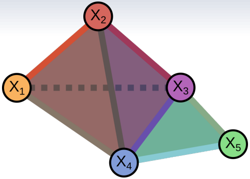


Information Topology: Statistical Physic of Complex Systems and Data Analysis

Pierre Baudot **Daniel Bennequin**

Jean-Marc Goaillard **Monica Tapia-Pacheco et al.**

August 29, 2017
TGSI-2017 - CIRM

"When you use the word information, you should rather use the word form" R.Thom

Contents

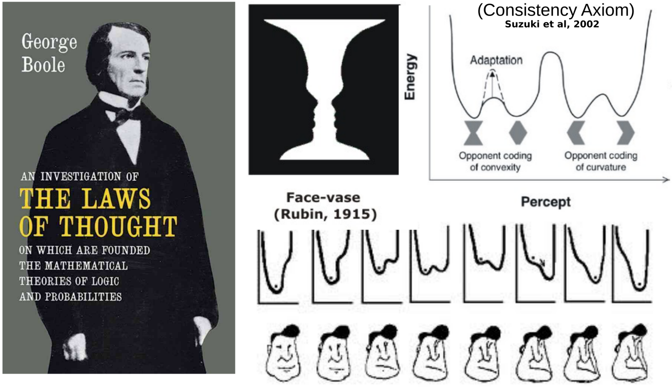
1. Introduction
   - Neuroscience-Cognition
   - Biology
   - Information functions
2. Information Cohomology
   - Information structures
   - Simplicial information cohomology
3. Information landscape and paths and Information topology of Gene expression
   - Information Landscapes
   - Information topology of genetic expression
   - Information paths
   - Minimum free energy complex
4. Conclusion

Neuroscience - Cognition

"It's old like the world, this, the novelty!" J.Prévert.

Cognition: Logic - Probability - Perception:

"It is impossible for any one to believe the same thing to be and not to be" Aristotle (Consistency Axiom)

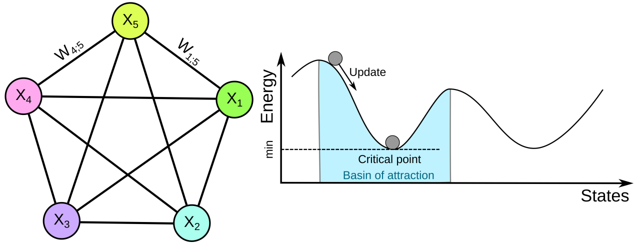


Aristotle
Leibniz
Boole :
Idempotence
$X.(1-X)=0$

Neuroscience - Cognition

Cognition: Neural Network - Machine Learning:

Hopfield
Hinton
Sejnowski
(Boltzmann - Helmholtz machines)
...

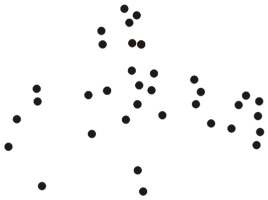


Hopfield Network, 1982

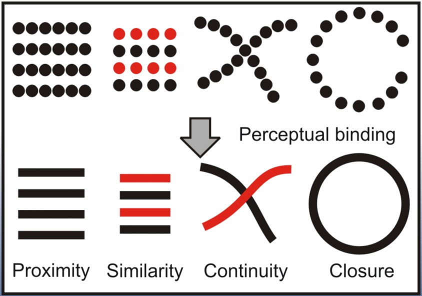
Neuroscience - Cognition

Neuroscience: Learning - Adaptation - Information sensory processing. *"Understanding is compressing"* Chaitin. **Efficient coding** (Attneave, 1952): the goal of sensory perception is to extract the redundancies and to find the most compressed representation of the environment. Any kind of symmetry and invariance are information redundancies and Gestalt principles of perception can be defined on information theoretic terms.

Gestalt
Barlow
Attneave
Laughlin
Linsker
Atick
Nadal
Sejnowski
Bialek...



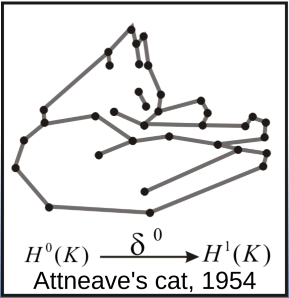
Attneave's cat, 1954

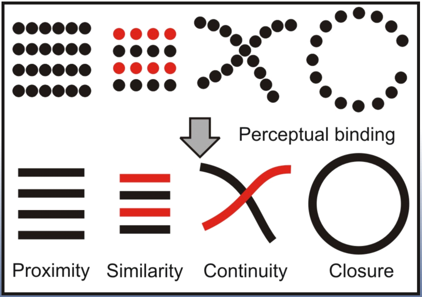


Neuroscience - Cognition

Neuroscience: Learning - Adaptation - Information sensory processing. *"Understanding is compressing"* Chaitin. **Efficient coding** (Attneave, 1952): the goal of sensory perception is to extract the redundancies and to find the most compressed representation of the environment. Any kind of symmetry and invariance are information redundancies and Gestalt principles of perception can be defined on information theoretic terms.

Gestalt
Barlow
Attneave
Laughlin
Linsker
Atick
Nadal
Sejnowski
Bialek...





Biology

Biology: Development - Evolution - Morphogenesis:

Waddington
Thom
Wieschaus
...
Knowledge seems to be energy, and there should be enough energy in this room to make nice things.

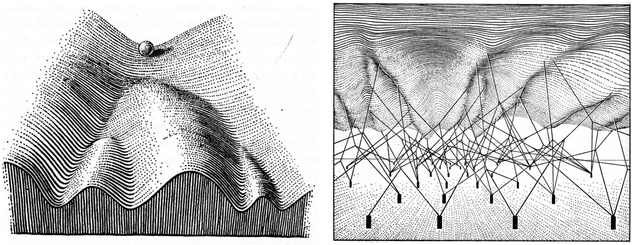


FIGURE 4
Part of an Epigenetic Landscape. The path followed by the ball, as

FIGURE 5
The complex system of interactions underlying the epigenetic landscape.

C.H. Waddington, 1957

Information functions

The information functions used here are defined by Shannon and generalized by Hu Kuo Ting and Yeung, (using $k = -1/\ln 2$, bit):

- The entropy of a single variable:

$$H_1 = H(X_j; P_{X_j}) = k \sum_{x \in [N_j]} p(x) \ln p(x)$$

  where $[N_j] = \{1, ..., N_j\}$ denotes the alphabet of X_j.

- The joint entropy:

$$H_k = H(X_1, ..., X_k; P_{X_1, ..., X_k}) = k \sum_{x_1, ..., x_k \in [N_1 \times ... \times N_k]}^{N_1 \times ... \times N_k} p(x_1 .... x_k) \ln p(x_1 .... x_k)$$

  where $[N_1 \times ... \times N_k] = \{1, ..., N_j \times ... \times N_k\}$ denotes the alphabet of $(X_1, ..., X_k)$ and $\mathbb{P}_{(X_1, ..., X_k)}$ the joint probability joint-distribution.

Information functions

- The 2-mutual information

$$I_2 = I(X_1; X_2; P_{X_1,X_2}) = k \sum_{x_1,x_2 \in [N_1 \times N_2]}^{N_1 \times N_2} p(x_1.x_2) \ln \frac{p(x_1)p(x_2)}{p(x_1.x_2)}$$

- generalized to k-mutual-information :

$$H_n(X_1, ..., X_n; P) = \sum_{i=1}^{n} (-1)^{i-1} \sum_{I \subset [n]; card(I)=i} I_i(X_I; P)$$

Ex: $I_3 = H(1) + H(2) + H(3) - H(1,2) - H(1,3) - H(2,3) + H(1,2,3)$, giving:

$$I_k = I(X_1; ...; X_k; P) = k \sum_{x_1,\dots,x_k \in [N_1 \times \dots \times N_k]}^{N_1 \times \dots \times N_k} p(x_1.....x_k) \ln \frac{\prod_{I \subset [k]; card(I)=i; i \text{ odd}} p_I}{\prod_{I \subset [k]; card(I)=i; i \text{ even}} p_I}$$

Ex: $I_3 = k \sum_{x_1,x_2,x_3 \in [N_1 \times N_2 \times N_3]}^{N_1 \times N_2 \times N_3} p(x_1.x_2.x_3) \ln \frac{p(x_1)p(x_2)p(x_3)p(x_1.x_2.x_3)}{p(x_1.x_2)p(x_1.x_3)p(x_2.x_3)}$. For $k \geq 3$, I_k can be negative.

Information functions

- Conditional entropy-information:

$$X_2.H_1 = H(X_1|X_2;P) = k \sum_{x_1,x_2 \in [N_1 \times N_2]}^{N_1 * N_2} p(x_1.x_2) \ln p_{x_2}(x_1)$$

- The conditional mutual information:

$$X_3.I_2 = I(X_1;X_2|X_3;P) = k \sum_{x_1,x_2,x_3 \in [N_1 \times N_2 \times N_3]}^{N_1 \times N_2 \times N_3} p(x_1.x_2.x_3) \ln \frac{p_{x_3}(x_1)p_{x_3}(x_2)}{p_{x_3}(x_1,x_2)}$$

Conditional mutual information generates the preceding information functions as subcases (Yeung). We have the theorem : if $X_3 = \Omega$ then it gives the mutual information, if $X_2 = X_1$ it gives conditional entropy, and if both conditions are satisfied, it gives entropy. Notably, we have $I_1 = H_1$.

Chain rules of information

- $$H_{k+1} - H_k = (X_1, ...X_k).H(X_{k+1}) \tag{1}$$

- $$I_{k-1} - I_k = X_k.I_{k-1} \tag{2}$$

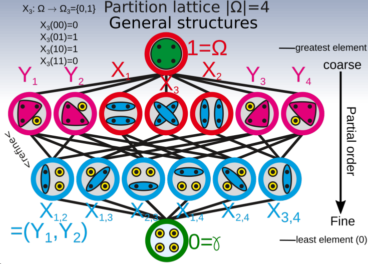
- by recurrence :

$$H_k = H(X_1, ..., X_k; P) = \sum_{i=1}^{k} (X_1, ..., X_{i-1}).H(X_i; P)$$

$$I_k = I(X_1; ...; X_k; P) = I(X_1) - \sum_{i=2}^{k} X_i.I(X_1; ...; X_{i-1})$$

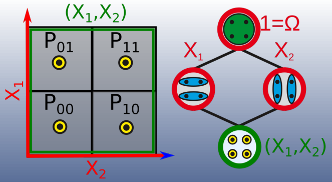
Information structures

Baudot and Bennequin, *The Homological nature of entropy*, Entropy, 2015.



- The random variables are **partitions** of the atomic probabilities of $(\Omega, \mathcal{B}, P)$ (equivalence classes).
- The **Joint-Variable** (X_1, X_2) is the less fine partition that is finer than X_1 and X_2 (gcd).
- The (general) **information structure** is the triple (Ω, Π, P) where Π is the lattice of all partitions.
- Information functions, $F(X_1, ..., X_k; P)$ is the real module of all **measurable functions** defined on the whole lattice of partitions. $(X_1, ..., X_k; P)$ is the image law of P by $(X_1, ..., X_k)$.

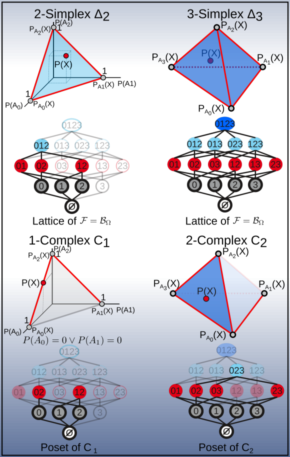
Atomic probabilities |Ω|=4
2 binary variables - simplicial structures

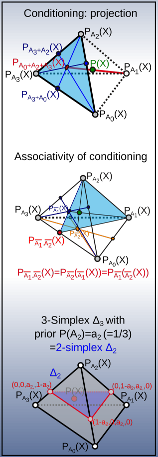


Probability Simplex

The probability space $(\Omega, \mathcal{B})$, $|\Omega| = N$ is a $(N-1)$-simplex of probability, implementing geometrically Kolmogorov axiomatic:

- $\sum_i P(A_i) = 1$ the geometry is affine
- $P(A_i) \geq 0$ convex
- Theorem of total probability: barycentric coordinate $P(X) = \sum_i P(A_i.X) = \sum_i P(A_i).P_{A_i}(X)$
- Conditioning is a projection on subsimplex.
- Complex of probability given by set of constraints of the form $P(A_0) = 0 \vee P(A_1) = 0$





Actions and coboundaries

Conditioning-expectation by Y, $Y.F(X_1,...,X_k;P)$, is the left action of Y on the functional module, $Y.F(X;P)=\sum_i P(Y=y_i)F(X;P_{Y=y_i})$. The action of conditioning is **associative**, we have $X.(Y.F(Z;P))=(X,Y).F(Z;P)$.

Complexes of random variables are $X^k=(X_1,...,X_k;P)$, and we consider cochain complexes (X^k,∂^k):

$$0 \rightarrow X^0 \xrightarrow{\partial^0} X^1 \xrightarrow{\partial^1} X^2 \xrightarrow{\partial^2} ...X^{k-1} \xrightarrow{\partial^{k-1}} X^k$$

Actions and coboundaries

3 coboundaries with left, trivial and symmetric action (Hochschild 1945, for associative and ring structures):

- The left action coboundary (Galois cohomology):

$$(\partial^k)F(X_1;X_2;...;X_{k+1};P) = X_1.F(X_2;...;X_{k+1};P)$$

$$+\sum_{i=1}^{k}(-1)^i F(X_1;X_2;...;(X_i,X_{i+1});...;X_{k+1};P) + (-1)^{k+1}F(X_1;...;X_k;P)$$

Actions and coboundaries

3 coboundaries with left, trivial and symmetric action (Hochschild 1945, for associative and ring structures):

- The left action coboundary (Galois cohomology):

$$(\partial^k)F(X_1;X_2;...;X_{k+1};P) = X_1.F(X_2;...;X_{k+1};P) + \sum_{i=1}^{k}(-1)^i F(X_1;X_2;...;(X_i,X_{i+1});...;X_{k+1};P) + (-1)^{k+1}F(X_1;...;X_k;P)$$

- The "topological-trivial" coboundary and cohomology (trivial left action $X_1.F(X_2;...;X_{k+1}) = F(X_2;...;X_{k+1})$) :

$$(\partial^k_t)F(X_1;X_2;...;X_{k+1};P) = F(X_2;...;X_{k+1};P) + \sum_{i=1}^{k}(-1)^i F(X_1;X_2;...;(X_i,X_{i+1});...;X_{k+1};P) + (-1)^{k+1}F(X_1;...;X_k;P)$$

Actions and coboundaries

3 coboundaries with left, trivial and symmetric action (Hochschild 1945, for associative and ring structures):

- The left action coboundary (Galois cohomology):

$$(\partial^k)F(X_1;X_2;...;X_{k+1};P) = X_1.F(X_2;...;X_{k+1};P)$$
$$+\sum_{i=1}^{k}(-1)^i F(X_1;X_2;...;(X_i,X_{i+1});...;X_{k+1};P) + (-1)^{k+1}F(X_1;...;X_k;P)$$

- The symmetric Hochschild-information coboundary (Gerstenhaber and Shack, symmetric left and right action of conditioning $X_1.F(X_2;...;X_{k+1}) = F(X_2;...;X_{k+1}).X_1$):

$$(\partial_*^k)F(X_1;X_2;...;X_{k+1};P) = X_1.F(X_2;...;X_{k+1};P) + \sum_{i=1}^{k}(-1)^i$$
$$F(X_1;X_2;...;(X_i,X_{i+1});...;X_{k+1};P) + (-1)^{k+1}X_{k+1}.F(X_1;...;X_k;P)$$

Cohomology in the first degree

For the first degree $k = 1$, we have the following results:

- The left 1-co-boundary is $(\partial^1)F(X_1;X_2) = X_1.F(X_2) - F(X_1,X_2) + F(X_1)$. The 1-cocycle condition $(\partial^1)F(X_1;X_2) = 0$ gives $F(X_1,X_2) = F(X_1) + X_1.F(X_2)$ which is the chain rule of information. Then following Kendall and Lee (1964), it is possible to recover the functional equation of information and to characterize uniquely, up to the arbitrary multiplicative constant k, the entropy as the first class of cohomology.

Main theorem [?]

The information co-homology space of degree one is one-dimensional and generated by entropy.

Cohomology in the first degree

For the first degree $k = 1$, we have the following results:

- The left 1-co-boundary is $(\partial^1)F(X_1;X_2) = X_1.F(X_2) - F(X_1,X_2) + F(X_1)$. The 1-cocycle condition $(\partial^1)F(X_1;X_2) = 0$ gives $F(X_1,X_2) = F(X_1) + X_1.F(X_2)$ which is the chain rule of information. Then following Kendall and Lee (1964), it is possible to recover the functional equation of information and to characterize uniquely, up to the arbitrary multiplicative constant k, the entropy as the first class of cohomology.

Main theorem [?]

The information co-homology space of degree one is one-dimensional and generated by entropy.

- Topological 1-coboundary $(\partial_t^1)F(X_1;X_2) = F(X_2) - F(X_1,X_2) + F(X_1)$ gives I_2: $(\partial_t^1)F(X_1;X_2) = H(X_1) + H(X_2) - H(X_1,X_2) = I(X_1;X_2)$

Cohomology in the first degree

For the first degree $k = 1$, we have the following results:

- The left 1-co-boundary is $(\partial^1)F(X_1; X_2) = X_1.F(X_2) - F(X_1, X_2) + F(X_1)$. The 1-cocycle condition $(\partial^1)F(X_1; X_2) = 0$ gives $F(X_1, X_2) = F(X_1) + X_1.F(X_2)$ which is the chain rule of information. Then following Kendall and Lee (1964), it is possible to recover the functional equation of information and to characterize uniquely, up to the arbitrary multiplicative constant k, the entropy as the first class of cohomology.

Main theorem [?]

The information co-homology space of degree one is one-dimensional and generated by entropy.

- Symmetric 1-coboundary:
  $(\partial^1_*)F(X_1; X_2) = X_1.F(X_2) - F(X_1, X_2) + X_2.F(X_1)$ is the negative of I_2
  $(\partial^1_*)F(X_1; X_2) = X_2.H(X_1) - H(X_1, X_2) + X_1.H(X_2) = -I(X_1; X_2)$.

Cohomology in the second degree

For the second degree $k = 2$, we have the following results:

- The left 2-co-boundary $\partial^2 F(X_1;X_2;X_3) = X_1.F(X_2;X_3) - F((X_1,X_2);X_3) + F(X_1;(X_2,X_3)) - F(X_1;X_2)$ is minus the 3-mutual information $\partial^2 F(X_1;X_2;X_3) = X_1.I(X_2;X_3) - I((X_1,X_2);X_3) + I(X_1;(X_2,X_3)) - I(X_1;X_2) = -I(X_1;X_2;X_3)$.
- The topological 2-coboundary is $(\partial_t^2)F(X_1;X_2;X_3) = F(X_2;X_3) - F((X_1,X_2);X_3) + F(X_1;(X_2,X_3)) - F(X_1;X_2)$, is $\partial_t^2 F(X_1;X_2;X_3) = I(X_2;X_3) - I((X_1,X_2);X_3) + I(X_1;(X_2,X_3)) - I(X_1;X_2) = 0$.
- The symmetric 2-coboundary is $(\partial_*^2)F(X_1;X_2;X_3) = X_1.F(X_2;X_3) - F((X_1,X_2);X_3) + F(X_1;(X_2,X_3)) - X_3.F(X_1;X_2)$ is $\partial_*^2 F(X_1;X_2;X_3) = X_1.I(X_2;X_3) - I((X_1,X_2);X_3) + I(X_1;(X_2,X_3)) - X_3.I(X_1;X_2) = 0$.

Cohomology in the third degree

For the second degree $k = 3$, we have the following results:

- The left 3-co-boundary
  $\partial^3 F(X_1; X_2; X_3; X_4) = X_1.F(X_2; X_3; X_4) - F((X_1, X_2); X_3; X_4) + F(X_1; (X_2, X_3); X_4) - F(X_1; X_2; (X_3, X_4)) + F(X_1; X_2; X_3)$ is
  $\partial^3 F(X_1; X_2; X_3; X_4) = X_1.I(X_2; X_3; X_4) - I((X_1, X_2); X_3; X_4) + I(X_1; (X_2, X_3); X_4) - I(X_1; X_2; (X_3, X_4)) + I(X_1; X_2; X_3) = 0.$
- The topological 3-coboundary
  $\partial_t^3 F(X_1; X_2; X_3; X_4) = F(X_2; X_3; X_4) - F((X_1, X_2); X_3; X_4) + F(X_1; (X_2, X_3); X_4) - F(X_1; X_2; (X_3, X_4)) + F(X_1; X_2; X_3)$ is
  $\partial_t^3 F(X_1; X_2; X_3; X_4) = I(X_2; X_3; X_4) - I((X_1, X_2); X_3; X_4) + I(X_1; (X_2, X_3); X_4) - I(X_1; X_2; (X_3, X_4)) + I(X_1; X_2; X_3) = I(X_1; X_2; X_3; X_4).$
- The symmetric 3-coboundary
  $(\partial_*^3) F(X_1; X_2; X_3; X_4) = X_1.F(X_2; X_3; X_4) - F((X_1, X_2); X_3; X_4) + F(X_1; (X_2, X_3); X_4) - F(X_1; X_2; (X_3, X_4)) + X_4.F(X_1; X_2; X_3)$ is
  $\partial_*^3 F(X_1; X_2; X_3; X_4) = X_1.I(X_2; X_3; X_4) - I((X_1, X_2); X_3; X_4) + I(X_1; (X_2, X_3); X_4) - I(X_1; X_2; (X_3, X_4)) + X_4.I(X_1; X_2; X_3) = -I(X_1; X_2; X_3; X_4).$

Cohomology in the higher degrees

It is possible to generalize to arbitrary degrees by remarking that we have:

- For even degrees $2k$: we have $I_{2k} = -\partial_t I_{2k-1}$ and then $I_{2k} = \partial_t \partial \partial_t \ldots \partial \partial_t H$ with $2k-1$ boundary terms.
- For odd degrees $2k+1$: $I_{2k+1} = -\partial I_{2k-1}$ and then $I_{2k+1} = -\partial \partial_t \partial \ldots \partial \partial_t H$ with $2k$ boundary terms.

Theorem [?]

Let X^n be an information structure, then:

- For even degrees $2k$: $\partial^{2k} = -I_{2k+1}$ and $\partial_*^{2k} = -\partial_t^{2k} = 0$
- For odd degrees $2k+1$: $\partial^{2k-1} = 0$ and $\partial_*^{2k-1} = -\partial_t^{2k} = -I_{2k}$.

Information double complex

Attempt to make a single cohomology following Gerstenhaber and Shack (Hodge decomposition of Hochschild cohomology) by constructing a double complex $X^{\bullet,\bullet}$, the triplet $(X^{\bullet,\bullet}, \partial, \partial_*) = (X^{k',k''}, \partial^{k',k''}, \partial_*^{k',k''})$. We have $\partial^k \partial_*^k + \partial_*^k \partial^k = 0$. The total complex is defined by $X_{\text{Tot}}^k = \bigoplus_{k'+k''=k} X_{\text{Tot}}^{k',k''}$, with coboundary $\partial_{tot}^k = \partial^k + (-1)^k \partial_*^k$ and then the coboundary of the total complex of information is $\partial_{tot}^k = (-1)^{k+1} I_{k+1}$

$$\begin{tikzcd}
X^{k,0} \arrow[r, "\partial^{k,0}"'] & X^{k,1} \arrow[r, "\partial^{k,1}"'] & X^{k,2} \arrow[r, "\partial^{k,2}"'] & X^{k,3} \arrow[r, "\partial^{k,3}"'] & \cdots \arrow[r, "\partial^{k,k-1}"'] & X^{k,k} \\
\vdots \arrow[u, "\partial_*^{k-1,0}"'] & \vdots \arrow[u, "\partial_*^{k-1,1}"'] & \vdots \arrow[u, "\partial_*^{k-1,2}"'] & \vdots \arrow[u, "\partial_*^{k-1,3}"'] & & \vdots \arrow[u, "\partial_*^{k-1,k}"'] \\
X^{2,0} \arrow[u, "\partial_*^{2,0}"'] \arrow[r, "\partial^{2,0}"'] & X^{2,1} \arrow[u, "\partial_*^{2,1}"'] \arrow[r, "\partial^{2,1}"'] & X^{2,2} \arrow[u, "\partial_*^{2,2}"'] \arrow[r, "\partial^{2,2}"'] & X^{2,3} \arrow[u, "\partial_*^{2,3}"'] \arrow[r, "\partial^{2,3}"'] & \cdots \arrow[r, "\partial^{2,k-1}"'] & X^{2,k} \arrow[u, "\partial_*^{2,k}"'] \\
X^{1,0} \arrow[u, "\partial_*^{1,0}"'] \arrow[r] & X^{1,1} \arrow[u, "\partial_*^{1,1}"'] \arrow[r] & X^{1,2} \arrow[u, "\partial_*^{1,2}"'] \arrow[r] & X^{1,3} \arrow[u, "\partial_*^{1,3}"'] \arrow[r] & \cdots \arrow[r] & X^{1,k} \arrow[u, "\partial_*^{1,k}"']
\end{tikzcd}$$

k-independence cocycle

Theorem 2-independence $\Leftrightarrow \partial^1_* = 0$ (Li, 1990)

X_1, X_2 are statistically independent if and only if $I_2 = I(X_1, X_2; P) = 0$

Moreover, $I(X_1, X_2) = 0 \Rightarrow \rho_{X_1,X_2} = 0,\ \rho_{X_1,X_2} = \frac{\mathrm{cov}(X_1,X_2)}{\sigma_{X_1}\sigma_{X_2}}$.

Definition k-independence

$X_1, ..., X_k$ are k-independent if $I_k = 0$

Theorem mutual-independence

$X_1, ..., X_n$ are mutually independent if and only if $\forall k \leq n,\ I_k = 0$.

As a probabilistic interpretation, information cohomology quantifies statistical dependences at all degrees, the obstruction to factorization: k-independence coincides with cocycles.

Simplicial information substructures

Tapia-Pacheco et al., Information topology of gene expression profile in dopaminergic neurons, bioRxiv 168740 (2017).

- **Computational problem:** complexity of the estimation of information functions: Bell's combinatoric $\mathcal{O}(exp(exp(N^n)))$ for n N-ary variables. At each degree k, the number of H_k and I_k to evaluate is given by Stirling numbers $S(N^n, k)$ with $B_{N^n} = \sum_{k=0}^{N^n} S(N^n, k)$. Ex: 16 variables, 8 values each: $|\Pi| \approx e^{e^{2^{48}}-1} > 2^{200}$ elements to compute.
- **Computational solution**: Data analysis is developed on the simplest sub-case of the general information structure, the simplicial information structure and the simplicial information cohomology with complexity in $\mathcal{O}(2^n)$ and $\binom{n}{k}$ elements for each k with $2^n = \sum_{k=1}^{n} \binom{n}{k}$.
- **Consequence:** some possible statistical dependences cannot be detected.

Simplicial information substructures

- **A simplicial information structure** is the triple (Ω, Δ^n, P) where Δ^n is the Boolean lattice of all subsets (2^n elements and $\binom{n}{k} = \frac{n!}{k!(n-k!)}$ elements at each degree k in one to one correspondence with the k-faces of the n-simplex of random variables.
- **Joint (X_1, X_2) and meet $(X_1; X_2)$ of variables** are the usual joint and meet of Boolean algebra and define two opposite-dual monoids $(X_1, .., X_n, \vee)$ $(X_1, .., X_n, \wedge)$, generating freely the semi-lattice of all subsets and its dual.

Simplicial information substructures

Theorem simplicial information

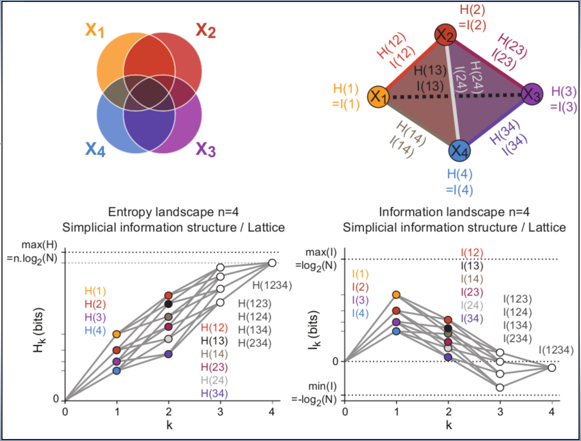
A simplicial information structure is a substructure of information structure.

Proof: using theorem of Pudlak (1980): any finite lattice is a sub-lattice of the partition lattice.

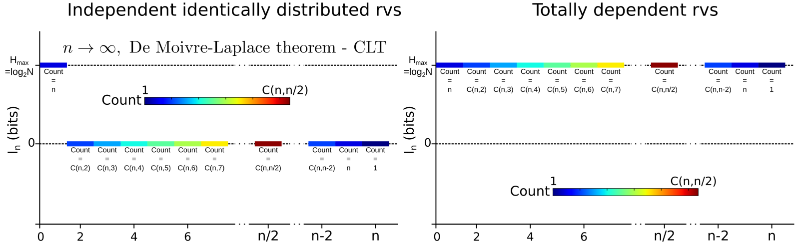
- A simplicial complex of random variables $X^k = (X_1, ..., X_k; P)$ is any subcomplex of the simplex Δ^n with $k \leq n$, and any simplicial complex can be realized as a subcomplex of a simplex (Steenrod, 1947).
- The simplicial information homology is a (usual simplicial) subcase of information homology and defined as previously (cochain complexes (X^k, ∂^k))
- In this ordinary homological structure, the degree obviously coincide with the dimension of the data space.

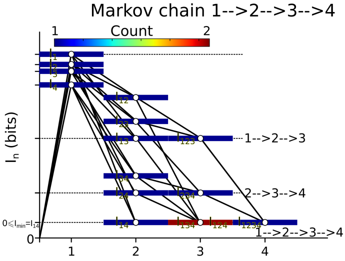
H_k and I_k Landscapes

- **Information landscapes:** represent the lattice of information structures with in abscissa the degrees k and in ordinate the values of H_k and I_k.
- H_k and I_k (real continuous functions): ranking of the lattices at each k.
- H_k quantify variability-randomness, I_k quantify statistical dependences.



Special cases of H_k and I_k Landscapes





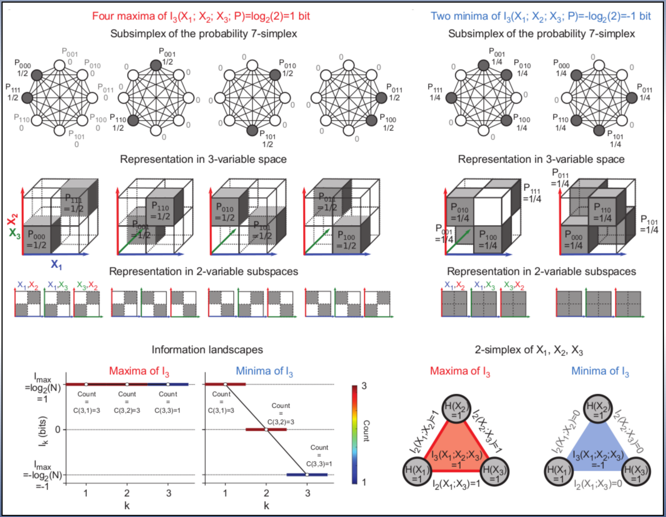
Theorem (Hu Kuo Tin, 1962):
(1-->...-->n)Markov chain <=> $I_2(1,n)=I_k(1,\dots,n)$
Positivity: (1-->...-->n)Markov chain => $0 \leqslant I_2(1,n) \leqslant I_k$
Data Processing ineq: $I_{13} \leqslant I_{12}$, $I_{13} \leqslant I_{23}$, $I_{24} \leqslant I_{34}$, $I_{24} \leqslant I_{23}$, $I_{14} \leqslant I_{24}$, $I_{14} \leqslant I_{13}$, $I_{124} \leqslant I_{234}$, $I_{124} \leqslant I_{123}$

I_k extrema and negativity - Special cases

Theorem (Hu Kuo Ting, 1962)

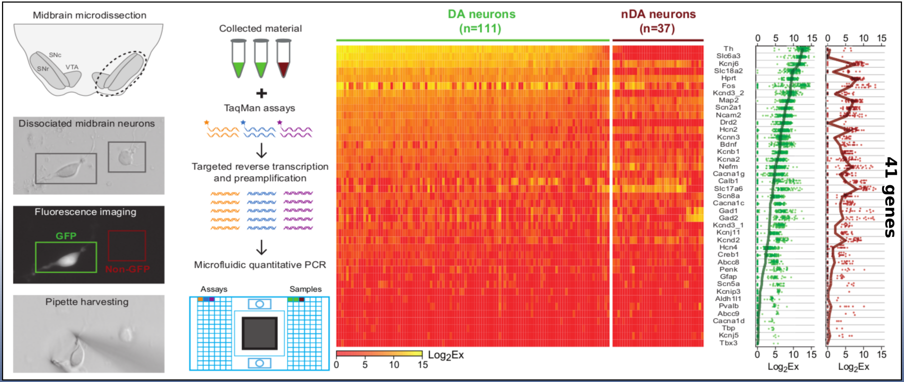
For $k \geq 3$ I_k can be negative.

- Schrödinger "what is life?": living system feed upon negentropy (free-energy)
- Synergy (Brenner et al.)
- Frustrated spin glasses (Matsuda)



Gene expression measures

- Quantitative PCR of single neurons in SNc (dopaminergic) and other midbrain nucleus (nDA)
- mRNA expression levels for $n = 41$ genes in $m = 111$ DA and $m = 37$ nDA



Probability estimation

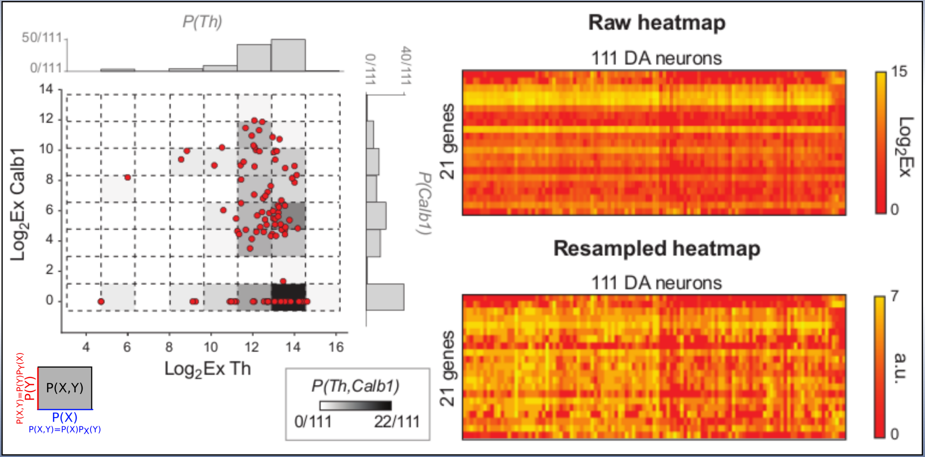
- **heatmap:** (m, n) matrix D with real/rational coefficients $x_{ij} \in \mathbb{R},\ i \in \{1..m\},\ j \in \{1...n\}$
- **Graining:** the intervals $[\min x_j, \max x_j]$ for each variable X_j is divided into $N_j = 8$ giving $N_1.N_2...N_n = 8^n$ boxes in n-dimensions.
- **Estimation of the atomic probabilities**: usual counting, defined as:

$$P\left(\mathrm{bmin}_1 \leq X_1 \leq \mathrm{bmax}_1, \mathrm{bmin}_2 \leq X_2 \leq \mathrm{bmax}_2, ..., \mathrm{bmin}_n \leq X_n \leq \mathrm{bmax}_n\right)$$

$$= \sum_{i=1}^{m} \frac{\delta_i}{m},\ \delta_i = \begin{cases} 0, & \text{if}\,\mathrm{bmin}_1 > x_{i1} \text{ or } x_{i1} > \mathrm{bmax}_1 \text{ ...or } \mathrm{bmin}_n > x_{in} \text{ or } x_{in} > \mathrm{br} \\ 1, & \text{if}\,\mathrm{bmin}_1 \leq x_{i1} \leq \mathrm{bmax}_1 \text{ and...and } \mathrm{bmin}_n \leq x_{in} \leq \mathrm{bmax}_n \end{cases}$$

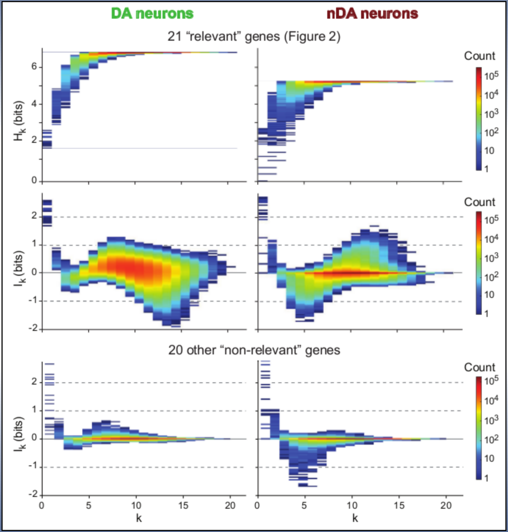
- **Marginalization (projection on lower dimensions, on subsets of variables)**: Conditioning $P_X(Y) = \frac{P(X.Y)}{P(X)}$ and in general the theorem of total probability $(P(X) = \sum_{i=0}^{N} P(A_i.X) = \sum_{i=0}^{N} P(A_i).P_{A_i}(X))$.

Probability estimation



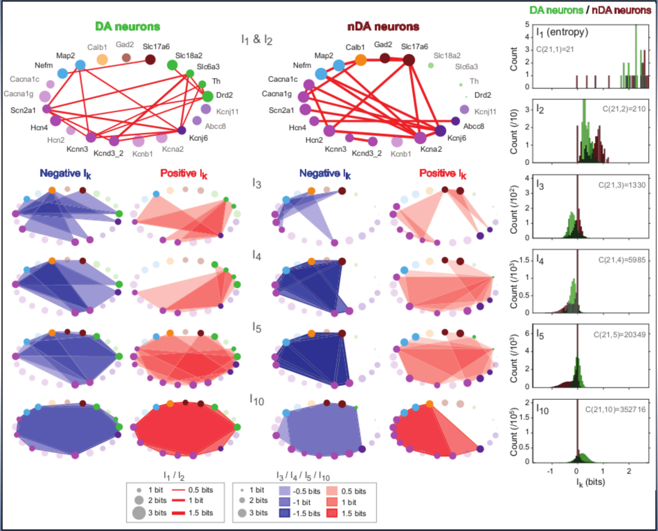
Information topology of genetic expression

- Computational restriction to $n = 21$ ($2^{21} \approx 2.10^6$ elements)
- Positive, negative I_k and k-independence $I_k = 0$ even for high k

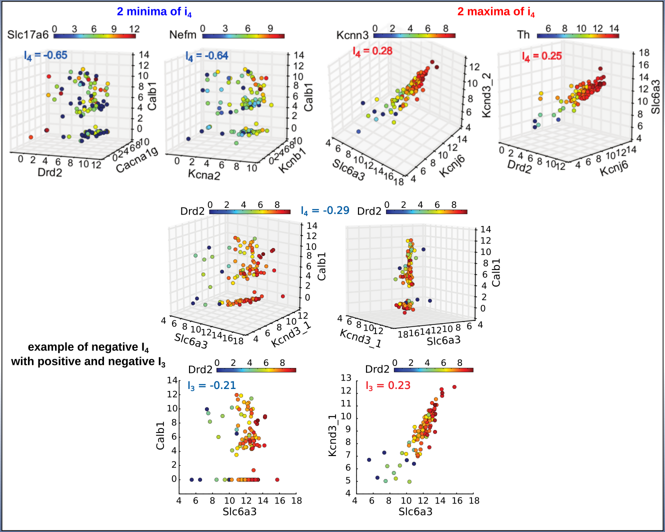


Maximum and minimum I_k "modules"

- I_2 qualitatively similar to $\rho_{X,Y}$ (Reshef, 2011).
- **Combinatorial complexity of interactions:** diversity and impressively numerous!
- I_k are nontheless specific to a given cell type: **cell identity signature**.



Maximum and minimum I_k "modules"



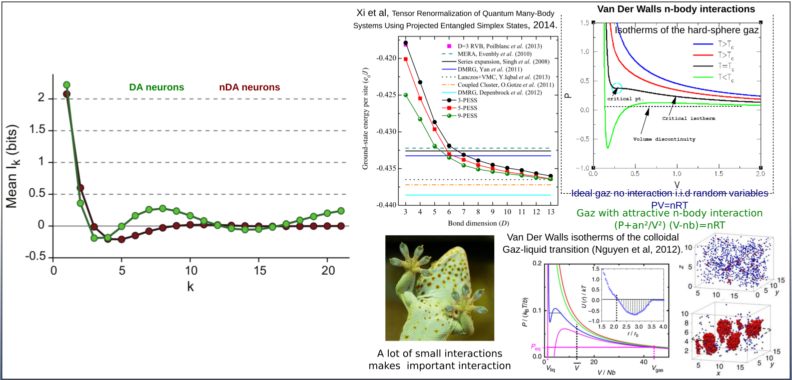
- Negative I_k detects clusters
- Positive I_k detects covariations

Mean H_k and I_k

Mean behavior of the information structure defined the mean H_k and I_k:

$$\langle H_k \rangle = \frac{\sum_{T\subset[n];card(T)=i} H_k(X_T;P)}{\binom{n}{k}}, \quad \langle I_k \rangle = \frac{\sum_{T\subset[n];card(T)=i} I_k(X_T;P)}{\binom{n}{k}}$$

Mean information correspond to ideal homogeneous structure X^n_{hom} with homogeneous k-body interactions.



Free information energy n-body interaction

Two different components in information structures:

- for $k = 1$, I_1 and $\langle I_1 \rangle$ are a self-interaction $I(X_i) = H(X_i)$. We call the first degree mutual information component $I(X_i) = H(X_i)$ the **self information or kinetic information energy**, in analogy to thermodynamic (Density Functional Theory, Kohn-Hohenberg). Here self-interaction does not diverge: no regularization or renormalization corrections.
- for $k > 1$ I_k and $\langle I_k \rangle$ quantifies the contribution of the k-body interaction. We call I_k and $\langle I_k \rangle$ the **k-free-information-energy**. Each I_k terms is as a free energy correction accounting for the k-body interactions.

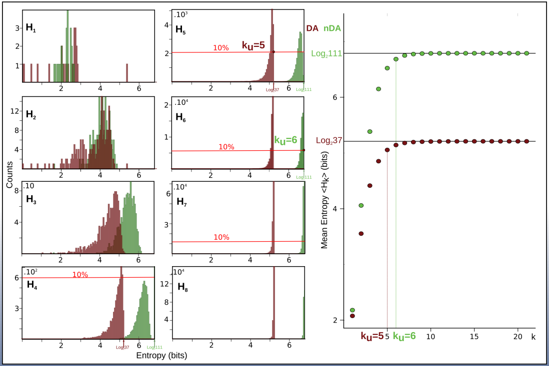
Defining self information energy by $T(X_1, ..., X_n; P_N) = \sum_{i=1}^{n} I_1(X_i; P_N)$ and free-information-energy by $G(X_1, ..., X_n; P_N) = \sum_{i=2}^{n} (-1)^{i-1} \sum_{I \subset [n]; |I|=i} I_i(X_I; P_N)$ we recover the usual isotherm thermodynamic relation:

$$H_n(X_1, ..., X_n; P_N) = T(X_1, ..., X_n; P_N) - G(X_1, ..., X_n; P_N)$$

Finite size effect - sampling problem

Dependence on m. Undersampling (curse of dimension/sampling problem (Strong 1998,Nemenman 2004)): when $N_1 ... N_n$ are such that only one data point falls in a box then $p = 1/m$ and $H_n = \log_2 m$.

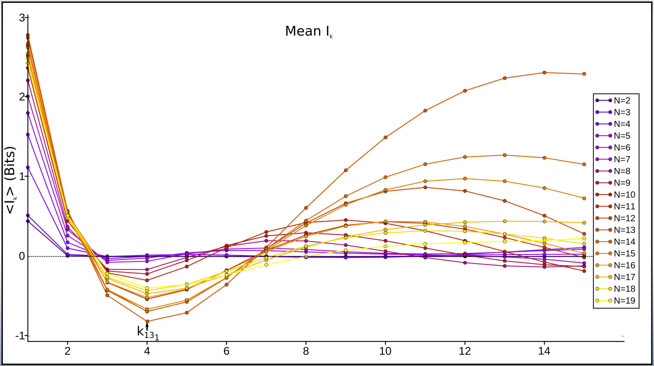
- Add a combinatorial number of 0 values to I_k landscape
- Equivalent to adding a deterministic variable 0 since the probability remains unchanged ($1/m$)
- Degree k_u for which more than 10% of the H_k are in $\log_2 m - 0.05 \leq H_k \leq \log_2 m$.
- Analysis holds well bellow usual undersampling regime.



Dependence on graining N

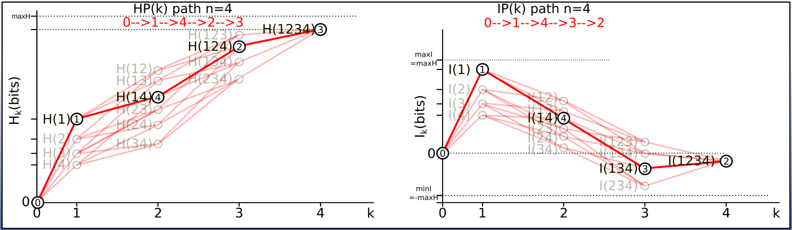
Dependence on N. Study of iso-graining landscapes and apparition of critical points in the same way as isotherms does in usual thermodynamic.

- For $N = 2$ the mean $\langle I_k \rangle$ is monotinicaly decreasing. This $N = 2$ iso-graining is analog to the non condensed disordered phase.
- All the other mean $\langle I_k \rangle$ paths have a critical point.



Information paths

Information path: Let (Ω, Δ^k, P) be a simplicial information structure, then a path of degree k in Δ^k is a sequence of edges of the lattice that begins at the leastest element of the lattice (the identity-constant 0), travels along edges from vertex to vertex of increasing degree of the lattice and ends at the greatest element of the lattice of degree k.
Information paths are defined dually on joint and meet-mutual information semi-lattice. The set of all information paths are noted HP_k paths for entropy paths and IP_k for mutual-information paths.

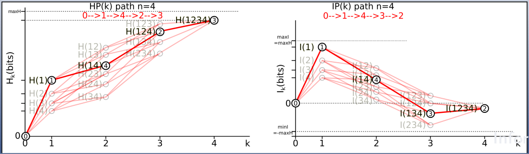


Information paths - Symmetric group

Theorem information path symmetric group

The set of all information paths HP_k and IP_k in Δ^k are both in bijection with the symmetric group S_k. Notably, there are $k!$ paths in Δ_k.

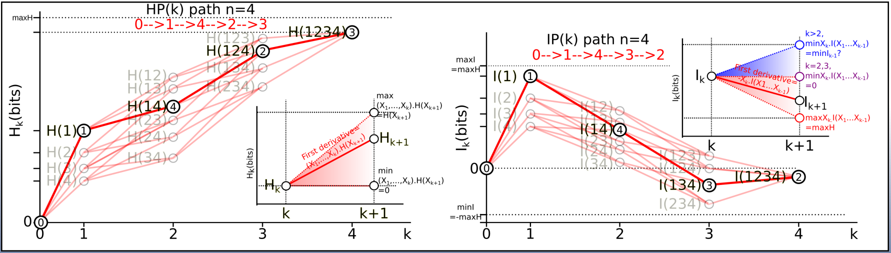
A path in Δ_4, noted $IP_i = (0 \rightarrow X_2 \rightarrow X_1 \rightarrow X_4 \rightarrow X_3)$ can be identified with **a permutation** or **a total order**:
$IP_i = 0 \rightarrow (0, X_2) \rightarrow (0, X_1, X_2) \rightarrow (0, X_1, X_2, X_4) \rightarrow (0, X_1, X_2, X_3, X_4)$ can be noted $\sigma : (01234) \xrightarrow{\sigma} (02143)$. Paths are seen as automorphism of $\{1, 2.....k\} = [k]$ and HP_k and IP_k paths can be endowed with the structure of two opposite symmetric group S_k.



Derivatives of information paths

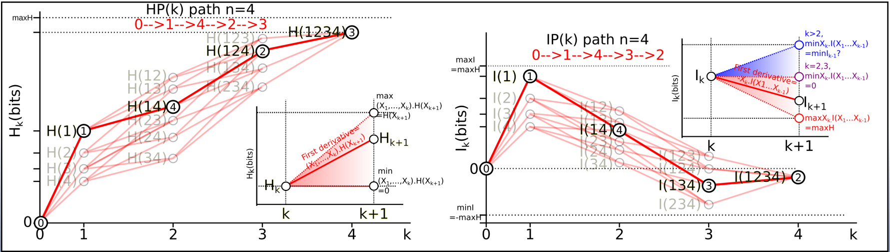
The paths HP_i and IP_i as piecewise linear functions, $IP_i(k) = I_k$.

- **First derivative of entropy path** $HP_i(k)$ is conditional entropy: $dHP_i(k)/dk = H_k - H_{k-1} = (X_1, ..., X_{k-1}).H(X_k; \mathbb{P})$
- **First derivative of mutual information path** $IP_i(k)$ is minus conditional information (coface map): $dIP_i(k)/dk = I_k - I_{k-1} = -X_k.I(X_1, ..., X_{k-1}; \mathbb{P})$



Derivatives Bounds, Information inequalities

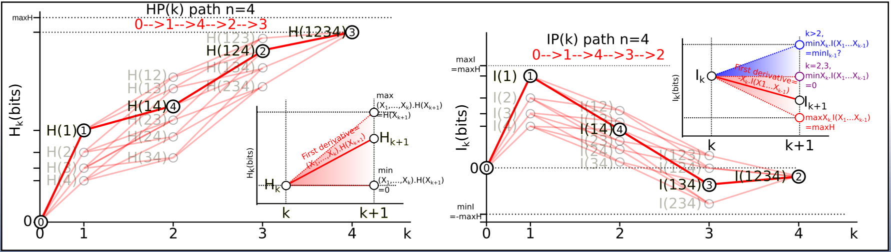
Derivatives bounds given by information inequalities and define cones (Yeung).
For entropy path we have:
determinism $0 \leq (X_1, ..., X_{k-1}).H(X_k) \leq H(X_k)$ independence.



Derivatives Bounds, Information inequalities

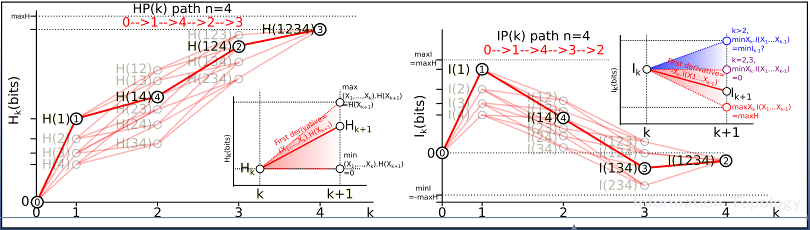
The bounds of mutual-information path are richer:

- For $k = 2$, $X_i.I(X_j) = X_i.H(X_j)$ and $0 \leq X_i.I(X_j) \leq I(X_j)$.
- For $k = 3$ the conditional mutual-information $0 \leq X_i.I(X_j; X_h) \leq \min(X_i.H(X_j), X_i.H(X_h))$ with right equality iff X_j and X_h are conditionally independent given X_i.



Derivatives Bounds, Information inequalities

- For $k > 3$, $X_k.I(X_1;..;X_{k-1})$ can be negative: $X_k.I(X_1;..;X_{k-1}) < 0$ iff $I_k < I_{k+1}$ (Matsuda). "Shannonian" inequalities: the set of inequalities that are obtained from conditional information positivity $X_i.I(X_j;X_h) \geq 0$ by linear combination, a convex "positive" cone after closure. Negativity gives "non-Shannonian" inequalities and cone (Yeung, Matus...).
  $\min I(X_1;..;X_{k-1})(?) \leq X_k.I(X_1;..;X_{k-1}) \leq \min_{i\in[k-1]} (X_k.H(X_i))$ with right equality given by the configurations for which the variable $X_1,..,X_{k-1}$ are equivalent $X_1 \sim ... \sim X_{k-1}$ when X_k is given.



Local minima and critical dimension

Lemma local minima of information paths

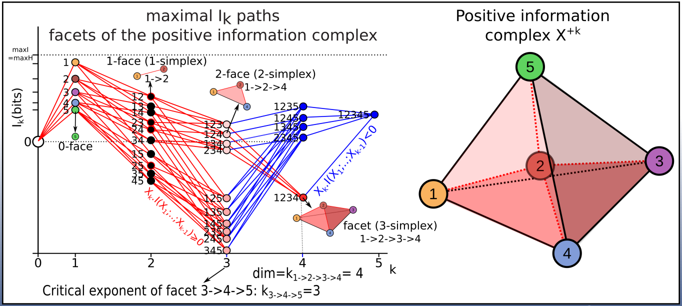
if $X_k.I(X_1;..;X_{k-1}) < 0$ then all paths from 0 to I_{k-1} have at least one critical point. The first critical point if it exists is a local minima. In order for an information path to have a critical point it is necessary that $k > 3$, the smallest possible degree of critical point being $k = 3$.

- **The first informational critical dimension** of the information path IP_i, noted k_{i_1} is the degree k of the first local minima of an information path.
- **Positive information path** is an information path from 0 to a given I_k corresponding to a given k-tuple of variable such that $I_k < I_{k-1} < ... < I_1$.
- **Maximal Positive information path** is a Positive information path of maximal length. More formally, a maximal positive information path is a positive information path that is not a proper subset of positive information path.

Minimum free energy complex

Theorem Minimum free energy complex

The set of all positive informations paths forms a simplicial complex, that we call the minimum free energy complex, noted $X^{+k_{i_1}}$. A necessary condition for this complex not to be a simplex is that its dimension $d \geq 4$.



Minimum free energy complex

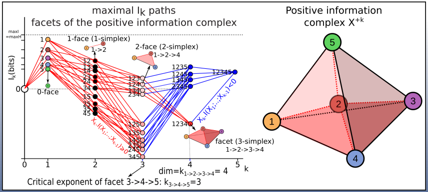
- Positive information path and maximal positive information path coincide with chain (face) and maximal chains (facet). The maximal faces encode all the structure of minimum free energy complex.
- **The dimension of the minimum free energy complex** is the maximum of the first informational critical dimension k_{i_1} if it exists or the dimension of the whole simplicial structure n.
- The set of first critical points of information paths would give a good description of the landscape, and of the complexity of the measured system. This complex is nothing but the formalization of **the minimum free energy principle in a degenerate case**.

Minimum free energy complex

- Without proof of its peculiar interest, we define the **minimum free energy characteristic** as:

$$H^{+k}(X^{+k};P) = \sum_{i=1}^{k} (-1)^{i-1} \sum_{I \subset X^{+}; card(I)=i} I_i(X_I;P)$$

- Analog to **paths sum**: it sums over all paths until they diverge; the divergence being the negativity of conditional mutual information.
- $H^{+k}(X^{+k}) = \sum^{5} I(X_i) - \sum^{10} I(X_i;X_j) + \sum^{10} I(X_i;X_j;X_h) - I(X_1;X_2;X_3;X4)$



Second law information topology

Random-stochastic process $\{X_t, t \in T\}$ is a collection of random variables on the same probability space $(\Omega, \mathcal{F}, P)$ and T is a totally ordered set.

Lemma stochastic processes - information paths:

Let (Ω, Δ^k, P) be a simplicial information structure, then the set of paths HP_k and IP_k are in one to one correspondence with the set of stochastic process $\{X_t, t \in T, |T| = k\}$

Theorem second Law (information topology)

Let (Ω, Δ^k, P) be a simplicial information structure, then the entropy of a stochastic process can only increase with time.

"You can't have something for nothing, not even an observation" Gabor.

Second law information topology

Theorem second Law (information topology)

Let (Ω, Δ^k, P) be a simplicial information structure, then the entropy of a stochastic process can only increase with time.

"You can't have something for nothing, not even an observation" Gabor.

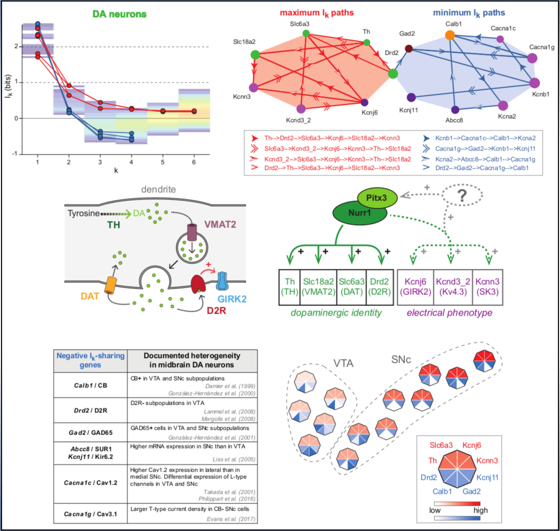
- The statement is equivalent to $H(X_1, ...X_k) \geq H(X_1, ...X_{k-1})$ which is a direct consequence of conditional entropy positivity and the chain rule of information with $k = -\ln 2$.
- Improves the result of Cover (1991) that assumes stationary Markov condition.
- Paths are automorphisms of $\{1, 2.., k\} = [k]$, initial minimal low entropy state $H(0) = 0$. No stationarity or ergodicity assumptions.

Computation of the Minimum free energy complex

- **Computational problem:** finding a global functional extrema or all the first critical dimension is NP-hard class ($\mathcal{O}(n!)$).
- **Computational solution:** At each element of the lattice, we start at one of the I_1 and at each element of the paths we explore only the two paths with lowest and highest positive values of $X_{k+1}.I(X1;..;X_k)$ (local), and iterate until it stops at the minima (whenever the conditional mutual information starts to be negative) and then rank the paths as a function of their length. It finds the maximal positive information paths that have highest and lowest I_k values at each element of a path. Computational complexity in $\mathcal{O}(n)$ but only give a **partial estimation of the minimum free energy complex** (can be richer and greater dimensionality).

DA Minimum free energy complex

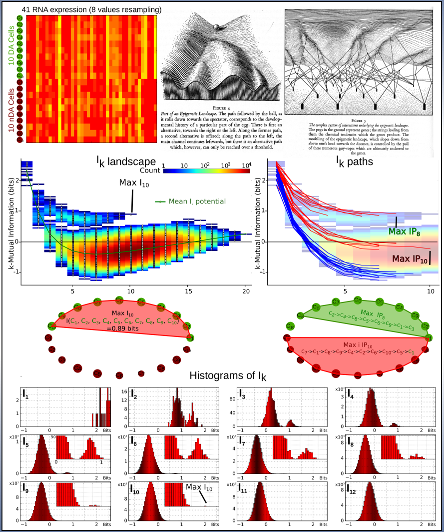
- Identifies functional module up to $k_i = 6$.
- Maximum path detect the metabolic chain of Dopamine, genes having common transcription regulators and unravel electrophysiological and neuromediator identity coupling.
- Minimum path detect heterogenity, suclasses and spatial differential expressions.



Negative I_k-sharing genes	Documented heterogeneity in midbrain DA neurons
Calb1 / CB	CB+ in VTA and SNc subpopulations *Damier et al. (1999)* *González-Hernández et al. (2000)*
Drd2 / D2R	D2R- subpopulations in VTA *Lammel et al. (2008)* *Margolis et al. (2008)*
Gad2 / GAD65	GAD65+ cells in VTA and SNc subpopulations *González-Hernández et al. (2001)*
Abcc8 / SUR1 *Kcnj11* / Kir6.2	Higher mRNA expression in SNc than in VTA *Liss et al. (2005)*
Cacna1c / Cav1.2	Higher Cav1.2 expression in lateral than in medial SNc. Differential expression of L-type channels in VTA and SNc *Takada et al. (2001)* *Philippart et al. (2016)*
Cacna1g / Cav3.1	Larger T-type current density in CB- SNc cells *Evans et al. (2017)*

Neuronal Minimum free energy complex

- Transpose matrix (egoist genes). $m = 41$ genes, $n = 20$ cells
- Preidentified 10 DA and nDA neurons.

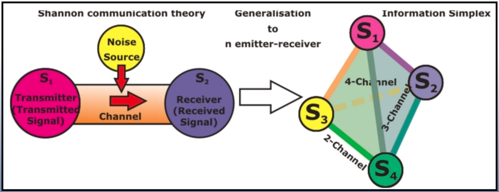


Information theory

- Landscape represents and implements all classical information functions, chain rules and inequalities: easy tool.
- Non-shannonian inequalities (cone) are related to the existence of critical point in information path.
- New methods for topological and statistical data analysis, with totally opensource tools (the python program allowing all quantifications and representations is available on Github)

Information theory

- The global picture: information communication is only partially accounted by pairwise exchange of information, formalized by a communication channel, that is a 1-simplex between two variables, the emitter and the receiver. By considering n emitters/receivers and defining k-communication channels as the k-face of a simplicial structure, with respective capacity $\max(I_k)$, the present topological formalism gives very preliminary basement for such a generalized communication theory. Moreover, it suggests refined data compression algorithm.



Statistical physic

- At least in genetic expression, but we propose that it is a generic feature of biological structures, high order than pairwise statistical interaction exist, can be non negligible, and moreover can be combinatorially numerous.
- Clustering of data points analog to matter condensation, a simple picture.
- Topological and informational formalization of the Potts model, negativity signature of frustration, multiplicity of local minima.
- mean information path is analog to DFT treatment of the n-body problem, but the formalism here is different, it is finite and discrete, it computes the cohomology group of measurable function, do not assume any metric (like an interaction distance r), nor Hamiltonian or Lagrangian structure, symplectic or contact structure, configuration or phase space (etc.). The main difference with classical statistical physic determinations of free energy and entropy is the absence of predefined metric and the finiteness-discreteness of the formalism (no assymptotic limit, no Stirling approximation).

Statistical physic

- Our theorem applied to 3n dimensions of a configuration space (like in DFT) implies that whereas the minimum free information energy complex of an elementary body can only be a simplex, the configuration space of n elementary body can be a complex with quite arbitrary topology (possible heterogeneity at large "scales").
- What should be done next: discrete analog of Noether theorem.

Ecology - Biology - Complex systems

- **Ecology** is the scientific analysis and study of interactions among organisms and their environment...
- Biology and ecology: the main interest of the present formalism is to capture and identify diversity, while yet allowing selectivity. It gives a quantitative framework the cellular identity and its differentiation.
- From complex network to ... complex.

Thank you! Thank Jean-marc, Monica, Daniel ... UNIS1072 inserm, ERC Chanelomics

Questions

Appendices

First appendix

Second appendix