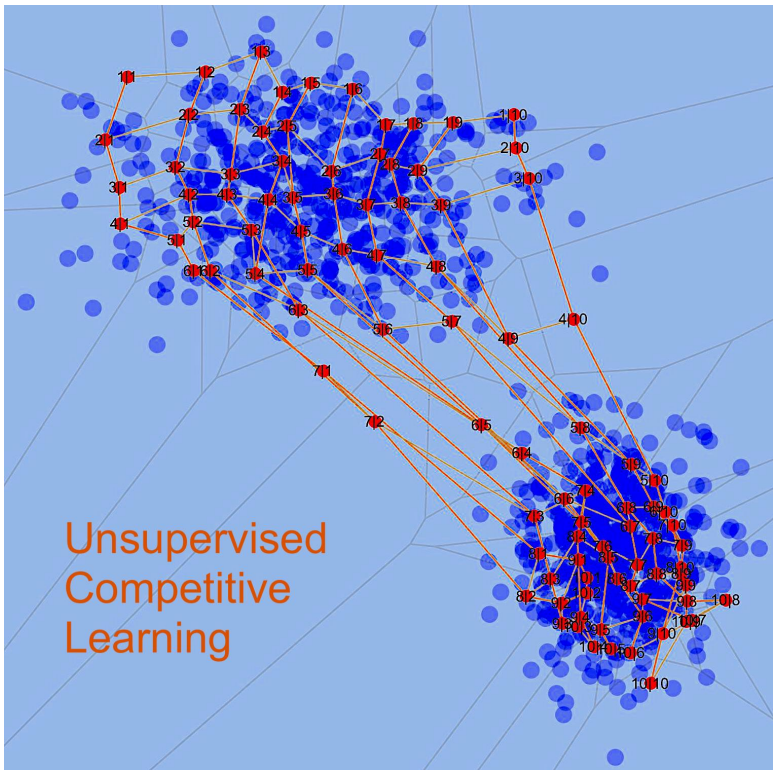
Mapping features onto categories
Unsupervised Competitive Learning

Jacques Ambühl

2023, July 07.



Abstract

This essay is devoted to Unsupervised Competitive Learning (USCL) in neural networks. It is divided in five sections, each describing specific traits of this learning procedure. The introduction presents some key features of USCL from an overarching perspective.

Making use of a toy-like network, fundamental properties of USCL are discussed in the second section. A full fledged application of USCL aimed at extracting societal similarities and differences among countries is developed in the third section. The section four and five highlight two properties of neural networks developed according to USCL.

How do we map features onto categories? An attempt of my own making, related to interactions putatively established between philosophy and artificial intelligence, concludes the essay.

Two preliminary warnings are in order:

It must be emphasized at the outset, in all modesty, that the objects presented in the following are infinitely simpler than biological systems of nerves and brains found in nature.

Authors usually avoid nouns such as neuron or synapse while designating software structures mimicking biological organs, and prefer surrogates such as units or connexions. Fully aware of my effrontery, I dare to use the biological wording in the sequel. A thesaurus provided in annex 7.3 aims at delivering matches between the terms and concepts introduced below.

Contents

1 Introduction

Bionic is the term used to describe the approach of an engineer who draws inspiration from living nature when planning and realizing a machine or technical system. Leonardo da Vinci observes bats when he imagines his flying machines. The first aircraft designers try to imitate the flapping of birds' wings. Numerous modern mechanical and architectural constructions are inspired by motifs from the vegetal and animal worlds. What about the structures of the brain? These were gradually revealed in the course of the 20th century, and the first models of neural networks inspired by these discoveries emerged in the 1950s as a result of the invention of the computer.

As virtual structures, neural networks obviously only exists in memories of computers, thus the question arises as to where their bionic peculiarity remains hidden. It consists, among other things, in the fact that the network behaves like a dynamic, integral, holistic memory in which the information is no longer stored byte by byte, as would be the case in a computer memory, but is distributed over the entire network.

Another essential bionic property is the learning capability that the network must display in order to fulfill its task. In this learning phase, the network is confronted with a set of patterns or stimuli, or situations (see annexes 7.2 and 7.4) extracted from the outside world. Memorization consists in gradually implementing the information gained from observing each of these patterns into the synapses of the totality of neurons in the network. In this type of so called unsupervised and competitive learning process, the network is not presented with ready-made answers it should have to reproduce, but it is forced to organize itself, obeying a simple scheme by which the neurons compete among themselves in an effort to grasp the essence of the patterns presented[1]. This type of learning was observed clinically in the 1960s, and Finnish engineer Teuvo Kohonen formalized it as an algorithm by the end of the 1980s.

[1]In contrast, supervised learning (usually dubbed Deep Learning) is another neural algorithm whose biological counterpart occurs primarily in the limbic regions of the brain. It consists of responding in an optimal way to a stimulus for which the response is known. Reflexes are based on this type of learning. The associated networks are called "perceptrons", and the learning type associated with them is called "back-propagation of the gradient".

1.1 Brief historical note

For me, everything started in 1992 at a third cycle course given at the Ecole Polytechnique Fédérale de Lausanne - EPFL - and titled "Réseaux de neurones naturels et artificiels", [2]. On this occasion, I developed toylike neural networks aimed at evaluating weather fields and classifying them with the help of a micro expert system written in Prolog. At that time, formal "artificial intelligence" based on logical and declarative languages such as Lisp on the syntactical, or Prolog on the semantical edge were praised and implemented in so called expert systems. Later, working together with Daniel Cattani and Pierre Eckert at the Swiss Meteorological Institute, we started to develop networks able to cluster meteorological ensemble forecasts, [3], whereby the seminal book by Herz, Krogh and Palmer, [1], was our ultimate reference in the field. Neural based assessment systems based on ECMWF ensemble forecasts were implemented in the operations suites of Meteo Swiss[2] by the end of the century and have been routinely operated since then.

Some of the properties of these meteorological neural networks are presented in Annex 7.1. In their simplest architecture (Figure 1), such neural networks consist in a square array of 12 x 12 neurons, each neuron being equipped with a bundle of 200 synapses that are both information sensors and analogue storage elements. The synapses are projected from each neuron onto the vertices of a grid of 10 x 10 points on a map of the whole of Central Europe. References [4 and 5] provide comprehensive information in this respect.

Lately, Making use of the time at disposal after retirement, I started refurbishing old and forsaken pieces of code into a comprehensive and modular structure written in the Wolfram's Mathematica language. First elements of this application had been gradually programmed during the 30 past years in the Pascal, C++ and finally Mathematica versions available at each corresponding epoch. All of this resulted in a clutter of inefficient pieces of software that I decided to refurbish during the autumn and the winter 2022-2023. All the figures hereafter presented have been established with this newly written

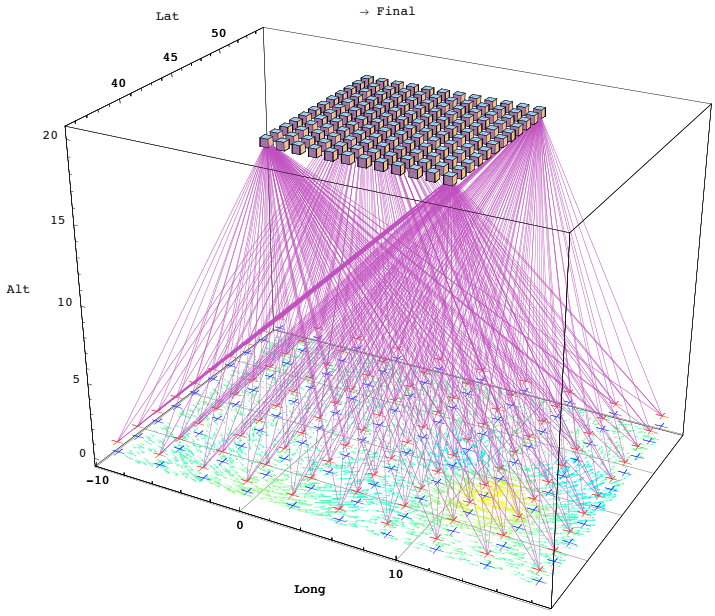


Figure 1: *Schematic representation of a neural network proceeding according to USCL. Each small cube symbolizes a neuron. Its synapses, simultaneously sensors and memory elements, are sketched as purple descending lines. In this depiction, the drawing "on the ground" represents a weather map.*

piece of software whose modular structure is sketched in Annex 7.4.

2 Fundamental properties from a toy network

The best possible way to describe USCL consists in starting with the toy example of a square network, similar to the one presented in Figure 1, this time, however, with 10 x 10 neurons.

Two surfaces, or "spaces", characterize the network. The first one is the network, or grid itself, as pictured in Figures 1 and the left panel of 2. The second space, on the right panel, is abstract. I name it synaptic, or semantic space. It is the vector space in which the synapses take their values. As each

[2] A new name for the venerable Institute

neuron has the same number of synapses, the geometrical dimension of the space is this number. As the synaptic weights are real numbers, it is defined on the field $\mathcal{R}^2$ of the real numbers. Each neuron is represented in this space as a point (a vector) whose coordinates are its synaptic weights. As we are now working with a network whose neurons are equipped with two synapses only, the synaptic space is two dimensional, as featured on the right panel of figure below. The edges drawn on both panels correspond, in a way that we now describe. A comprehensive interpretation of the synaptic space as a semantic space will take its sense later in Section 5, with more complex applications of the algorithm.

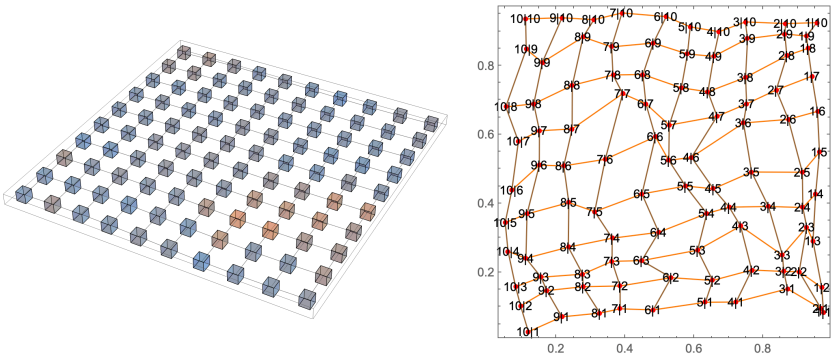


Figure 2: *Left, Neural Network. Right, synaptic, or semantic Space. Two distance will be implemented, the first one, designed by δ, will operate on the network. The second one, named $\mathcal{D}$, will operate on the synaptic space. This figure is the exact correspondent of the Figure 9.*

In this almost trivial configuration, the task to be learnt is a set of 1000 couples of numbers randomly generated in a $[0, 1] \times [0, 1]$ square, as pictured on the upper left panel of Figures 3 and 4. This dataset is provided as a matrix X whose dimensions is 1000×2 in our case. These points or coordinates are located in the synaptic space related to the task to be learned.

As one notices, the distribution of the points, or stimuli, is uniform on Figure 3, and it is bimodal on Figure 4.

The upper right and lower left panels of both figures show the state of the network after the termination of the learning process. To each vertex on the panels corresponds a neuron, $\{i, j\}$ as labelled on the diagrams, located

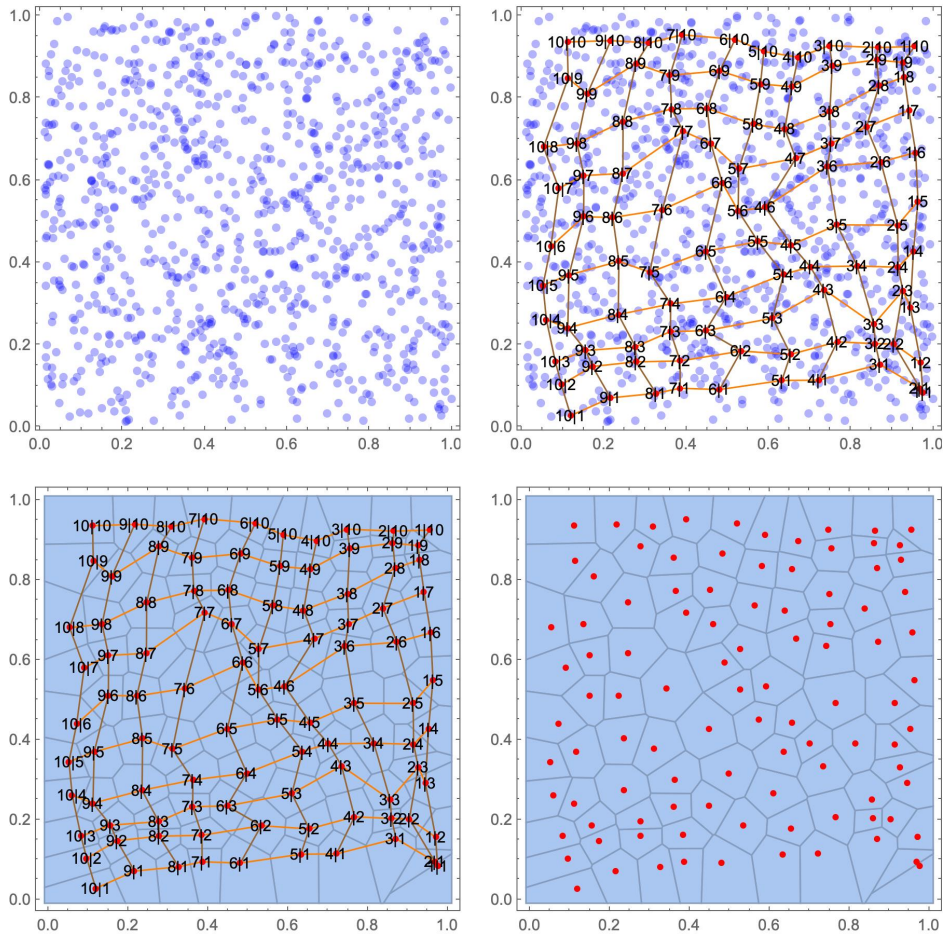


Figure 3: *Uniform distribution of the learning task. 2 synapses network.*

in the synaptic space at the point defined by its synaptic weights. The edges between neurons describe the neighborhood relation among neurons, as pictured on Figure 1 and the left panel of Figure 2.

Progressing in the description of the figures, the lower bluish panels describe one again the state of the network after completion of the learning process, this time, however, without the stimuli distributions. The red dots represent the locations of the neurons in the synaptic space, the edges on the left panels the connections among them. What matters now on the lower panels of Figures 3 and 4 is the Voronoï tessellation in the background. Each

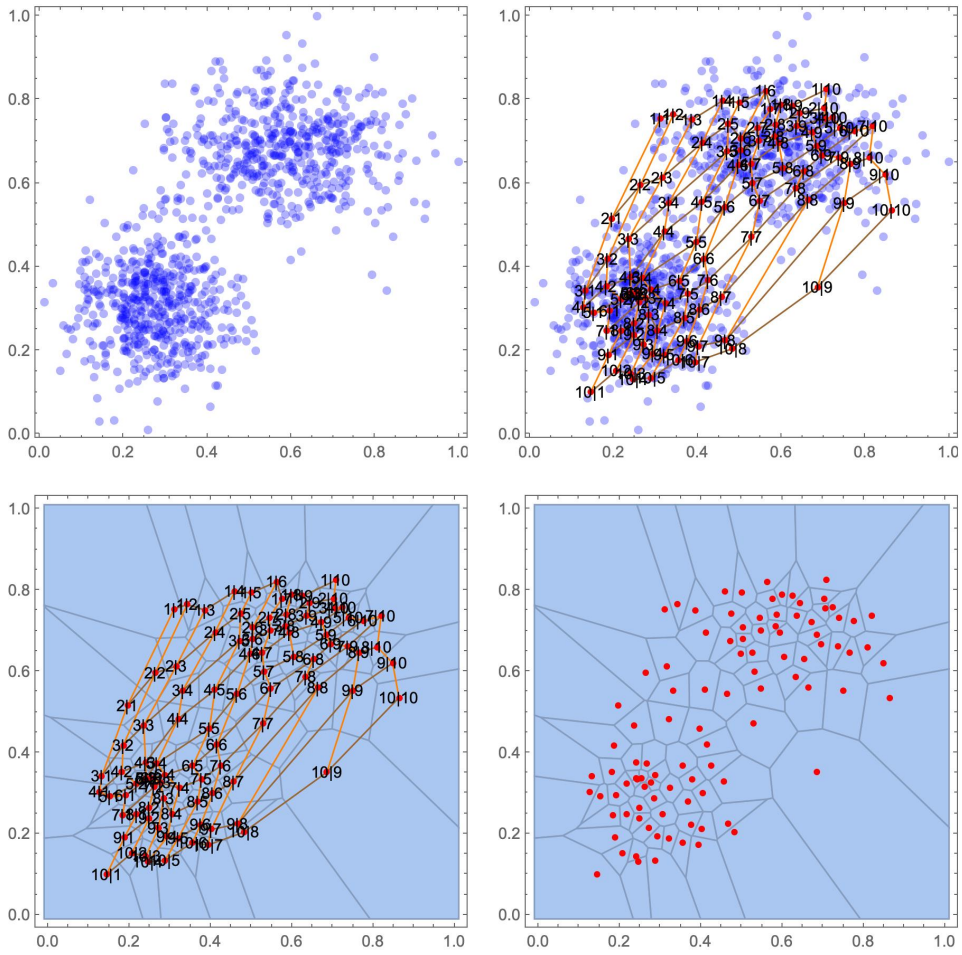


Figure 4: *Bimodal distribution of the learning task. 2 synapses network.*

tile represents the "domain on influence" of a corresponding neuron in the synaptic space and the whole tessellation is akin to a partition of the space in districts allocated to the neurons.

The spread of the districts happens to be uniform on Figure 3, in accordance to the uniformity of the the task having been learned. On the contrary, two clusters of crowded districts appear on Figure 4, together with a ring of almost empty districts towards the frontiers, well reflecting the bi-modality of the initial task.

The network happens to satisfy a property closely related to biological

reality: as a simple example, the somato-sensorial map of our body located in the cortical area of our brain is much more accurate for frequently simulated and thus more sensitive areas (the tongue), than for less sensitive areas (the back)[3]. As shown in previous figures, if the statistical distribution fields Ψ presented at the time of the learning is not uniform in the (synaptic) input space, the spatial organization of the synaptic weights reflects this statistical inhomogeneity. The location of the synaptic weights aggregate in areas where stimuli have been frequent and they are sparse in areas rarely visited during the learning process. Thus the network $\mathcal{N}$ is more accurate in frequently simulated areas, a property that can be formalized in the following way: if the Voronoï tile associated to a neuron $\{i, j\}$ is named $V_{\{i,j\}}$ and if $prob(\Psi \in d\sigma)$ is the probability of finding a stimulus of the Ψ field in the elementary volume $d\sigma$ of the synaptic space, then the area (volume) of a the Voronoï tile is small in a frequently visited area, large in seldomly visited areas, and satisfies:

$$\forall \{i, j\} \in \mathcal{N} \ : \ \int_{V_{\{i,j\}}} prob(\Psi \in d\sigma)\, d\sigma = constant$$

The size of each $V_{\{i,j\}}$ domain is the tuning parameter in the previous expression. As a consequence, the system happens to be suited to distinguish minute differences between patterns that are common in the learning task, and not so able to accurately discriminate seldom features.

This approach will enable us to define in following Section 3.4 the entropy of the learning process, that will be implemented as a measure of its efficiency.

2.1 Learning sequence

Following Figure 5 exhibits three phases of the learning process for both distributions, uniform on the left, bimodal on the right column. The initial state of the network is shown in the upper row: the synaptic weights have been given small random values. Accordingly, they are all crowded in the center of the square. In the middle row, at one third of the learning process, the networks are unfurling and spreading according to the respective distributions of the stimuli in the synaptic space. On the bottom row, at nine tenths of the learning process, the networks have almost reached their final states and cover mostly regularly their respective stimuli distributions. The

[3]The somato-sensorial cortex is located as a diadem beneath the skull, from ear to ear.

final states are those presented in Figures 3 and 4.

The "competition" among neurons occurs according to the following stochastic scheme in which a stimulus is defined by its coordinates (x, y) in the synaptic space, defined as $s_{p,(x,y)}$, (and represented as a blue point on previous Figures 3 and 4).

At each step p of learning process:

1. a stimulus s_p, determined by its coordinates (x, y) in the synaptic space is randomly chosen among the set of stimuli to be learned. The closest neuron to it is selected according to the euclidian metric $\mathcal{D}$ and is dubbed "the elected neuron" $\{i, j\}^*$:

$$\min_{\{k,l\}\in\mathcal{N}} \mathcal{D}[(x, y), \{k, l\}] \quad \rightarrow \quad \{i, j\}^*$$

The metric $\mathcal{D} : \mathcal{S} \times \mathcal{S} \rightarrow \mathcal{R}^+$ operates on the synaptic space and measures the actual distance between one neuron (its synaptic weights) and one stimulus, or between two neurons, located by their respective synaptic weights. It can be euclidian or defined using another measure, as for example $\mathcal{D}_{[X,Y]} = (X - Y)^T V^{-1}(X - Y)$, where V is the variance-covariance matrix of the input dataset: $V = X^T X$. This distance will be used later and named $\mathcal{D}_m$. The euclidian distance $\mathcal{D}_e$ is simply realized when $V = I$, the variance-covariance matrix equals the identity matrix.

2. A so-called ball of activity, or neighborhood $\mathcal{B}_{\{i,j\}^*,(p)}$, is then determined around the elected neuron $\{i, j\}^*$ on the network:

$$\mathcal{B}_{\{i,j\}^*,(p)} \;=\; \left\{\{k, l\} \in \mathcal{N} \; such\, that \; \delta[\{i, j\}^*, \{k, l\}] \leq \eta_{(p)}\right\}$$

The metric $\delta : \mathcal{N} \times \mathcal{N} \rightarrow \mathcal{R}^+$ operates on the network $\mathcal{N}$, as sketched in Figures 1 and 2, and measures the actual distances between neurons, on the network. The network can be given topologies that are different than that of the square, providing it with a toroidal shape. This option will be exploited in following section 3.2. The function $\eta_{(p)}$ solely depends on the step p and decreases as the learning process goes on, thus gradually reducing the radius of the ball.

3. All the neurons located in the ball, as defined above, are forced to move their synaptic weights toward the stimulus (x, y) that is presented at

this step p in the input space:

$$\forall\{k,l\} \in \mathcal{B}_{\{i,j\}^*,(p)} \ : \ \{k,l\} \ + \ \tau_{(p)}[(x,y) - \{k,l\}] \ \ \rightarrow \ \ \{k,l\}$$

The function $\tau_{(p)}$ solely depends on the step p and decreases as the process goes on, thus gradually reducing the height of the ball, or equivalently, the intensity of the learning action.

4. This closes the step p, that is incremented by 1 and the cycle is repeated.

Both functions $\eta_{(p)}$ and $\tau_{(p)}$ gradually reduce the impact of the learning process and eventually froze it, thus leaving the network in its learned state. From a biological point of view, they may be perceived as neurotransmitters, acting on the whole network at once. They are exhibited in following Figure 12. In this scheme, the unsupervised character of the process emanates from the random selection of the stimuli. The competition arises from the selection of an elected neuron (step 1), which then forces its neighbor neurons located in the activity ball (step 2) to imitate the stimulus presented (step 3)[4]. Figure 5 shows the topological impact of this action on the synaptic space.

All these operations will be repeated and extended in the next section, with data emanating from the real world, in stimuli spaces whose dimensions will be larger than two. Furthermore, the topology of the network will be morphed from a square to a torus.

[4]The religious correspondence to all of this may become clear to any vigilant reader. In this constellation, a prophet or a pundit corresponds to the elected neuron. He - or she - takes the lead and forces his - or her - disciples to endorses his - or her - teaching. People who do not belong to the circle (for us the activity ball) are ignored or expelled. Unmistakably, the units are no longer neurons in this dared interpretation, but people. The semantic space becomes the ideological mindset prevailing in this society.

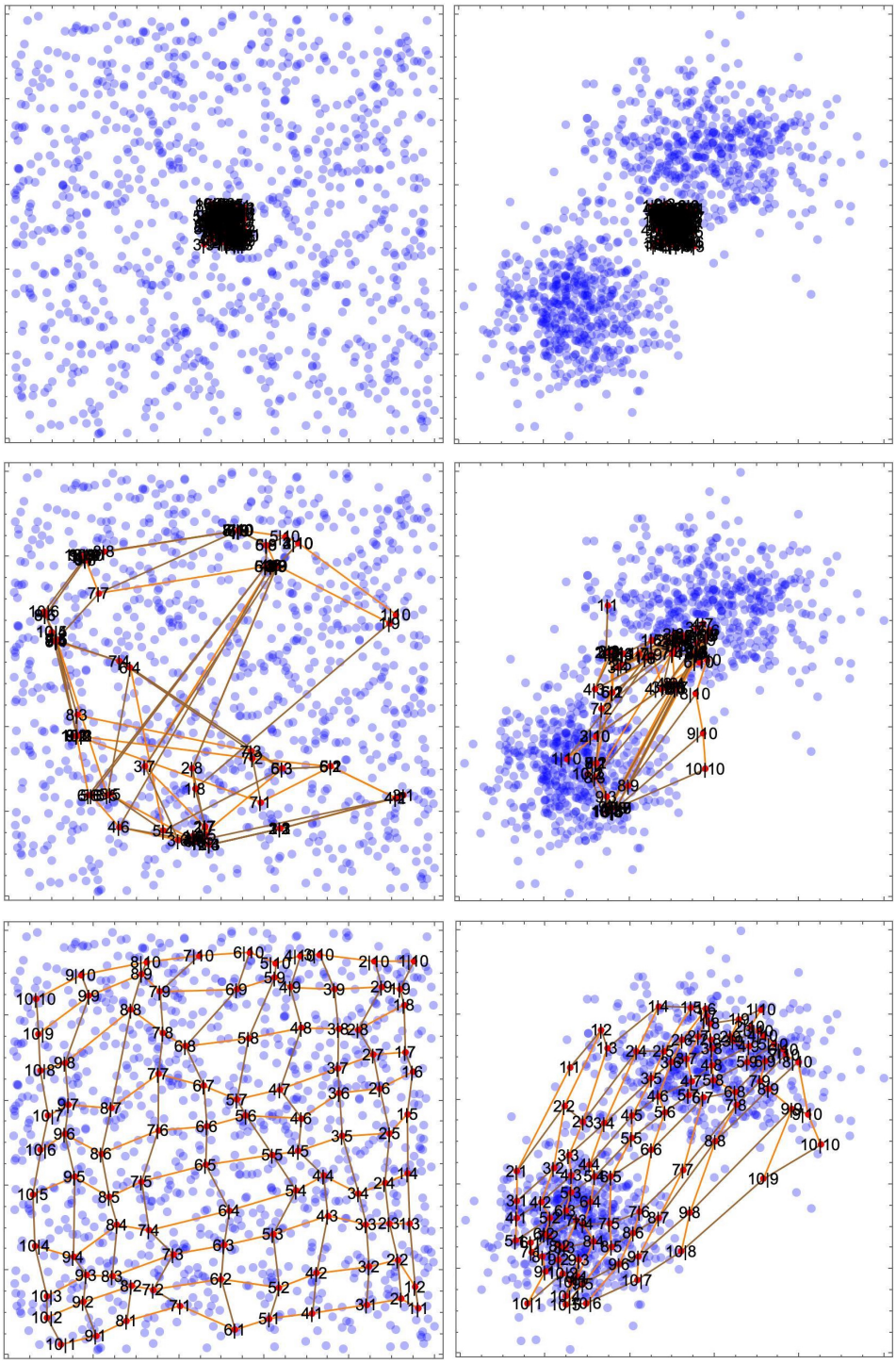


Figure 5: *Sketch of the temporal evolution of the learning process. 2 synapses network*

2.2 Assignment and Decision, first approach.

Deciding consist in identifying an yet unknown stimulus presented to the network, and consecutively attributing it an identity. "Unknown" means that this stimulus was not present in the learning task submitted to the network. The decision occurs according to the rule applied in the step 1 of the previous learning cycle, repeated below. To the unknown stimulus $\tilde{s}$, determined by its coordinates $(\tilde{x}, \tilde{y})$, is associated the closest neuron $\{i,j\}_{selected}$, selected according to the euclidian metric $\mathcal{D}_e$:

$$\min_{\{k,l\}\in\mathcal{N}} \mathcal{D}_e[(\tilde{x},\tilde{y}),\{k,l\}] \;\; \rightarrow \;\; \{i,j\}_{selected}$$

The notion of "domain on influence" of the $\{i,j\}_{selected}$ neuron takes its sense: all unknown stimuli falling in its corresponding Voronoï tile, or district, receive the traits potentially attributed to this neuron. This notion will be further discussed and applied in section 4.1.

3 Extension, complexification

After this initial presentation, we increase the complexity of the project in four directions:

Firstly, we start to work with a vastly more complex and realistic dataset, related to the characteristics of a bunch of western eurasian countries described by a array of more than twenty geographical, societal, governmental and financial parameters.

At the same time, we modify the topology of the neural network, from a square to a torus, with the purpose of avoiding the existence of boundary neurons on the square. Located on a torus, each neuron will be connected to the full set of its neighbors, thus cancelling any boundary effect from the network on the learning process.

A standard algebraic scheme called principal components analysis, (an application of a broader, more abstract theory called spectral analysis), will enable us to project the synoptic space onto our familiar three dimensional space. This projection will enable us to better explain the relationship between synaptic and semantic space.

Finally, extending the considerations developed in the previous section related to imbalance between frequently and infrequently visited areas in the synaptic space, we develop two measures, the first one entropic in its

essence, the second geometric, both suited to measure the quality of the learning process.

3.1 Extended Dataset

Formulated in the Wolfram Language, the Socioeconomic and Demographic Data provides seamless access to the curated and continuously updated Wolfram Knowledge base, which includes a wide range of types of socioeconomic and demographic data. Free-form linguistics provides a convenient mechanism for accessing all available data; more common categories also have specific associated Wolfram Language functions. The data presented in this work was collected from this source in April 2023. The 27 countries and 24 parameters are described in Annex 7.2. In this new situation, the dataset X is a 27×24 matrix with coefficients in $\mathcal{R}$.

In our neural parlance, the countries are now dubbed the "items", the "stimuli", or the "input". The geo-political parameters are named "properties" or "characteristics", according to the thesaurus provided in the annex. They build a relational matrix that may be pictured as a bipartite graph, or prosaically as an excel table. They are going to be encoded in the synaptic weights of the neurons.

The crafty, holistic representation of this relational matrix as a bipartite graph is presented in Figure 6. It is realized on the Poincare disk[5]. The vertices connect countries and properties. Colors code the intensity of a parameter for a given country. An arc is bluish if the corresponding parameter is relevant to the connected country, reddish if it is not.

The Poincare disk is a representation of the 2-dimensional hyperbolic space. The arcs are the geodesics in this space and the boundary circle of the disk is the set of the points located at infinity. This hyperbolic structure builds the groundwork of the M.C. Esher disks[6]. I adopted it solely in accordance to my aesthetic inclination.

As one notices, no arrow is connecting two countries on the left side, or two properties on the right side, thus the name "bipartite graph". An

[5] wikipedia.org_wiki_Poincare_disk_model

[6] web.colby.edu/thegeometricviewpoint/2016/12/21/tessellations-of-the-hyperbolic-plane-and-m-c-escher/

aim of the present work consist in detecting such secluded connections, or homologies between countries, making to this purpose use of the information transported backwards from the right to the left side of the graph through the learning process. Last, but not the least, the choice of the dataset is agnostic: no religion, ideology, politics or connection to the European Union has been taken into consideration here.

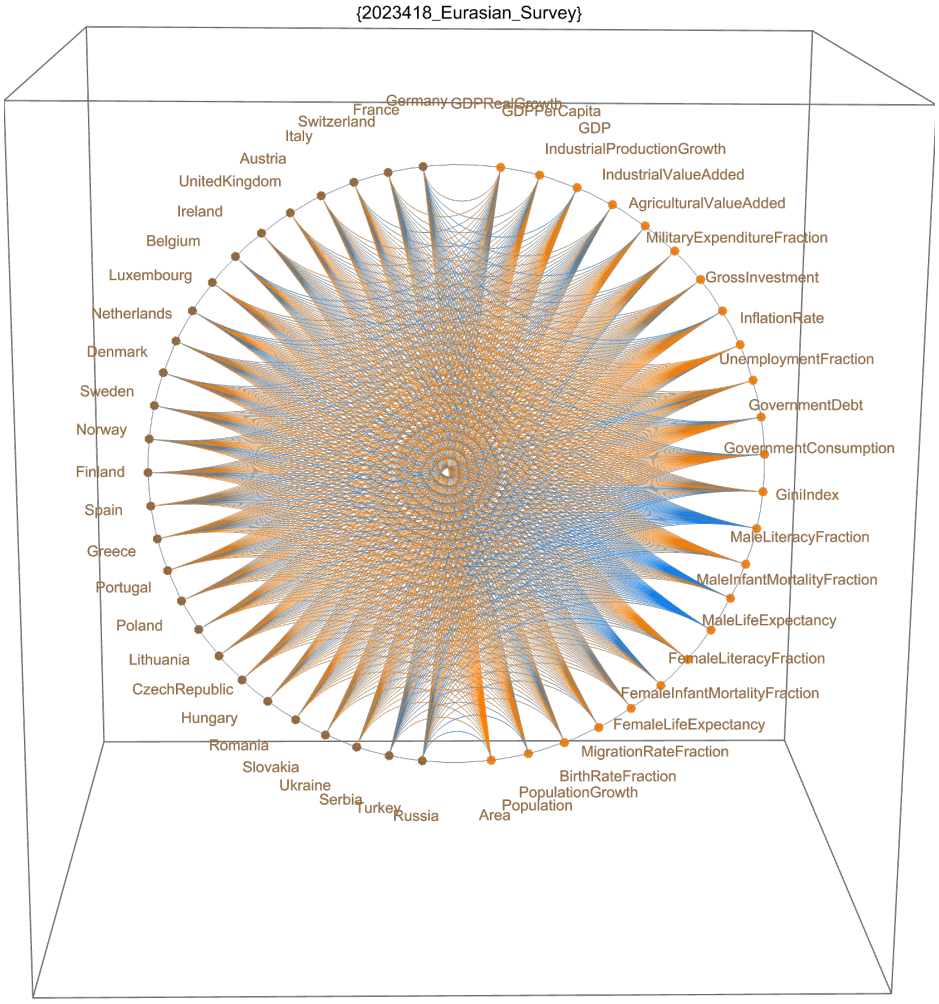


Figure 6: *Holistic representation of the Eurasian dataset as a bipartite graph drawn on a Poincare disk. This dataset was collected in April 2023.*

3.2 Toroidal network

We modify the topology of the neural network from a square to a torus, with the purpose of avoiding the existence of boundary neurons on the square. Located on a torus, each neuron will be connected to the full set of its neighbors, thus cancelling boundary effects induced on the learning process.

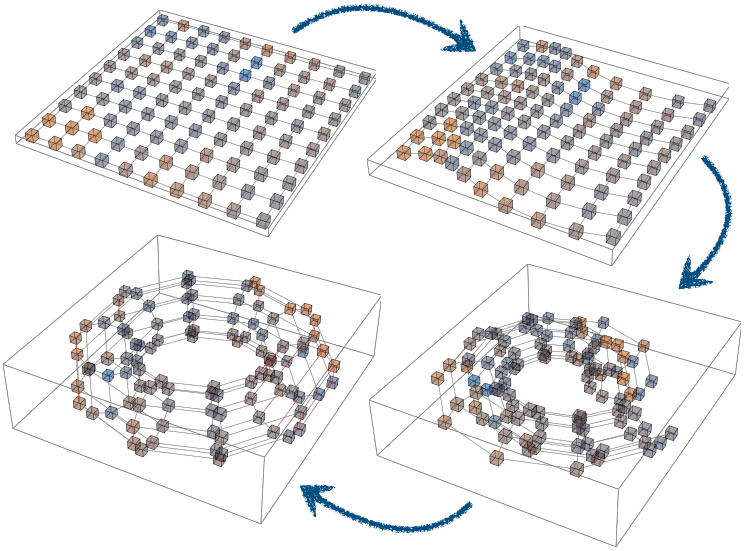


Figure 7: *From a square to a toroidal network.*

This operation is algebraic in its essence, the modular metric δ_{tore} being substituted to the euclidian metric δ_{eucl} initially defined on the square network in Figure 2. $\{i_1, j_1\}$ and $\{i_2, j_2\}$ define the locations of two neurons on the network, the size of which is $[1, i_{max}] \times [1, j_{max}]$. The distance between the two neurons is then provided by:

$$
\begin{aligned}
\delta_{eucl}[i_1, j_1, i_2, j_2] &= \sqrt{[(i_1 - i_2)^2 + (j_1 - j_2)^2}\ ; euclidian \\
\delta_{tore}[i_1, j_1, i_2, j_2] &= \sqrt{{Min_i}^2 + {Min_j}^2}\ ; toroidal \\
with &\rightarrow \\
Min_i &= Min[Abs[i_1 + i_{max} - i_2], Abs[i_2 - i_1], Abs[i_1 - i_{max} - i_2]] \\
Min_j &= Min[Abs[j_1 + j_{max} - j_2], Abs[j_2 - j_1], Abs[j_1 - j_{max} - j_2]]
\end{aligned}
$$

One notices on the lower panels of following Figure 8 that both vertical and horizontal boundaries of the network are "glued" together. This enables the boundary-free neighbor relations among neurons induced by the topology of the torus. The upper panels present the euclidian metric on the square as a reference. On the lower panels, the toroidal distances are computed from the most bluish neurons and the brightest spots may be considered as their respective "antipodal" locations on this fanciful toroidal planet[7]. On the upper panels, euclidian distances are computed from the most bluish neurons. Last but not the least, had we returned the terms in the $Abs[i_1 + i_{max} - i_2]$ in the Min_i or Min_j expressions, then the network would had taken the shape of a Klein Bottle.

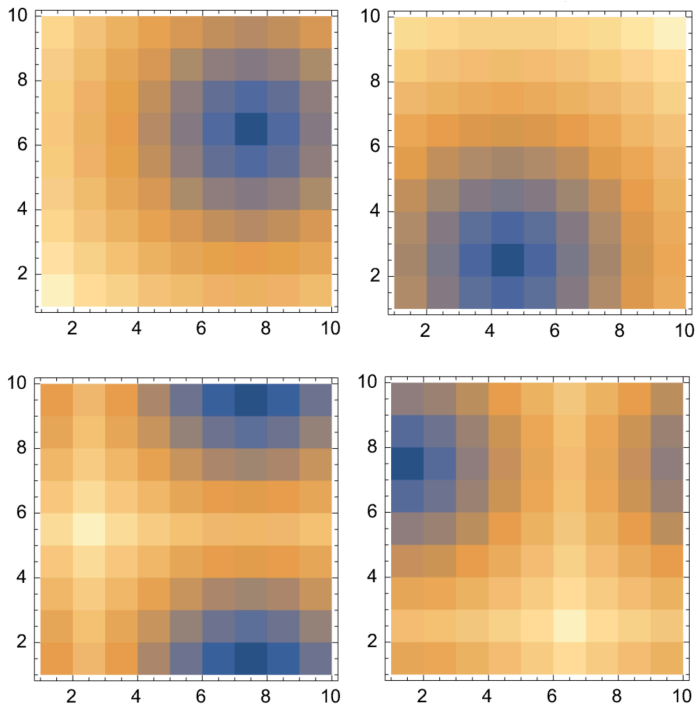


Figure 8: *Top panels, euclidian distances. Lower panels, toroidal distances.*

[7]New Zealand and Switzerland are almost antipodal to each other on spherical planet earth.

3.3 Integration and first computation

The neural network, now equipped with the structures just described, and the learning process are operated on the dataset described in the section 3.1. The learning sequence is exactly the same as presented in section 2.1: the algorithm and the corresponding software are unchanged, now working on a toroidal network whose neurons are furnished with 24 instead of 2 synapses. To the Figure 2 corresponds now the following Figure 9 with the network on the left and the synaptic or semantic space on the right. Both panels deserve accurate explanation.

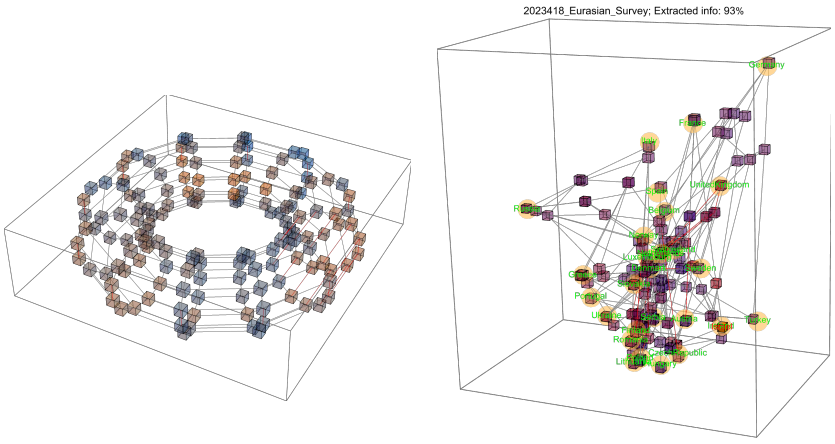


Figure 9: *Left, Toroidal Neural Network. Right, synaptic, or semantic Space. Two distance are implemented, the first one, designed by* δ_{tore}*, operates on the network. The second one,* $\mathcal{D}_m$*, operates on the synaptic, or semantic space. As mentioned in the text, this figure is the exact correspondent of the Figure 2.*

Considering the network first (left panel) and recalling the discussion related to the attribution decision in section 2.2, we now expect to attribute an

identity to each neuron, provided by the closest item to its synaptic location in the synaptic space, as provided by the $\mathcal{D}_m$ metric.
In our case, as items are countries, neurons are labelled as such and the "districts" evoked in previous section 2 may now be pictured on Figure 10 below.

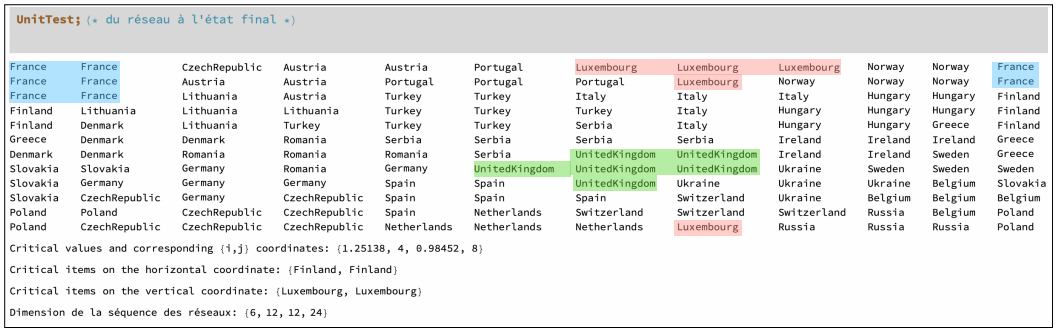


Figure 10: *Topological gathering on the network. Few districts are colored: United Kingdom in green. France in blue and Luxemburg in pink illustrate the toroidal geometry.*

Another example of topological gathering, applied on meteorological fields, is presented in the references [4 and 5]. What about the synaptic, or semantic space pictured in the right panel of Figure 10? Let us start with an analog description.

In quantum physics, a physical item, as for example an atom or a molecule, is described by a mathematical object named the *Hamiltonian*, describing how the energy is distributed in the item. Being called an *auto adjoint operator* in algebra, the hamiltonian has the faculty to generate a basis in the quantum space in which the item, atom or molecule, lays. The elements of this basis, called eigenvectors of the auto-adjoint operator, describe the eigenstates of the physical item (the oscillating modes of the molecule). To them correspond eigenvalues prescribing the amount of energy attributed to each eigenstate (perceived as the intensity of the oscillation). They are called the energy levels attributed to the eigenstates of our molecule. All of this corresponds to a standard algebraic scheme in data science named *Principal Components Analysis*. Applied here, it enables us to project the representation of the neural network, located in the $\mathcal{R}^n$ synaptic space, onto our

familiar three dimensional space.

Let $\mathcal{Y}$ be the $N \times n$ matrix of $N = imax \times jmax$ neurons, each furnished with n synapses. (Each neuron is described as a vector $\bar{n}$ of dimension 24 (synapses per neuron in our case)). Then the variance covariance matrix $\Sigma = \mathcal{Y}^T\mathcal{Y}$ is symmetric and happens to be our auto-adjoint operator. The projection of the $\bar{n}$ synaptic vectors on three eigenvectors of Σ delivers the three dimensional representation of the network presented in the right panel of the Figure 10. But, how do we choose those few three among the 24 eigenvectors? In ordering their corresponding eigenvalues and and selecting the three largest.

Eigenvalues quantify energy levels in quantum physics, in our field, data science, they quantify information. The miniature text featured on the top of the right panel in Figure 11 reads: "*202211_Eurasian_Geopolitics.* ***Extracted Information:*** 93%". This means that 93% of the information available in the representation of the network in dimension 24 is carried by the three first eigenvectors. The 7% of information carried by the following eigenvectors remains secluded in the higher dimensions of the semantic space and is simply disregarded.

3.4 Entropic & geometrical quality assessment

Further developing the considerations related to imbalance between frequently and infrequently visited areas of the synaptic space, we develop two measures, the first one entropic in its essence, the second geometric, both suited to measure the quality of the learning process.

We start with the entropic approach, based on the famous expression for the information entropy formulated in 1948 by Claude Shannon:

$$\mathcal{H} \;=\; -\sum_{i,j=1,1}^{imax,\,jmax} p_{(i,j)}\; Log(p_{(i,j)})$$

Let we name $N = imax\, jmax$ the number of neurons on the network $\mathcal{N}$, indeed its size. $v_{(i,j)}$ the number of visits to the neuron $\{i,j\}$ during the learning process of m steps. One has $m = \sum_{i,j} v_{(i,j)}$, then $p_{(i,j)} = \frac{v_{(i,j)}}{m}$, $\sum_{i,j} p_{(i,j)} = 1$ and $\forall \{i,j\} \in \mathcal{N} \;:\; p_{(i,j)} \leq 1$. Thus, $\mathcal{H} \geq 0$, all of this as

expected.

The maximal entropy $\mathcal{H}_{max}$ corresponds to a state of the learning process in which all the neurons have received an equal number of visits. This is expressed as $\forall\{i,j\} \in \mathcal{N} : \ v_{(i,j)} = \frac{m}{N} = constant$ and implies that $p_{(i,j)} = \frac{v_{(i,j)}}{m} = \frac{1}{N}$. Hence $\mathcal{H}_{max} = -\sum_N \frac{1}{N}\ Log(\frac{1}{N}) = +\ Log(N)$. Finally, choosing N as the logarithmic base, one gets $\mathcal{H}_{max} = Log_N(N) = 1$, thus normalizing the entropic measure:

$$\mathcal{H} \ = \ - \sum_{\{i,j\}\in\mathcal{N}} p_{(i,j)}\ Log_N(p_{(i,j)})$$

This definition is coherent with the assertion presented earlier, repeated below, and enables the assessment of a learning process.

$$\forall\{i,j\} \in \mathcal{N} \ : \ \int_{V_{\{i,j\}}} prob(\Psi \in d\sigma)\, d\sigma = constant$$

As already mentioned, the size of each $V_{\{i,j\}}$ domain is the tuning parameter in the previous expression. This property, that appears on the right panel of Figure 4, is still valid in the present situation. However, the hyper-Voronoï cells exist in the full dimensional space and I do not know how to project them onto our familiar three dimensional space[8].

Entropy is named an *extensive* quantity in classical thermodynamics. It is proportional to the size of the object or the amount of stuff under consideration[9]. As for example, in astrophysics, the entropy of a black hole is proportional to the area of its cosmological horizon. In our modest case, the entropy of the neural network is proportional to its size, given by the number $N = imax\, jmax$ of its neurons, in consistency with the formal definition.

The geometric approach is straightforward. It consists in collecting at each learning step p the distance $\mathcal{D}[(\tilde{x},\tilde{y}),\{k,l\}]$ in the synaptic space between the stimulus presented and the elected neuron, as already given by the expression in subsection 2.2:

[8]Mathematicians and among them primarily topologists know that in higher dimensions, the volume of a sphere is mostly located near its surface. For us, this means that our hyper-Voronoï cells form in the synaptic space a kind of foam in which the information is indeed located.

[9]Contrarily, temperature is an *intensive* thermodynamical quantity.)

$$\min_{\{k,l\}\in\mathcal{N}} \mathcal{D}_m[(\tilde{x},\tilde{y}),\{k,l\}] \quad \rightarrow \quad \{i,j\}_{selected}$$

Everything can now be collected in a synthetic view pictured in following Figure 12. The learning procedure described in section 2.1 is operated on the toroidal network, applied on the Eurasian dataSet pictured in Figure 6. The dynamical evolution of the network, as projected in the semantic space, is shown in the three lower panels at learning steps 0, 300 and 600. These panels correspond to those presented in Figure 5. The final state of the network is presented on Figure 11

Finally, the two decreasing curves are the graphs of the functions $\eta_{(p)}$ (blue curve) and $\tau_{(p)}$ (gold curve) introduced in section 2.1 and identified as a kind of numerical neurotransmitters.

It is worth mentioning that the synchrone tuning of all the involved parameters is a little tricky, trial and error being required. However, after having reached an empirical optimum, one notices at the experience that the learning process happens to be quite resilient. In all this tuning business, I have followed - so far as I could - my favorite tenet consisting in aiming at the greatest possible simplicity, just beyond triviality.

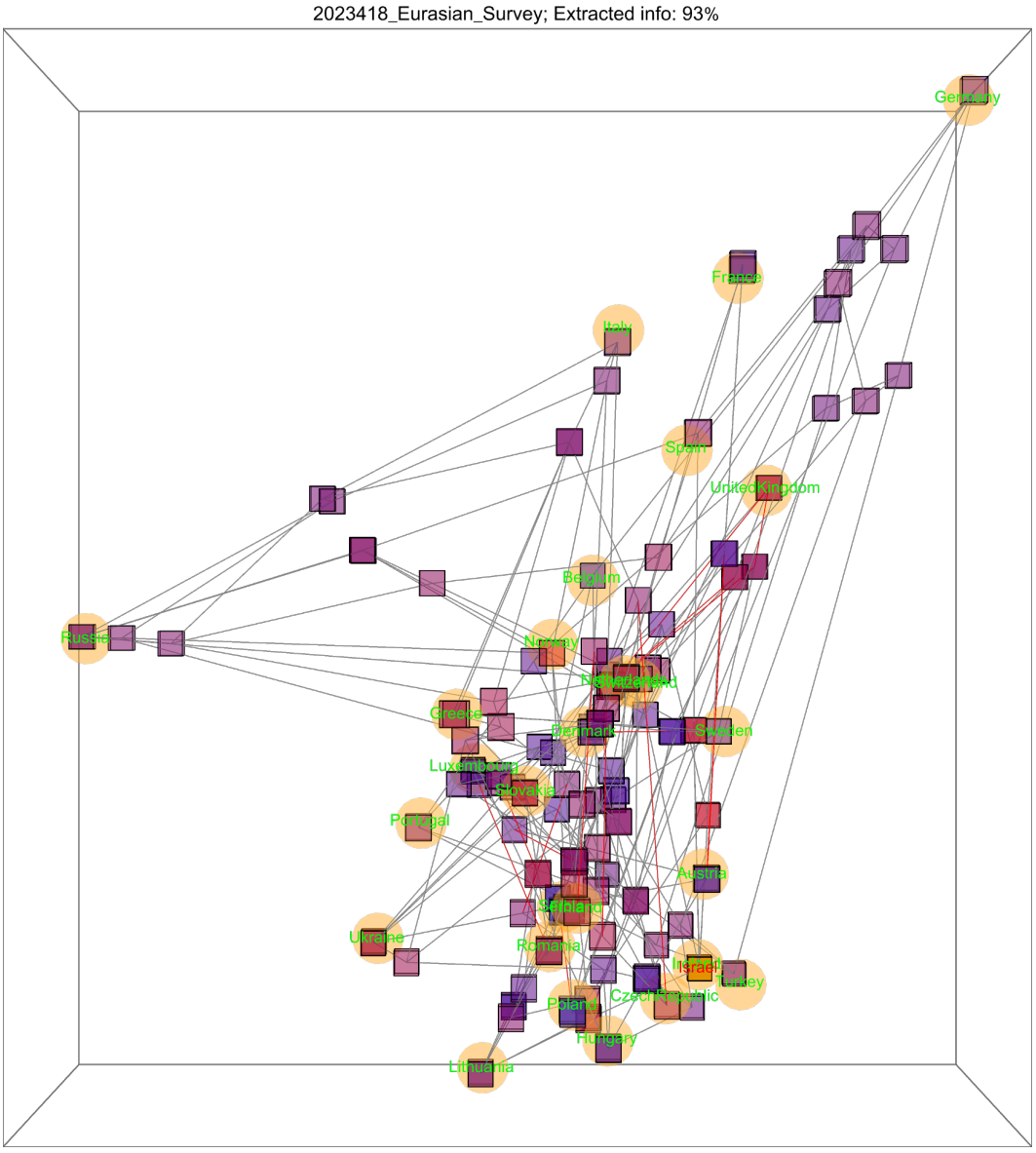


Figure 11: *Semantic space and grid. It ought to be put in relation with Figure 6. 93% of the information secluded in Figure 6 is projected in this 3-dimensional representation.*

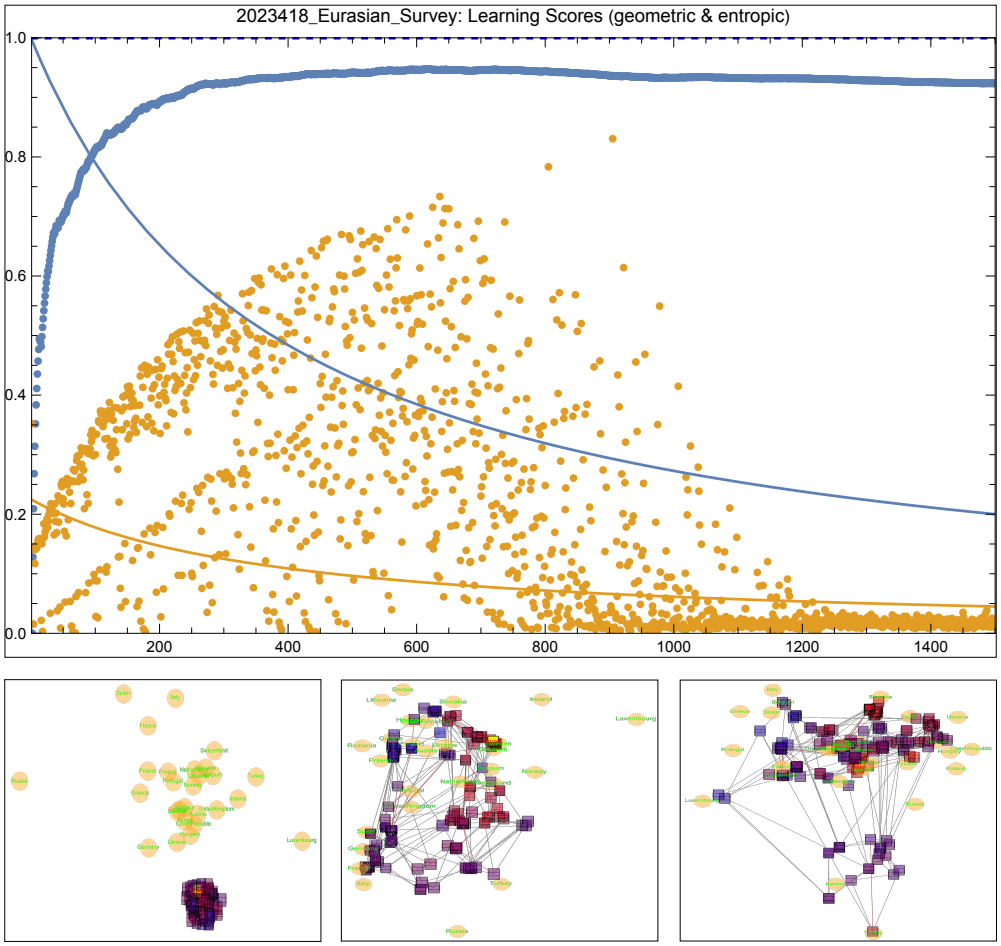


Figure 12: *Evolution of the scores as learning progresses. Abscissa: learning steps. Ordinate: learning scores. Blue curve: entropic, gold dots, geometric scores. The slight decreasing of the entropy by the end of the process is probably caused by some overfitting. Three lower panels: state of the network at learning steps 0, 300 and 600. These panels correspond to those presented in Figure 5.*

4 Stochasticity, Assignment, & neural crowds

We now extend the discussion initiated in Section 2.2, "Assignment and Decision", whereas the stochasticity of the learning process is now driving our reflexion. Unsupervised Competitive Learning is fundamentally stochastic. Indeed, the initial state of the network, at step zero, is randomly determined in the tiny initial domain. Then as the selection of the items presented to the network during the learning process occurs randomly, each new task delivers a new, unique network, indeed a genuine individual. Figure 13 shows two semantic networks consecutively generated by the same learning process and epitomizes, together with the network presented in figure 11, the stochasticity of the process. All three networks, or semantic spaces, are siblings.

4.1 Assignment & Neural crowds

Networks generated under equal preconditions are clones among themselves. Similar but not identical to their siblings, they deliver coherent projections and answers. Thus, instead of developing one single network, it is wiser to develop a crowd of them and use any kind of voting system in order to infer an answer provided by this virtual crowd instead of considering an unique answer that would be akin to "the winner takes it all" philosophy.

Indeed, although the learning process is demanding, the operation of the network is not. Taking advantage of this property, it would be easy for an engineer to implement a crowd of networks in a robot or any kind of operational device, together with a voting system. The robot would then take democratic decisions delivered by the crowd and behave accordingly, with its cognitive capability remaining secluded to its environment. Such a kind of decision making would provide it with a putatively subtlety enhanced behavior.

An example of assignment, or decision, is presented in Figure 14. Kazakhstan, that did not belong to the learning task presented to the network, is assimilated to Lithuania by the network. The elected neuron is located in position $\{4, 12\}$ on the upper panel, Then, the answer from the network to the stimulus is presented in the lower left panel. The selective ability of the network is presented in the lower right panel: the district associated to Sweden is well featured in red. Figure 15 exhibits the answer of four networks,

all trained under the same preconditions, to the task of assigning a known item (from the learning process), here a country, to the unknown item named "Kazakhstan". Three networks assimilate it to Sweden, one to Slovakia.

It may be noted here that no artificial intelligence is at work here, but solely sturdy algorithmic steered by a lot of stochasticity. However, the algorithm manages to mimic some intelligence when required to identify unknown items and assign an identity to them. This assignment realizes the expectation formulated in the title in mapping features on categories. One thinks here at Daniel C. Dennett's book "From Bacteria to Bach and Back" [6] in which he shows that smart algorithms, and even more crowds of them, may well exhibit some intelligent behavior without being by themself intelligent.

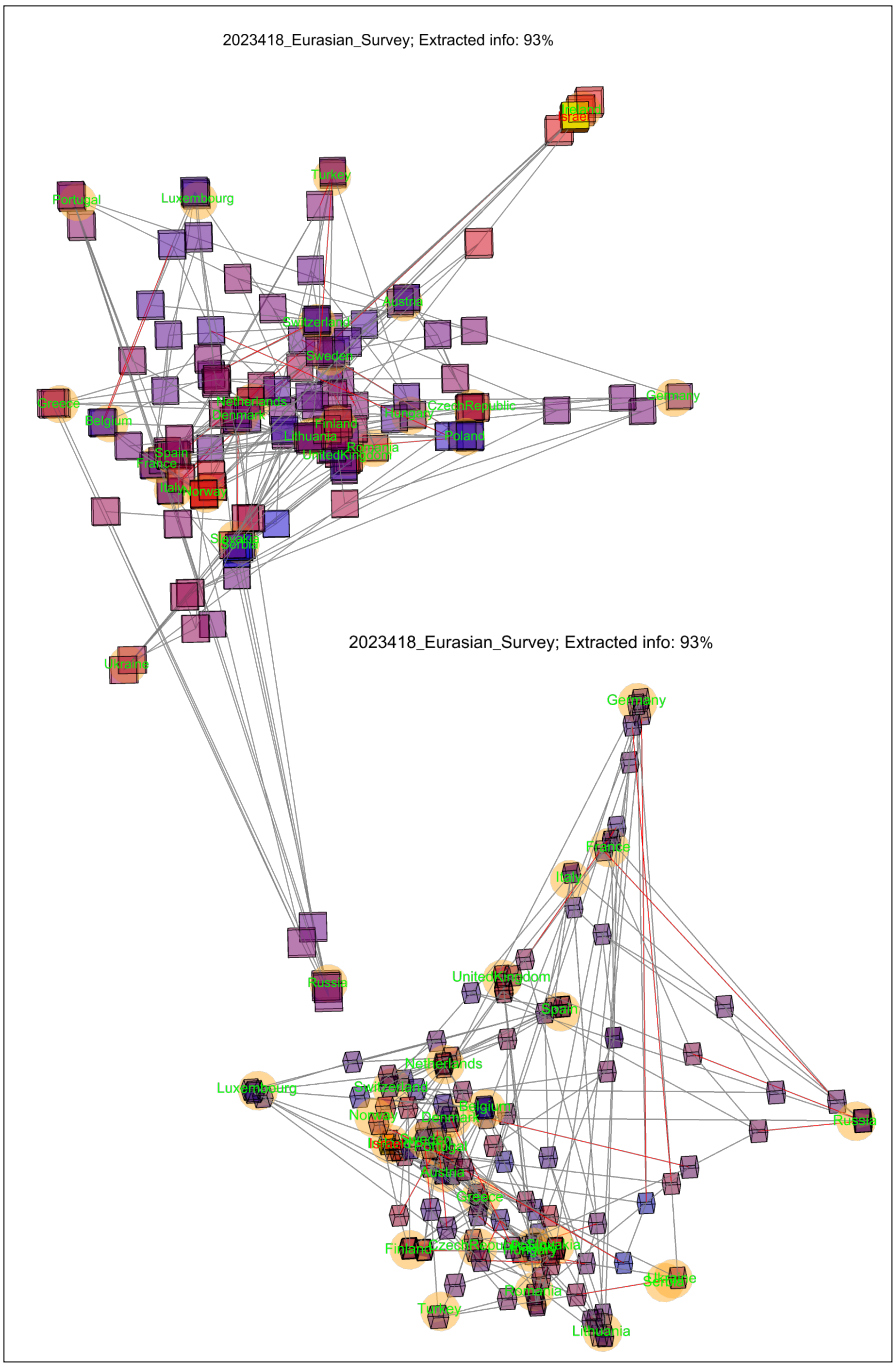


Figure 13: *Two instances of the learning process deliver two different semantic grids. Figure 11 shows another one.*

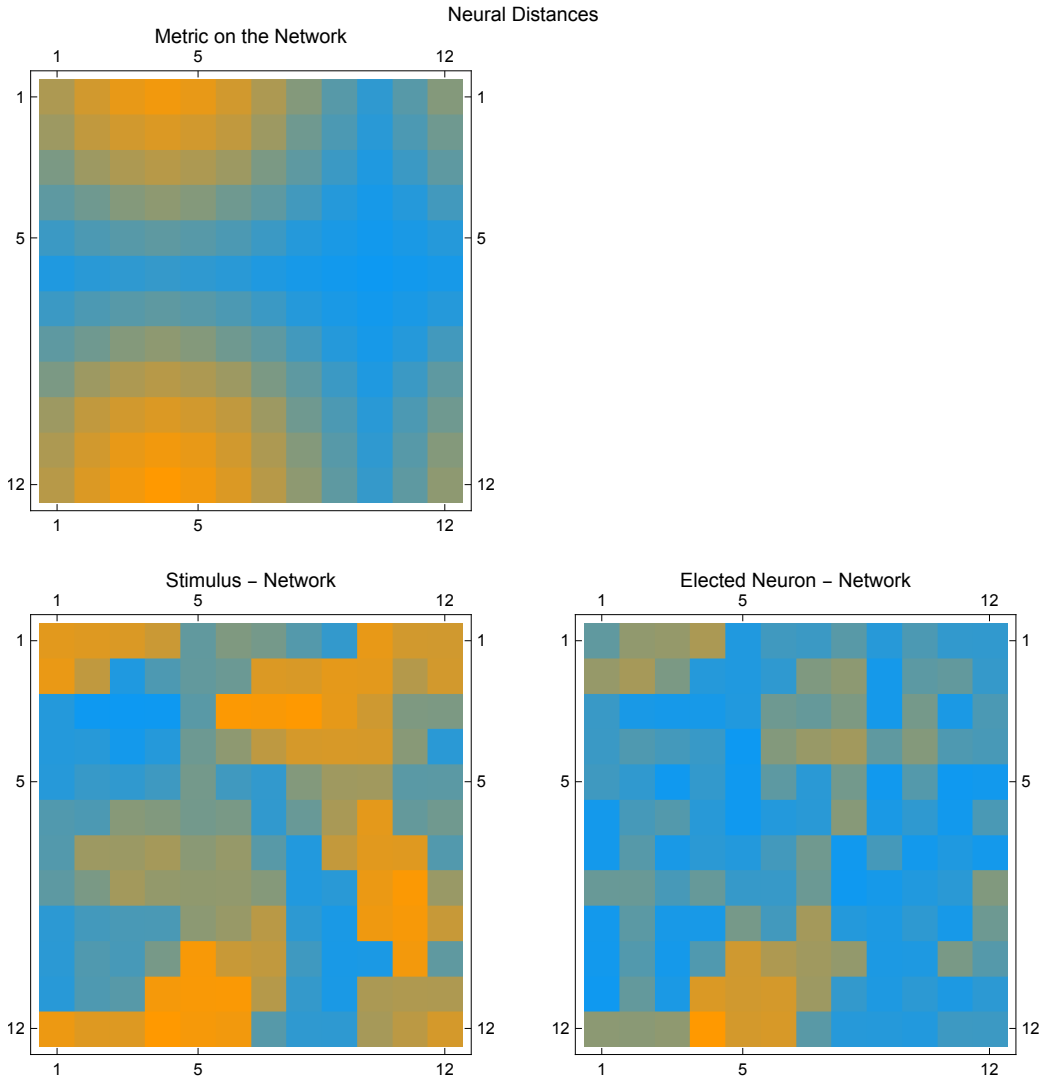


Figure 14: *Decision: Kazakhstan is assimilated to Lithuania by the network. The toroidal topology is well visible on each of the three panels. This Figure corresponds to the left panels of Figures 2 and 9.*

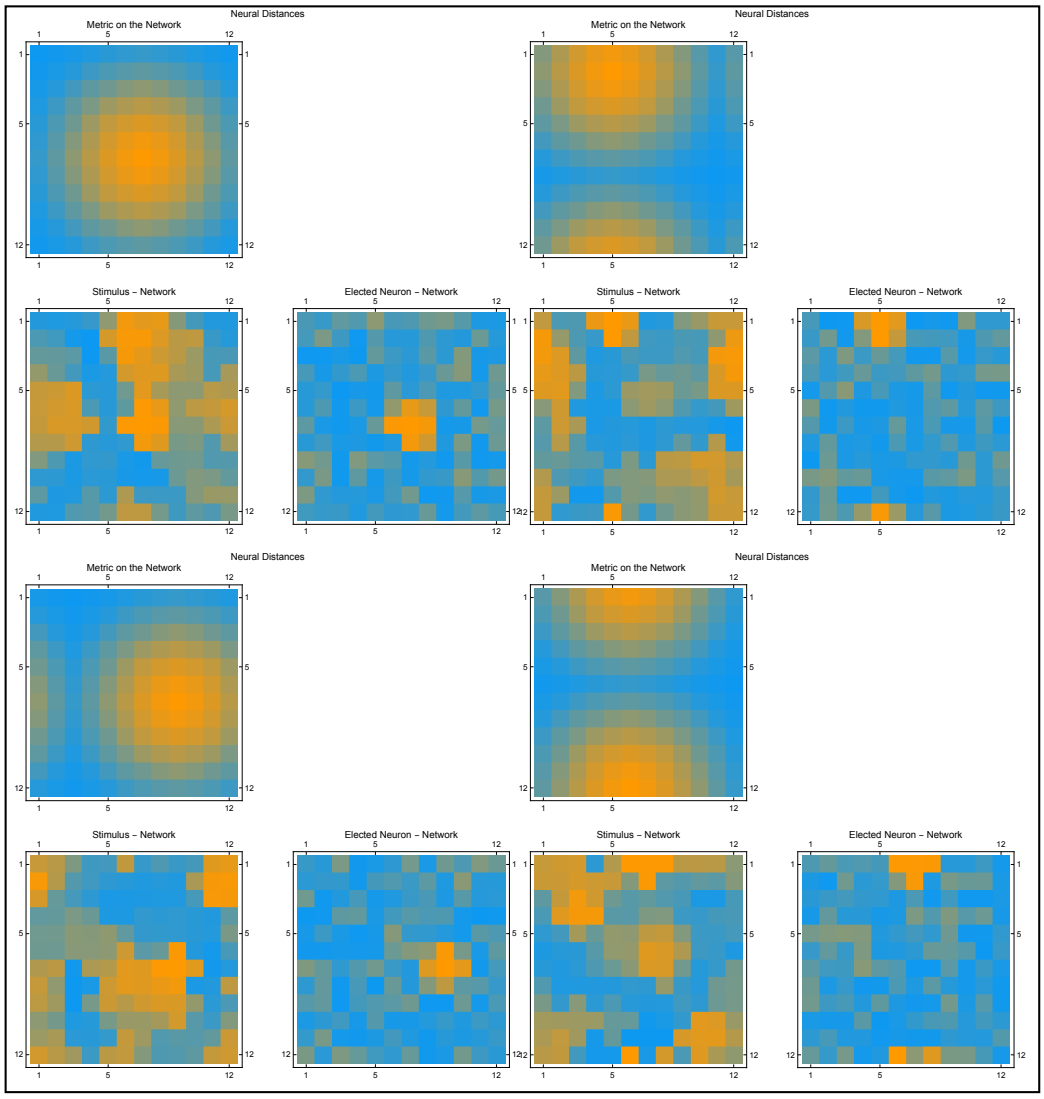


Figure 15: *Four clones of the network at work: Israel is assimilated four times to Ireland by these four networks. The four panels correspond to the left panels of Figures 2 and 9.*

5 Interpretation

Considering again Figure 6, one notices that we have dealt with the left part of the disk so far. The items, indeed the countries, have been featured, but their properties, distributed on the right side of the disk, remained hidden, coded in the synaptic weights of the network. Our next (last) action in this essay will now consist in making them visible. At this level of the work, this is solely an exercise in data visualization consisting in the following steps:

1. Establish a two dimensional picture by the projection of the previous 3 dimensional diagrams on their 1+2 , 1+3 or 2+3 eigenvectors (principal components). The edges of the networks will then no longer be pictured.

2. Select two parameters among the set of available parameters on the right side of the disk in Figure 6 (or, equivalently, from the annex 7.3).

3. Exhibit the neurons on the 2D projections as disks according to the following convention:
- The surface of the disk is proportional to the relevance of the first parameter.
- The color of the disk codes the second parameter with low relevance in red, high relevance in blue. Of course, a parameter can be selected twice, then either producing a large blue, or a tiny red disk.

All of this is programmed in an executable piece of Mathematica software that enables the selection of the projection and the parameters. Following four figures, drawn with this tool, exhibit few among many options. Obviously, the cluster of countries located right to the centre of the figure, together with the Scandinavian cluster, happens to be emerging as a stronghold in the western eurasian constellation.

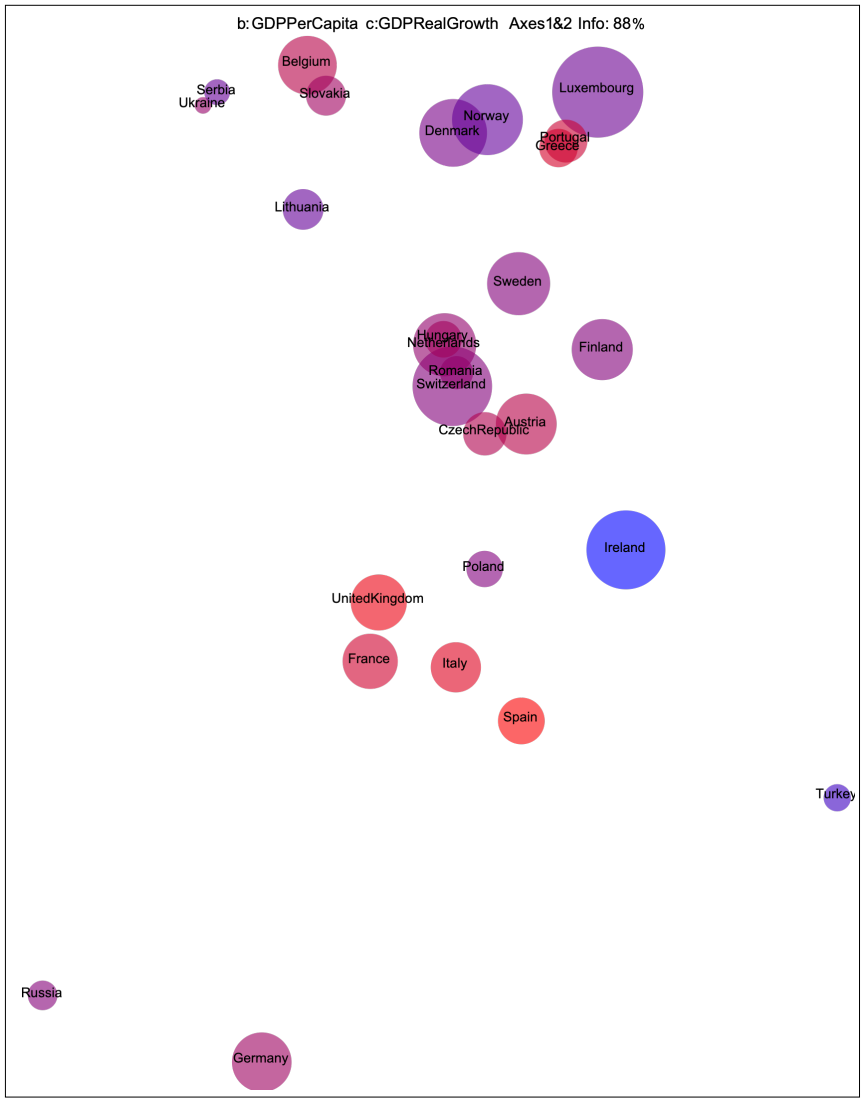


Figure 16: *Projection on the 1+2 axes. Information collected: 88%. Selected parameters: Gross Domestic Product per capita: size of the disk. GDP Real Growth: color. Similarly countries such as {United Kingdom, France, Italy, Spain} are located in corresponding clusters. Be aware of the fact that closely located countries may be distant in the third, here not presented dimension, as for example {Denmark, Norway and Luxemburg} versus {Portugal and Greece} in the upper cluster. Compare with the following Figures 18 and 19.*

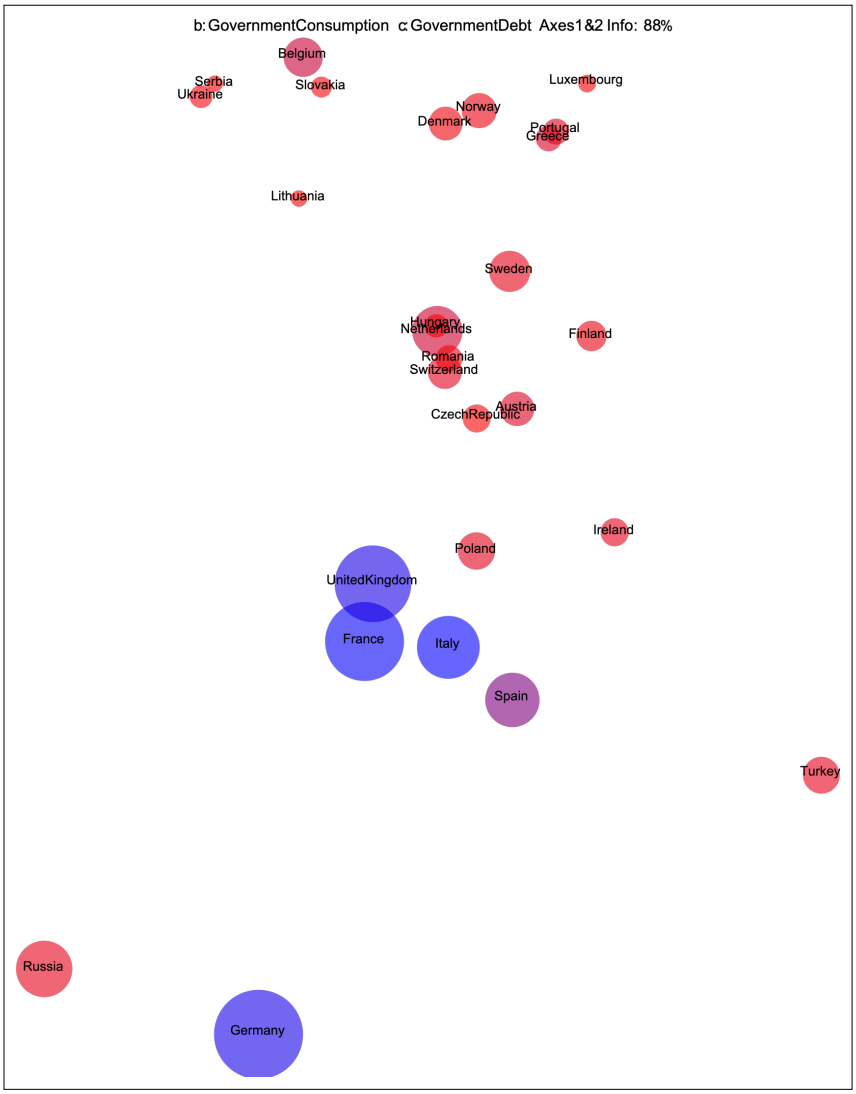


Figure 17: *Projection on the 1+2 axes, indeed the same projection as the previous figure. Information collected: 88%. Selected parameters: Government Consumption: size of the disk. Government Debt: color. One notices two clusters, the first one bluish with "old" western European countries below, and the second one reddish above, whose countries obviously implement other governmental procedures.*

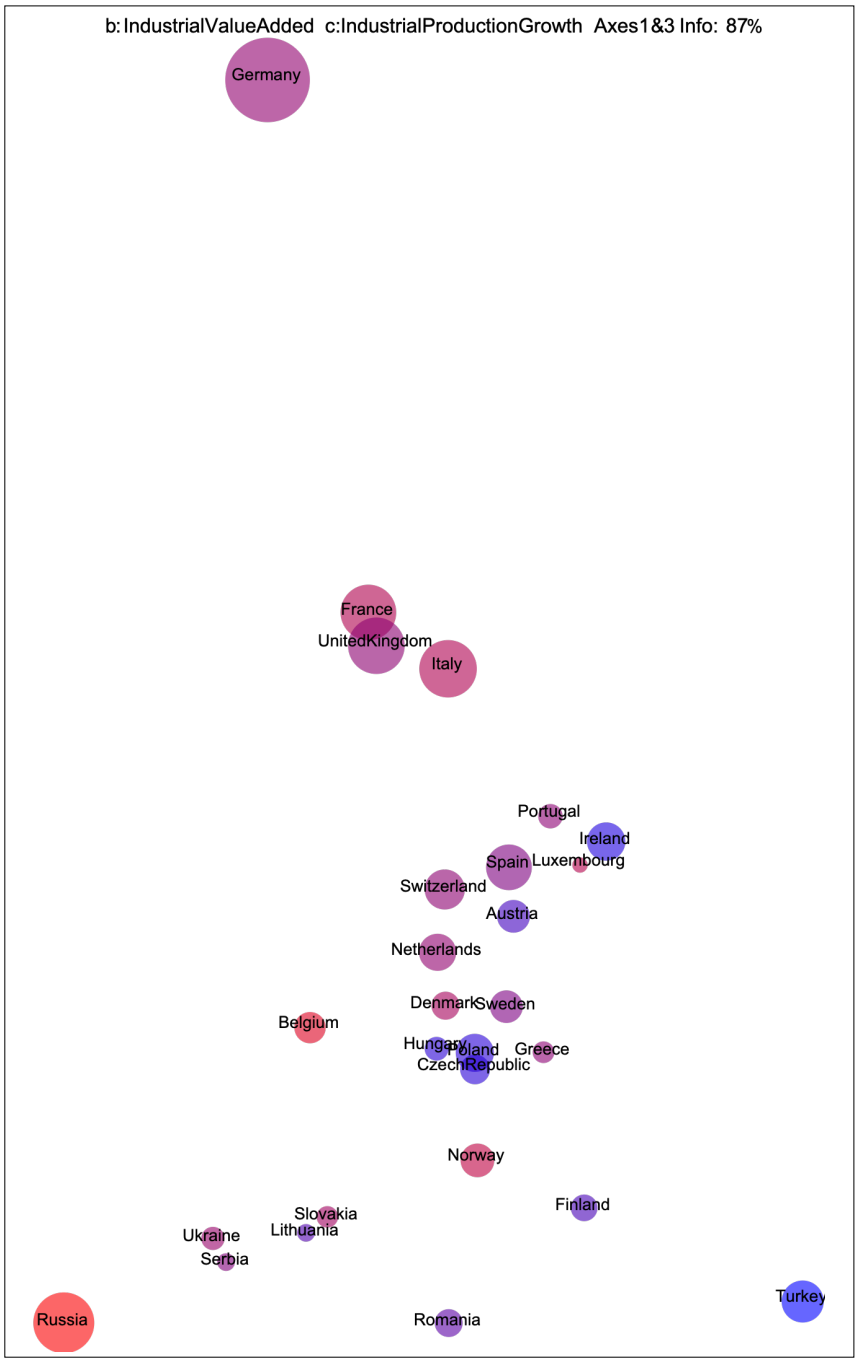


Figure 18: *Projection on the 1+3 axes. Information collected: 87%. Selected parameters: Industrial Value Added: size of the disk. Industrial Production Growth: color. {Ireland, Austria, Hungary, Poland, Czech Republic and Turkey }, to a less extend {Lithuania and Finland} are spotting momentum in this respect.*

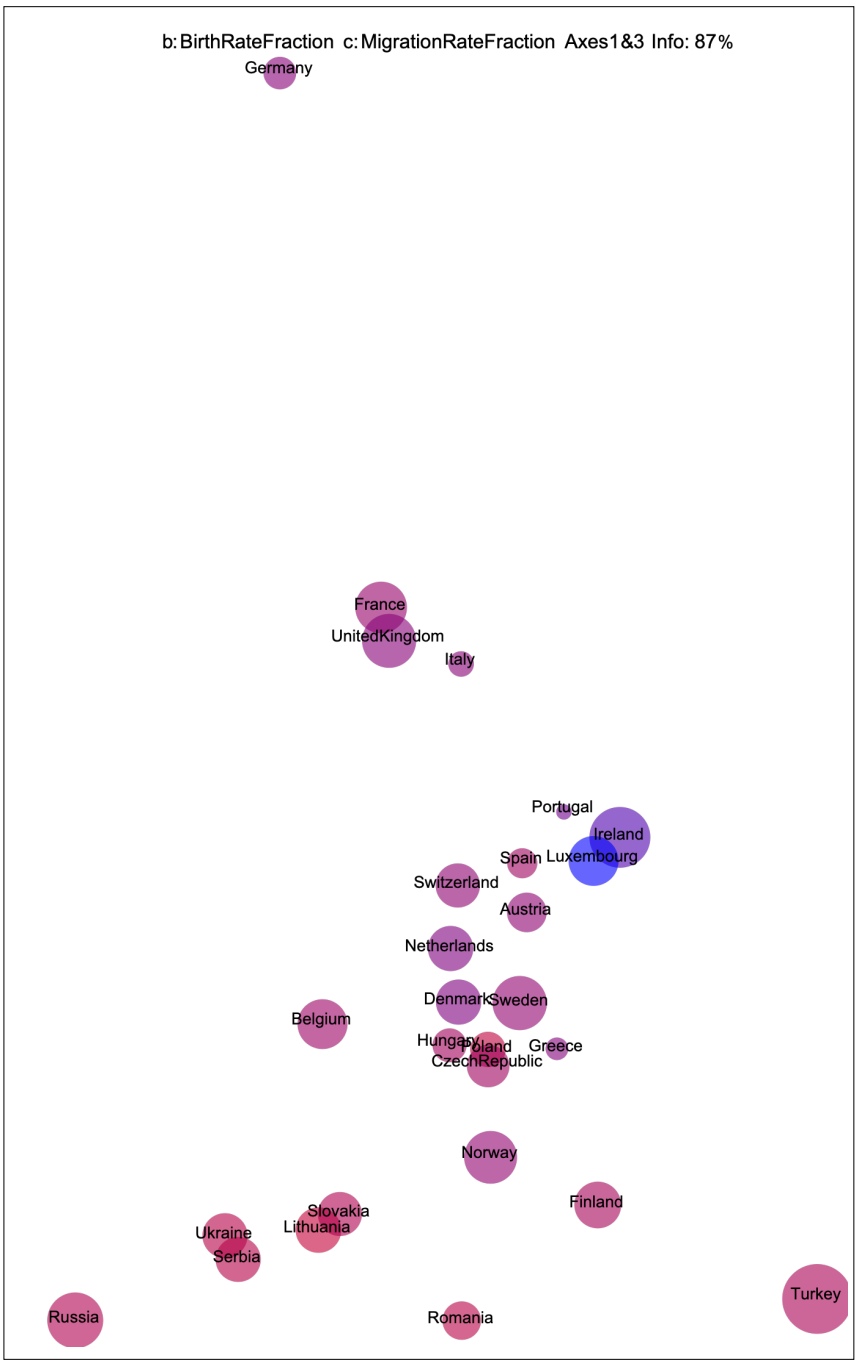


Figure 19: *Projection on the 1+3 axes, indeed the same projection as the previous figure. Information collected: 87%. Selected parameters: Birth Rate Fraction: size of the disk. Migration Rate Fraction: color. In all these representations, {Germany, Russia and Turkey} are rejected in the corners and happen to be seen as outliers within the eurasian crowd.*

6 Conclusion

Humbly considered, no artificial intelligence has been at work here, but solely sturdy algorithmic steered by a good deal of stochasticity. However, the USCL algorithm manages to mimic some intelligence when it is required to identify unknown items and assign an identity to them. This assignment realizes the expectation "Mapping features onto categories" formulated in the title.

Considered from a higher level, the overarching process presented in this essay bears the topology of a loop.

One starts with the contemplation of a yet informal object, in our case a bunch of western eurasian countries. This contemplation is given a formal and, as far as possible, rational structure by the use of well curated statistics. A first hint of arbitrariness is introduced by the choice of the parameters considered, according to appendix 7.2.

USCL, conceived to detect homogeneities and correlation among this statistical material, delivers a topological structure that we dubbed the syntactic or semantic grid, presented in Figures 9 and 11. Further two-dimensional projections presented in Figures 16 to 19 and the graphical implementation of parameters make the information secluded in the statistical data accessible to the human eye, and brain.

At this point, an expert, or pundit, and his - her - brain, has to interpret the graphical features exhibited on the Figures, and he projects his - her - interpretation back onto the initial items, here the western eurasian countries. It is worth stressing that the interpretation occurs on the features projected by the semantic grid, and not on the initial items, the western eurasian countries.

This closes the loop. The countries are given attributes and properties, as well as relationships among themselves, that result from the interpretation by an intelligent pundit of the features exhibited on the semantic grid. Ergo, smart algorithms, and even more crowds of them, may well exhibit some intelligent behavior without being by themself intelligent. Solely the expertise of the pundit, tainted by her experience, inclination and subjectivity, injects some interpretative intelligence into the loop.

Two remarks are in order. The first one is philosophical in its essence, the second brings us back to bionics.

Categories have been a philosophical topic since Aristoteles, about 25 centuries ago. Leibniz and his monades, Kant, Hegel and many others have refined the categorical concept over all the centuries. In twentieth century, categorical thinking has reached mathematics, mainly through algebraic topology thanks to the seminal work by Saunders and Mac Lane. In our 21 century, Emily Riehl and Dominic Verity have raised the concept to unexpected abstraction with their $\infty - Category\ Theory$. Introductions to this subject are provided in references [7 and 8]. The categorical ideas presented in this essay have been pragmatic and lead us back to bionics via an aeronautical analogy.

Until the end of the 19th. century, aviation pioneers tried to imitate flapping bird flight. Starting in the 20th. century, engineers designed their aircraft by moving away from this paradigm. A century later, Airbuses fly over the oceans with 500 people on board, a feat unattainable by nature, but these planes are no longer conceived to perch on an electric wire, something any swallow accomplishes with ease.

A similar development is underway in the artificial intelligence sciences. While systems invented in the 20th century attempted to approximate human intelligence[10], thus following a bionic approach, current research often liberates itself from that reference.

Thus, we will increasingly be confronted to machines and algorithms that will exhibit autonomous and non-human learning capabilities. These machines will not only find their references in the real world through feelers, sensors and observation systems, but mostly in the virtual world of internet and the metaverse. Current artificial intelligent systems already access unfathomable amounts of data and are operated on gigantic computer farms. Any comparison between them and and the puny software presented here is purposeless.

[10] According for example to the "Turing Test."

7 Annexes

7.1 Meteorological Network

In its simplest architecture (Figure 1), the neural network consists of a square array of 12 x 12 neurons, each neuron equipped with a bundle of 200 synapses that are both information sensors and analogue storage elements. The synapses are projected from each neuron onto the vertices of a grid of 10 x 10 points on a map of the whole of Central Europe. The measured meteorological data on each of these vertices corresponds to an altitude of air pressure of 500 millibars (geopotential 500 hecto-pascals) and temperature at altitude of 850 HPa. Each neuron is equipped with $10 \times 10 \times 2 = 200$ synapses and the network registers a weather situation through a system of $12 \times 12 \times 200 = 28'800$ synapses. In practice, these data are provided by a numerical forecast model as a matrix of numbers. The network executed by Meteo Swiss, whose toroidal shape is more complex, and is described in section 3.2.

Figure 20 describes the pattern learned by neuron classified 2,5 on the network (2th. line, 5th. column on Figure 1). The general air circulation over Europe is represented, on the one hand, by the 500 hPa contour lines and, on the other hand, by the areas of temperature distribution, where blue and green represent cold regions, yellow and red warm regions. The map covers a zone extending from Portugal (lower left) to Poland (upper right), the sea coasts and national borders are outlined, with Switzerland a little to the right of the center. Coordinates are latitude and longitude. Meteorologically, a cyclonic depression over the Sicily can be seen, as well as the general north-easterly flow north of the Alps and the outskirt of a High located over the North Atlantic. On the lower panel, the synaptic field of the same neuron 2,5 is shown in three dimensions. Coordinates are latitude and longitude. The vertical axis represents the geopotential height of the pressure field of 500 hPa expressed in hectometers. The wind field is computed from the geopotential field with the cyclonic depression over Sicilia. The temperature distribution and pressure level are indicated by the color and height of the arrows at each location. The basic characteristic of this representation is that it is not a meteorological map, but a description of the network's perception of such a situation. This is another encounter of the dual concept of the synaptic $\leftrightarrow$ semantic space. Neural based assessment systems based

on ECMWF[11] ensemble forecasts were implemented in the operation suites of Meteo Swiss by the end of the century and have been routinely operated since then. These algorithms are mainly used to assess the predictability of ensemble numerical weather forecasts. Consult reference [4] for more information.

What happens to the people who are confronted with such systems? While in the weather services of the past meteorologists worked in the production chain of weather information, from now on they are often placed above this chain and monitor it. A similar transition took place in ancient shipping: On a galley, propulsion was provided by the oarsmen. On sailing ships, however, sailors rig the sails and the wind propels their ship. Applied to our subject, where the intelligent work, rather than the power, is done by the machines, this shift triggers a humbling question: "either my work was intelligent, and the machine that replaces me must be at least as intelligent as I was, or else the machine is stupid, but then how was my work...? ". The reality, however, is by no means so Manicheean; in fact, meteorologists remain valuable advisors to their clients. However, their role has changed and requires outstanding communication skills.

[11] European Center for Medium Range Weather Forecast.

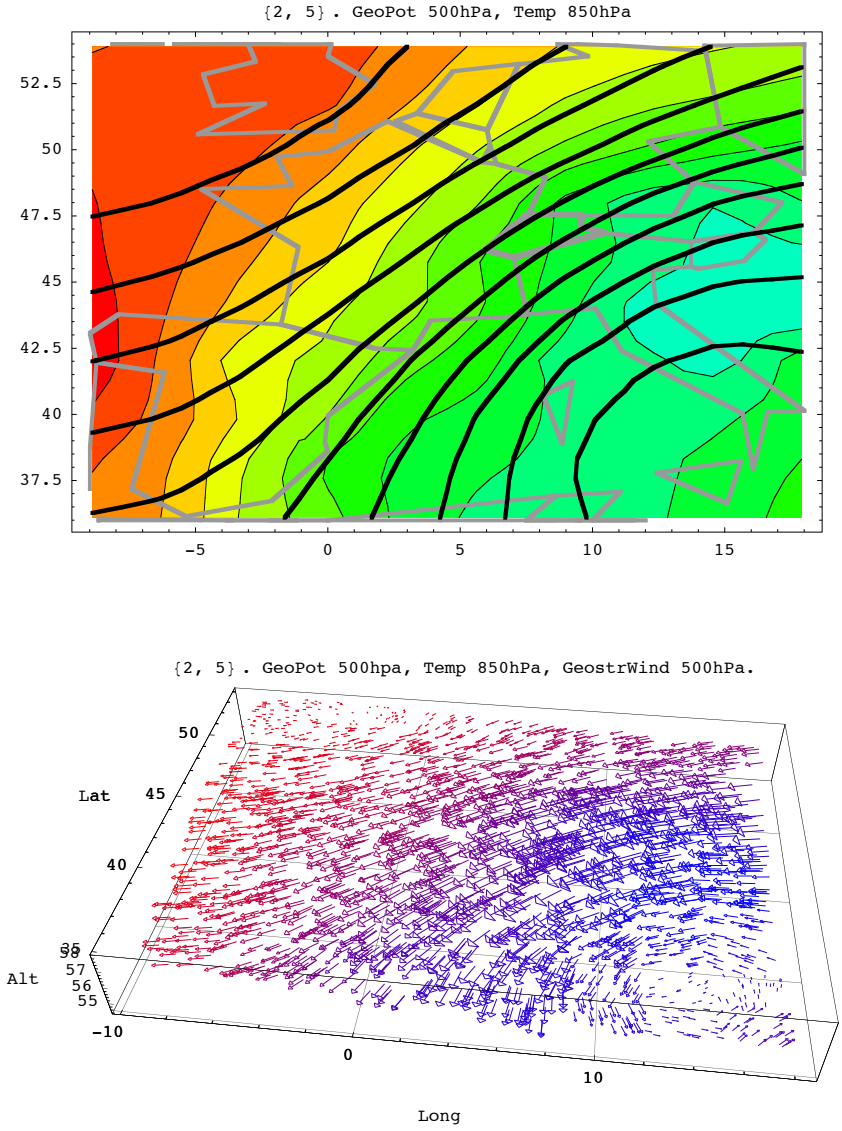


Figure 20: *Upper panel: On this neuron, classified 2,5 (2th. line, 5th. column in Figure 1), the weather circulation over Europe is represented, on the one hand, by the 500 hPa contour black lines and the temperature (red warm, green cold). Lower panel Synaptic field of the same neuron 2,5 in three dimensions. Coordinates are latitude and longitude. The vertical axis represents the geopotential field at 500 hPa expressed in hectometers.*

7.2 Geopolitical Dataset

Formulated in the Wolfram Language, the Socioeconomic and Demographic Data provides seamless access to the curated and continuously updated Wolfram Knowledge base. 27 western eurasian countries were arbitrarily selected in April 2023 for this exercise, as well as two extra countries serving as benchmarking for assignment trials:

- Germany, France, Switzerland, Italy, Austria, United Kingdom, Ireland, Belgium, Luxembourg, Netherlands, Denmark, Sweden, Norway, Finland, Spain, Greece, Portugal, Poland, Lithuania, Czech Republic, Hungary, Romania, Slovakia, Ukraine, Serbia, Turkey, Russia.
- Kazakhstan, Israel.

Each country is characterized by 24 parameters:

- Land Area, Population, Population Growth, Birth Rate Fraction, Migration Rate Fraction.
- Female Literacy Fraction, Male Literacy Fraction, Female Life Expectancy, Male Life Expectancy, Female Infant Mortality Fraction, Male Infant Mortality Fraction.
- Military Expenditure Fraction, Government Consumption, Government Debt, Unemployment Fraction, Inflation Rate, Gini Index.
- Gross Investment, Agricultural Added Value, Industrial Added Value, Industrial Production Growth.
- Gross Domestic Product, Gross Domestic Product Per Capita, Gross Domestic Product Real Growth.

Few examples describe the nature of the data:

- Switzerland, land area: 41'247 square kilometers.
- Belgium, industrial production growth: 0.2% per year.
- Israel, gross investment: $8.75\ 10^{10}$ US Dollars per year.

- Italy, Government debt: 2.57 10^{12}, US Dollars.
- France, unemployment fraction: 9.3972%.
- Turkey, GDP per capita: 8588.2 US Dollars per person per year.
- United Kingdom, male infant mortality fraction: 0.00567 people per people.

As already mentioned, the choice is totally agnostic: no religion, ideology, politics or connection to the European Union has been taken into consideration here.

Figure 21: *A fanciful representation of the dataset. The items - countries - are randomly spread on the surface, with the size of the writing proportional to one of their properties, in this case the Gross Domestic Product Real Growth.*

7.3 Thesaurus

Following designations used in the essay cover almost identical concepts:

- Neuron, Unit.
- Synapse, Connection.
- Item, Stimulus, Input.
- Synaptic or Semantic Space, Semantic grid.
- Properties, Characteristics, Synaptic Weights.
- Neural Network, Neural Grid, Neural Array.
- Voronoï Tile, District.
- Decision, Affectation, Interpretation.
- Expert, Pundit, Prophet.

7.4 The application and its modular structure.

At its current state of development, the use of the software remains a challenge. Next figure provides a glance on the operative sequence. Grey containers are data intermediary lagers. Blue boxes depict software operators. Terminator grey arrows point to the pictures exhibited in this essay.

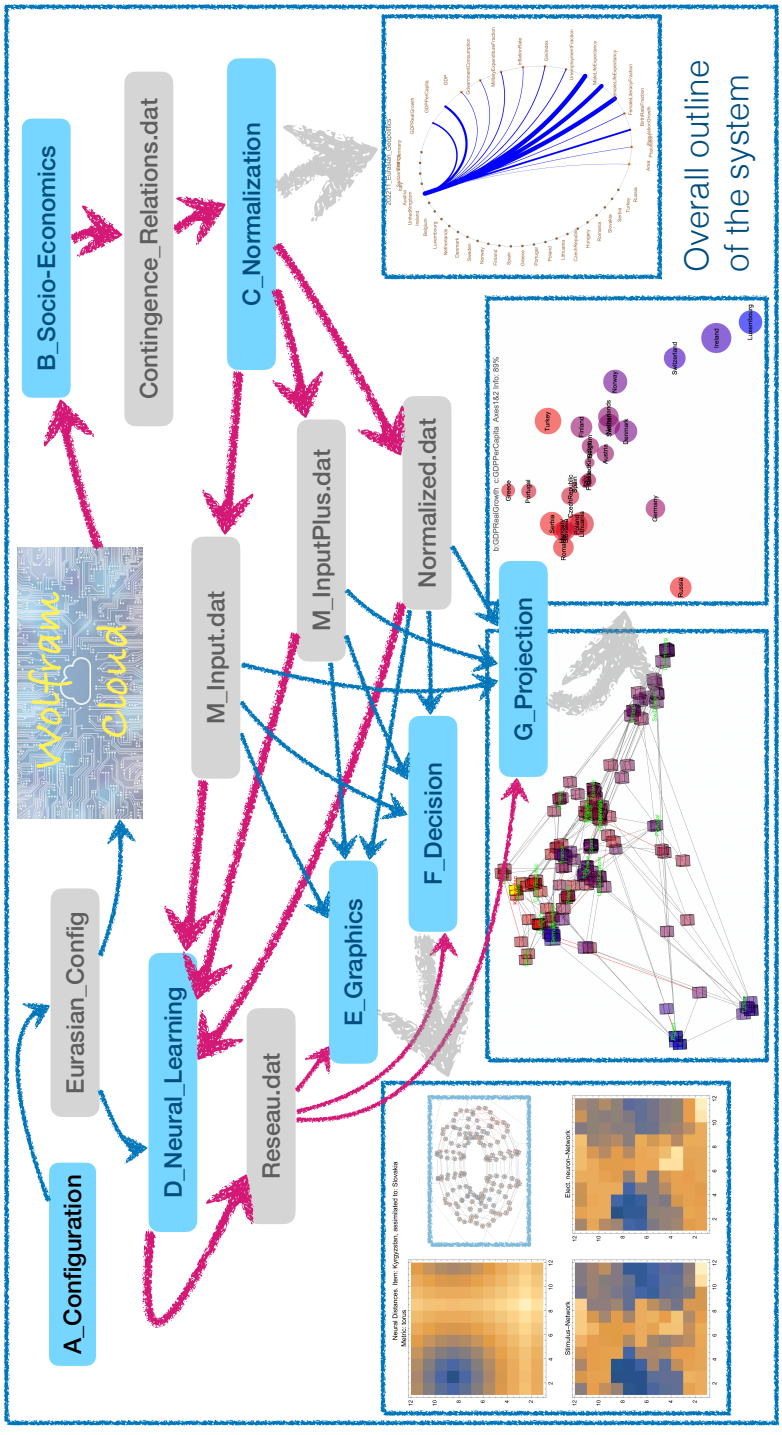


Figure 22: *Data Flow and software modules. Follow the purple paths in order to grasp the overall algorithmic stream.*

8 References

1. Herz J. Krogh A. Palmer R. G. 1991. Introduction to the theory of neural computation. A lecture notes volume in the Santa Fe Institute studies in the sciences of complexity.

2. Ambühl J. 1993. Réseaux de neurones en météorologie: une perspective d'application. (Postgrade EPFL 1992: Réseaux de neurones biologiques et artificiels). Arbeitsbericht Nr. 174 der schweizerischen meteorologischen Anstalt.

3. Eckert P. Cattani D. Ambühl J. 1996. Classification of ensemble forecasts by means of an artificial neural network. Meteorological Applications 3, 169-178.

4. Ambühl J. Cattani D. Eckert P. 2010. Neural interpretation of ECMWF Ensemble Predictions. https://www.meteoschweiz.admin.ch/service-und-publikationen/publikationen/wissenschaftliche-publikationen/2010/neural-interpretation-of-ecmwf-ensemble-predictions-.html.

5. Ambühl J. 2016. Neuronale Netze in der Meteorologie, ein Beispiel. https://www.visualambuhl.ch/201609_Bionik_Neurones_D.pdf.

6. Dennett D. C. 2017. From Bacteria to Bach and Back. The evolution of minds. Allen Lane, an imprint of Penguin Books.

7. Cheng E. 2023. The Joy of Abstraction. Cambridge University Press.

8. Riehl E. 2016. Category Theory in Context. Dover Publications Inc.

Author: Jacques Ambühl. *www.visualambuhl.ch*. Were you to have comments about this essay, I would appreciate a message at *ambuhl@icloud.com*.