra-TimeDistance* decreases.

Temporal entropy: from low to high values

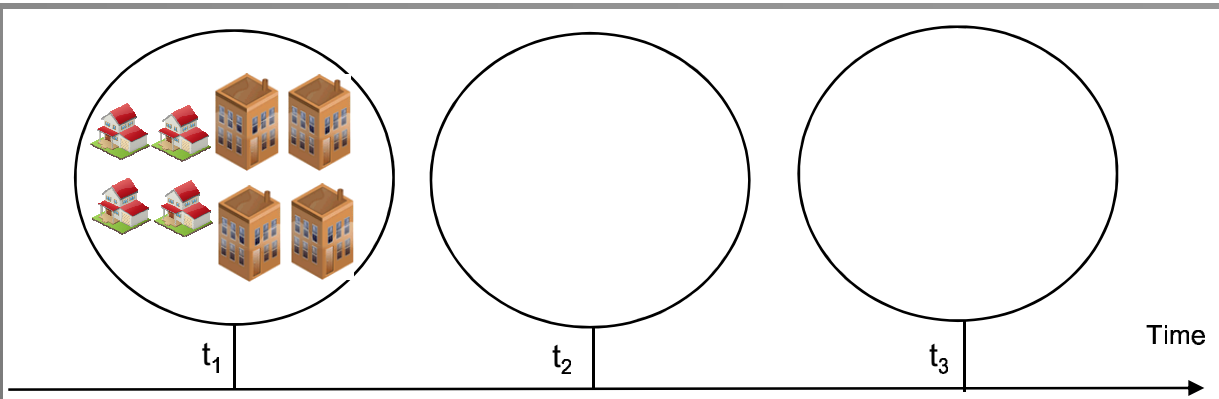


Figure 2.a $H_T = 1$

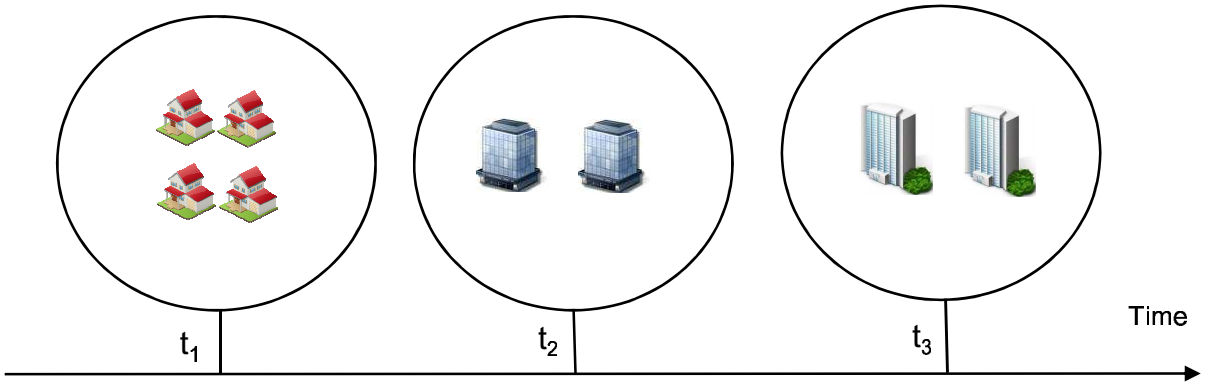


Figure 2.b $H_T = 0.53$

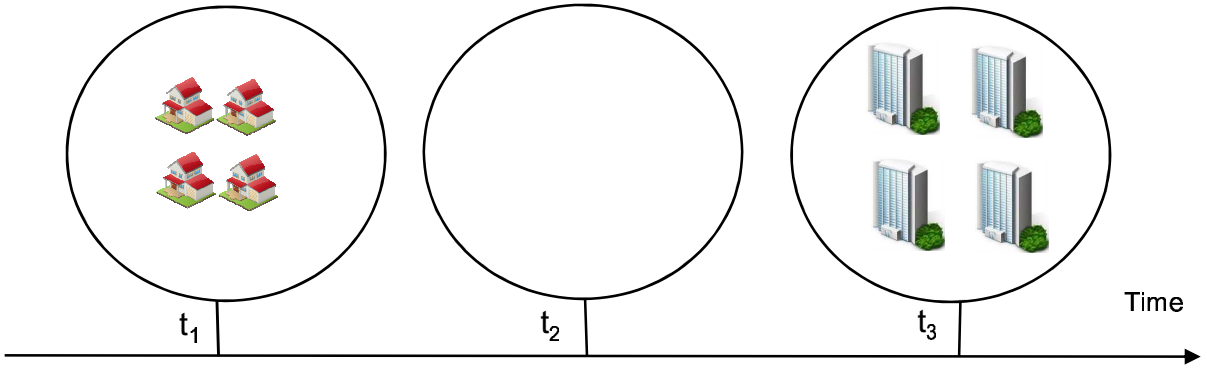


Figure 2.c $H_T = 0.33$

When considering space & time, and by extension

rule 1:*when different entities are closer in space and time, diversity increases*

rule 2:*when similar entities are closer in space and time, diversity decreases*

Intra- vs. extra-distances

- The spatio-temporal distances are defined as a cumulated influence and as a product of the spatial and temporal distances. They are given as follows.
  - Intra-SpatioTemporalDistance $std_i^{*\text{int}} = td_i^{*\text{int}} \times d_i^{*\text{int}}$
  - Extra-SpatioTemporalDistance $std_i^{*ext} = td_i^{*ext} \times d_i^{*ext}$

Spatio-temporal entropy

- The *spatio-temporal entropy* is then given as

$$H_{ST} = -\sum_{i=1}^{n} \frac{std_i^{*\text{int}}}{std_i^{*ext}} p_i \log_2(p_i)$$

- The *spatio-temporal entropy* H_{ST} is semi bounded by the real positive interval [0,+∞[.
- For some given *Intra-* and *Extra-SpatioTemporalDistance* values, H_{ST} is maximum when the classes are evenly distributed.
- For a given distribution of classes, the *spatio-temporal entropy* increases when either the *Intra-SpatioTemporalDistance* augments, or the *Extra-SpatioTemporalDistance* decreases.

Spatio-temporal entropy: from low to high values

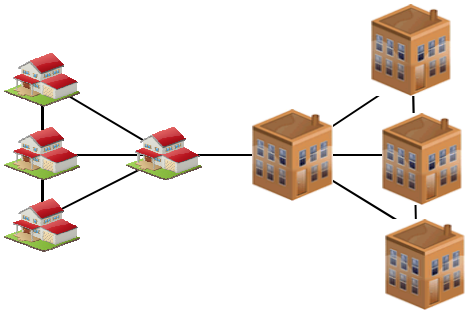


Figure 3.a $H_{ST} = 0.41$

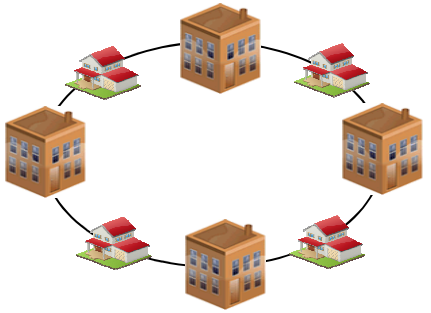


Figure 3.d $H_{ST} = 1.33$

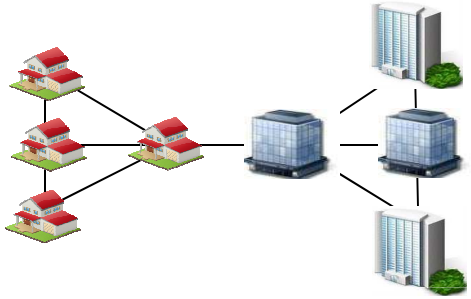


Figure 3.b $H_{ST} = 0.21$

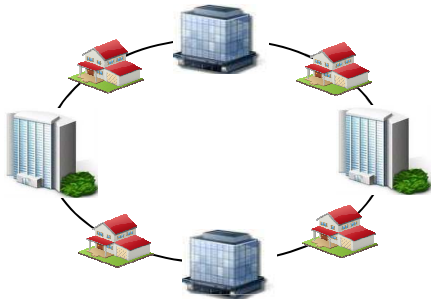


Figure 3.e $H_{ST} = 0.60$

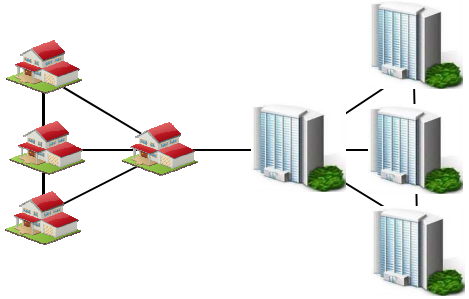


Figure 3.c $H_{ST} = 0.13$

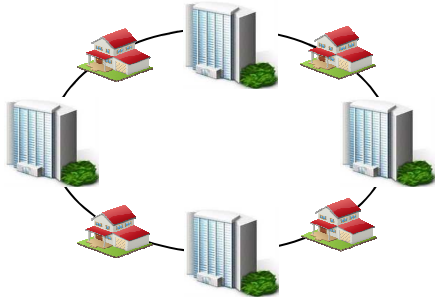
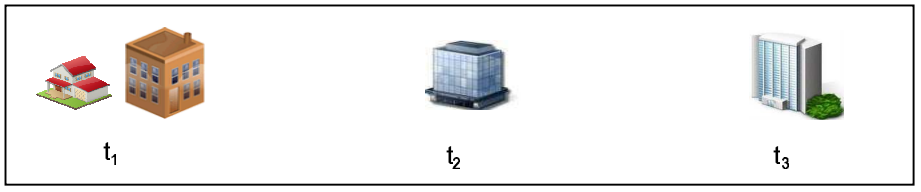


Figure 3.f $H_{ST} = 0.38$



Concluding remarks

- The analysis of the distribution of things in space and time is still a domain that requires the development of intuitive values that reflect the spatial and temporal structural properties.
- The research presented introduces a series of spatial, temporal and spatio-temporal measures that complete entropy, diversity measures and the way those are applied to space and time
- Those measures should support additional reasoning mechanisms which are still left for further exploration