Exploring and exploiting intrinsic synergies of multivariate systems

Fernando E. Rosas

Research Fellow

Centre of Complexity Science

Department of Mathematics

Department of Electrical and Electronic Engineering

Imperial College London

Acknowledges:

This work is the result of the fortunate collaboration with:

Pedro A.M. Mediano
Department of Computing
Imperial College London

Henrik J. Jensen
Department of Mathematics
Imperial College London

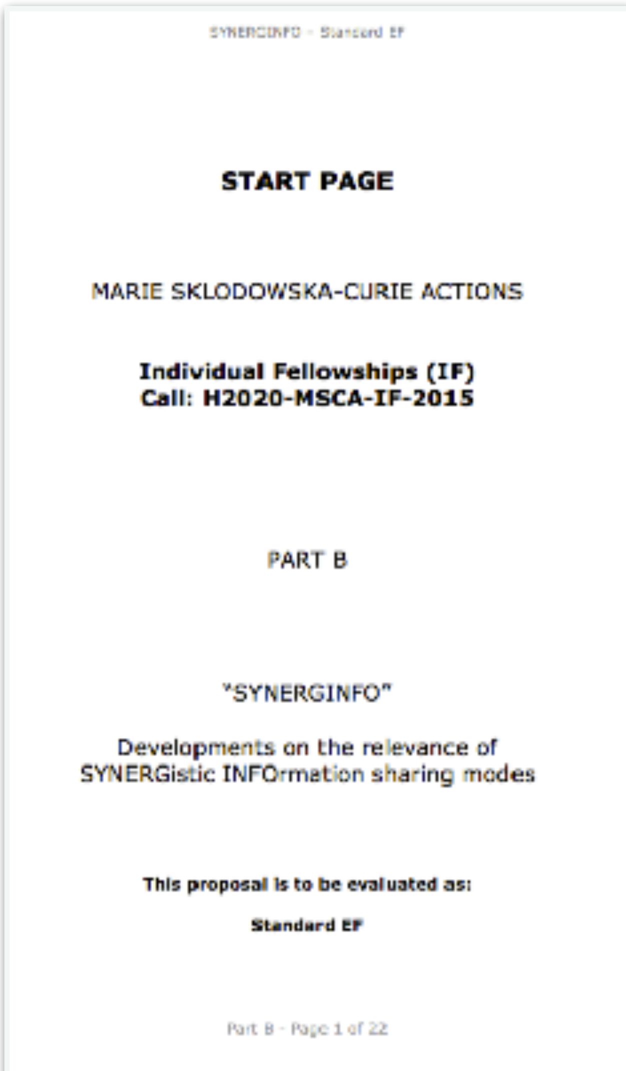
Michael Gastpar
School of Computer and Communication Sciences
EPFL

Martin Ugarte
Department of Computer Science
Université Libre Bruxelles

Borzoo Rassouli
School of Computer Science and Electronic Engineering
University of Essex

Deniz Gündüz
Department of E&E Engineering
Imperial College London

Acknowledges: Marie Slodowska-Curie Actions


SYNERGINFO - Standard EF

START PAGE

MARIE SKLODOWSKA-CURIE ACTIONS

Individual Fellowships (IF)
Call: H2020-MSCA-IF-2015

PART B

"SYNERGINFO"

Developments on the relevance of
SYNERGistic INFOrmation sharing modes

This proposal is to be evaluated as:

Standard EF

Part B - Page 1 of 22

Objectives:

(a) to understand and exploit **intrinsic statistical synergies** of multivariate systems

(b) apply this to **data privacy** and **neural data analysis**

Today's menu
1. Self-organisation
2. O-information
3. Data privacy
4. Summary and current work

1. Self-organisation: what is a pattern

What is a pattern/structure?

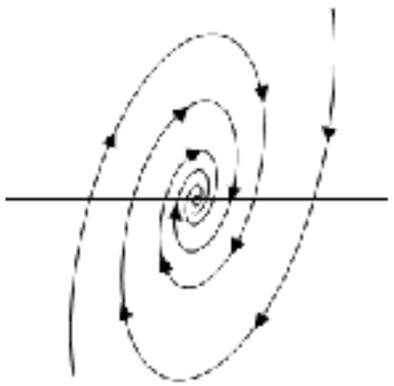
Shannon —> *compressibility of an statistical source*
Regularities / interdepedencies (i.e. deviations from statistical independence)

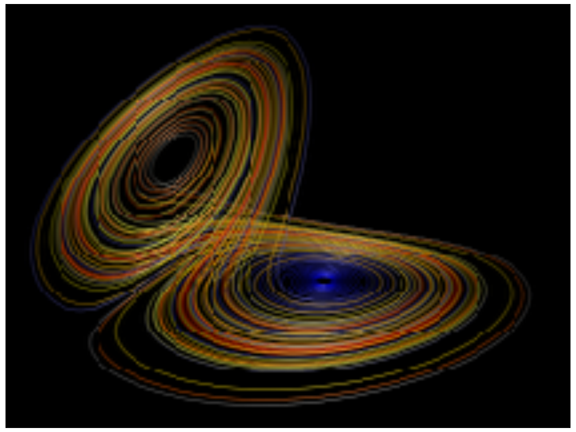
Organisation as statistical interdependency

1. What is self-organization?

Attractor: destination/result of the evolution

Linear recurrence / evolution is simple ⟷ simple/uninteresting attractors
Non-linear evolution ⟷ interesting/strange attractors





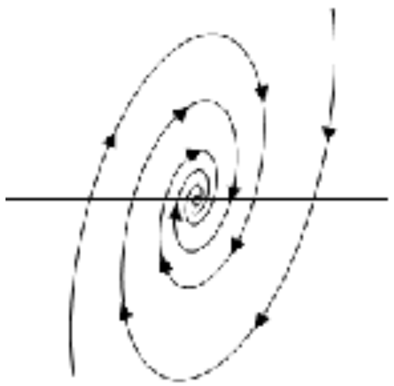
Challenge:
*Is it possible to relate the **attractors** properties to **organization** properties?*

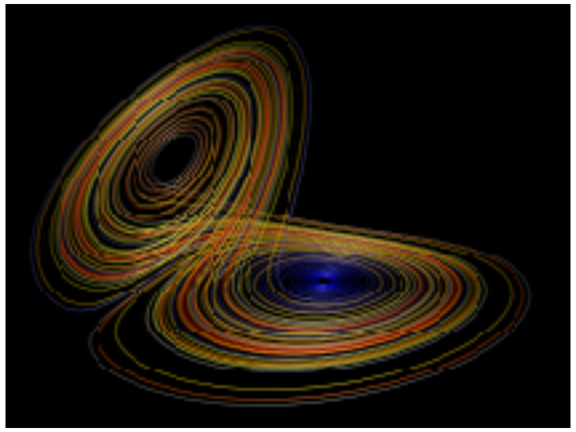
1. What is self-organization?

Attractor: destination/result of the evolution

Linear recurrence / evolution is simple ⟷ simple/uninteresting attractors

Non-linear evolution ⟷ interesting/strange attractors





Idea: high-order correlations allow to distinguish qualitatively different arrangements…

redundancy!

synergy!

1. Self-organisation: our approach

Approach to study spontaneous creation of correlations:

— Consider an flat initial distribution μ_0 (independence)

— Compute it's time evolution (μ_t) using the corresponding "master equation"

— Iterate for many steps, or until it reach a stationary state $\mu_{t_{\mathrm{st}}+1} = \mu_{t_{\mathrm{st}}}$.

— Analyse the "structure of the correlations" of $(X^1_{t_{\mathrm{st}}}, \ldots, X^N_{t_{\mathrm{st}}})$.

Why start from an uniform distribution?

Because any correlation found is due to the evolution and not the initial condition.

1. Self-organisation: dynamics as sculpture…

Φ_t

Φ'_t

"The Atlas"
Michelangelo
Circa 1530-1534

"The Awakening Slave"
Michelangelo
Circa 1520-1523

1. Self-organisation: entropy is not enough

Is easy to prove that, because of determinism, the joint Shannon entropy of the system is non-increasing.

$$H(X_{t+1}^1, \ldots, X_{t+1}^N) \leq H(X_t^1, \ldots, X_t^N)$$

Information is being dissipated

It takes more information to specify a random state than a point in an attractor

This not sufficient to guarantee self-organization!! (example: fix point attractors)

1. Self-organisation: entropy is not enough

Is easy to prove that, because of determinism, the joint Shannon entropy of the system is non-increasing.

$$H(X^1_{t+1},\dots,X^N_{t+1}) \le H(X^1_t,\dots,X^N_t)$$

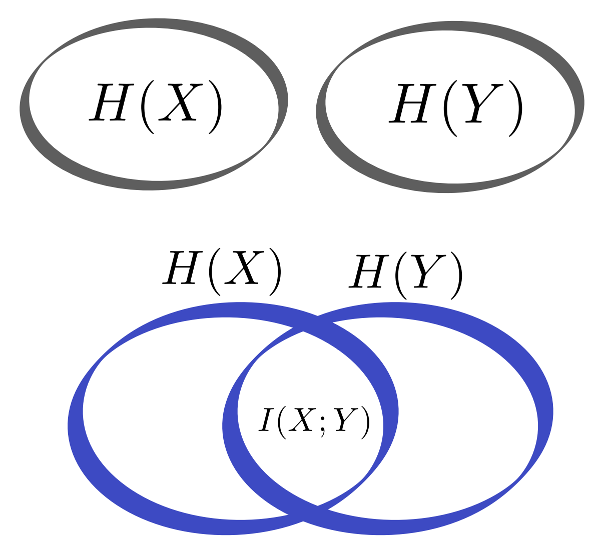
Information is being dissipated

It takes more information to specify a random state than a point in an attractor

This not sufficient to guarantee self-organization!! (example: fix point attractors)

Key idea:

The destruction of information can create correlations!!



1. Self-organisation: proposed framework

Our proposed total entropy decomposition:

"Residual entropy"

$$R(\mathbf{X}_t) := \sum_{j=1}^{N} H(X_t^j | \mathbf{X}_t^{-j})$$

"Binding entropy"

$$B(\mathbf{X}_t) := H(\mathbf{X}_t) - R(\mathbf{X}_t) = H(\mathbf{X}_t) - \sum_{j=1}^{N} H(X_t^j | \mathbf{X}_t^{-j})$$

Definition 1. *Let us consider a coupled dynamical system whose evolution is autonomous and whose matrix of interactions has a bound over the number of non-zero elements per row. Then, the system is self-organising if $B(\mathbf{X}_t)$ is an increasing function of t. Moreover, the value of $B(\mathbf{X}_t)$ is used as a metric of organisation strength.*

F. Rosas, P.A. M. Mediano, M. Ugarte, H.J. Jensen, "*An information-theoretic approach to self-organization: emergence of complex interdependencies in coupled dynamical systems*", Entropy 20, no. 10 (2018): 793.

1. Self-organisation: proposed framework

Decomposition by ***sharing modes:***

$$B(\mathbf{X}_t) = \sum_{n=2}^{N} \sum_{i=1}^{n-1} I_i^n(\mathbf{X}_t) = \sum_{i=1}^{N-1} m_i(\mathbf{X}_t)$$

by measuring *i* agents
can guess the state of *n*…

1. Self-organisation: proposed framework

*Decomposition by **sharing modes:***

$$B(\mathbf{X}_t) = \sum_{n=2}^{N} \sum_{i=1}^{n-1} I_i^n(\mathbf{X}_t) = \sum_{i=1}^{N-1} m_i(\mathbf{X}_t)$$

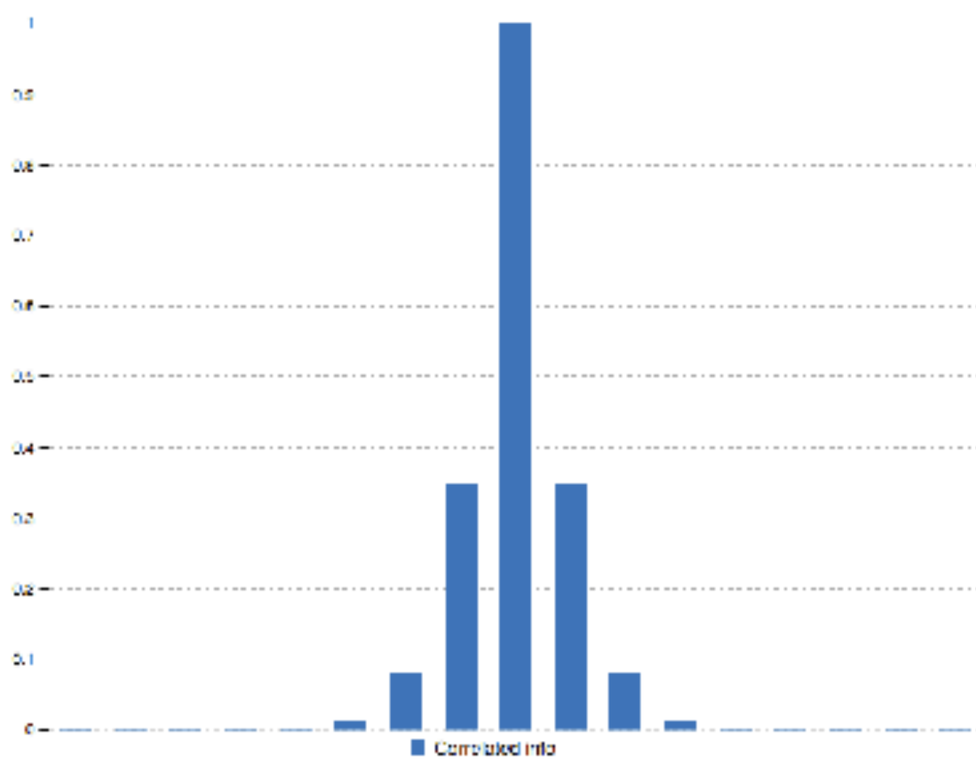
$$\psi_L(t) \leq \sum_{i=1}^{L} m_i(\mathbf{X}_t) \leq \sum_{j=1}^{N} \sum_{\substack{\boldsymbol{\alpha}_L \in \mathcal{I}_L \\ \alpha_i \neq j}} I(X_{\boldsymbol{\alpha}_L}(t); X_j(t)) \leq N \binom{N-1}{L} \psi_L(t)$$

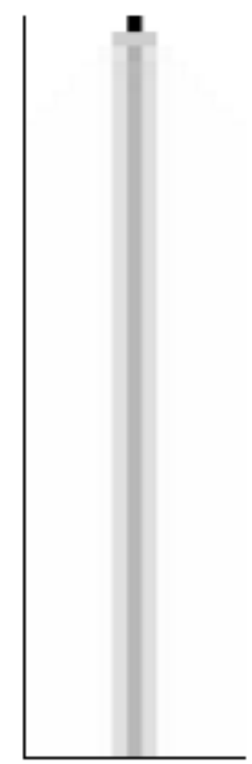
where

$$\psi_L(t) := \max_{j \in \{1,\ldots,N\}} \max_{\substack{\boldsymbol{\alpha}_L \in \mathcal{I}_L \\ \alpha_i \neq j}} I(X_{\boldsymbol{\alpha}_L}(t); X_j(t))$$

1. Self-organisation: proof of concept

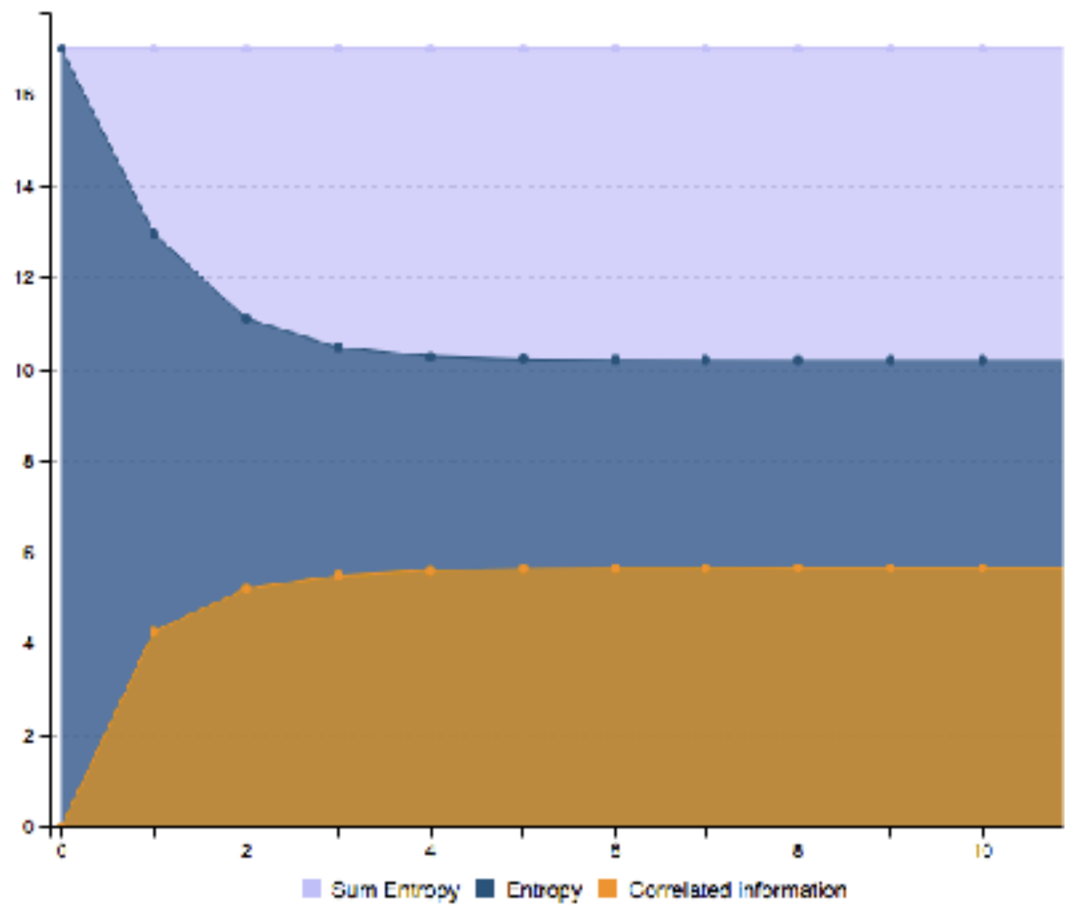
Rule 232

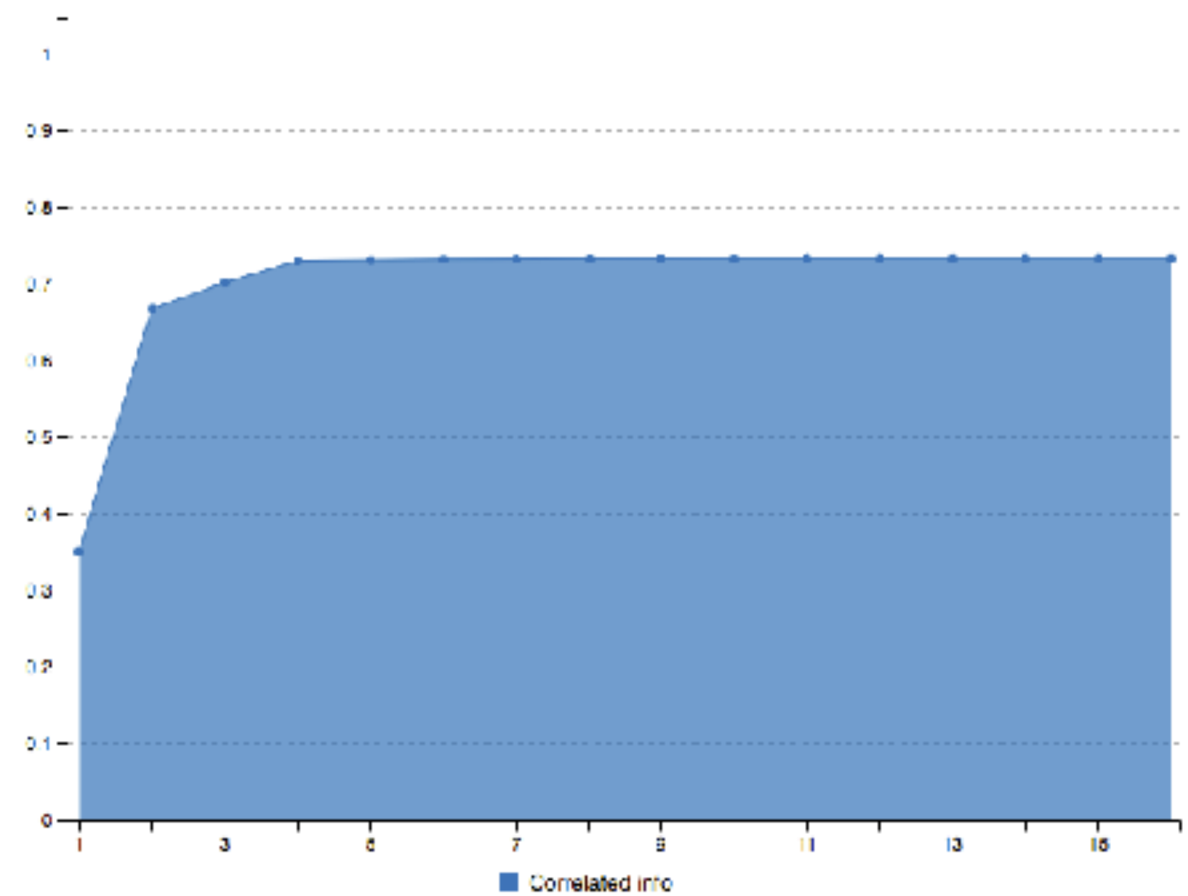




1. Self-organisation: proof of concept

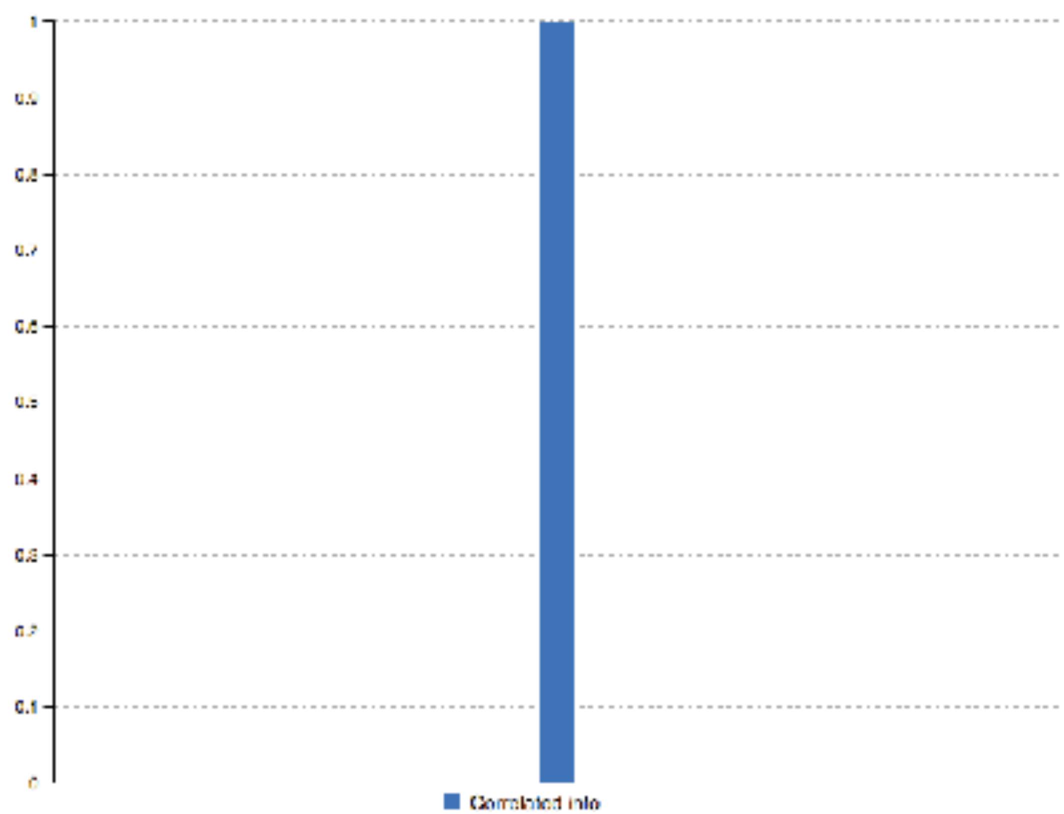
Rule 232

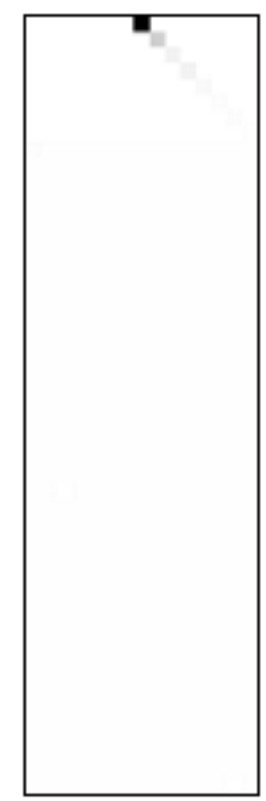




1. Self-organisation: proof of concept

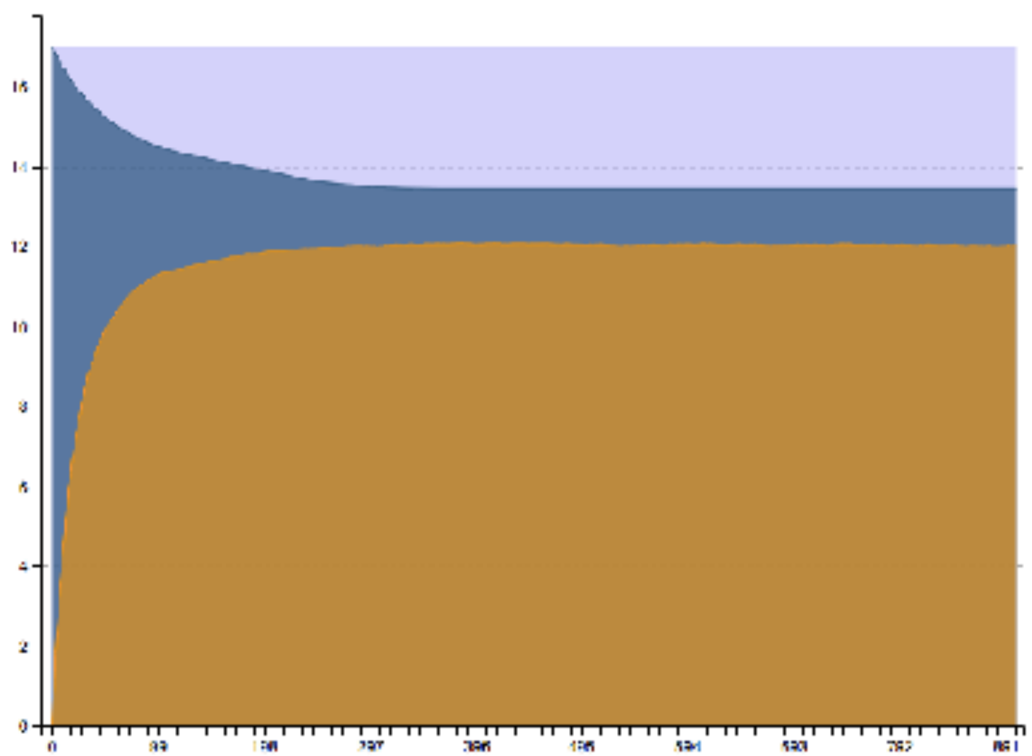
Rule 30

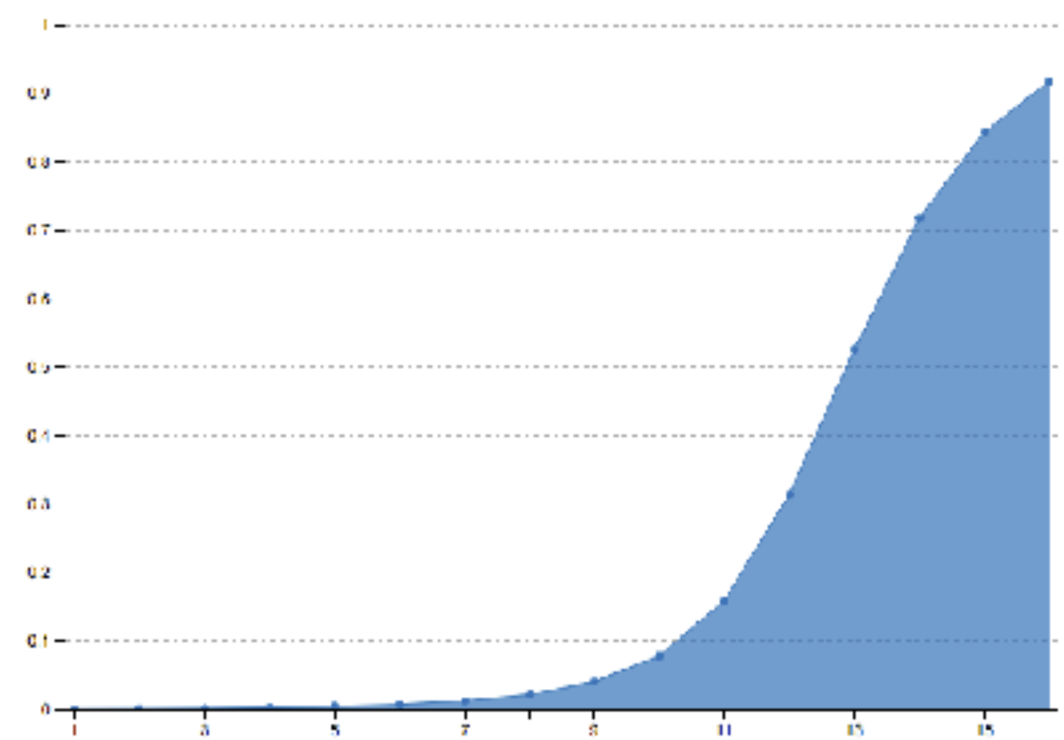




1. Self-organisation: proof of concept

Rule 30



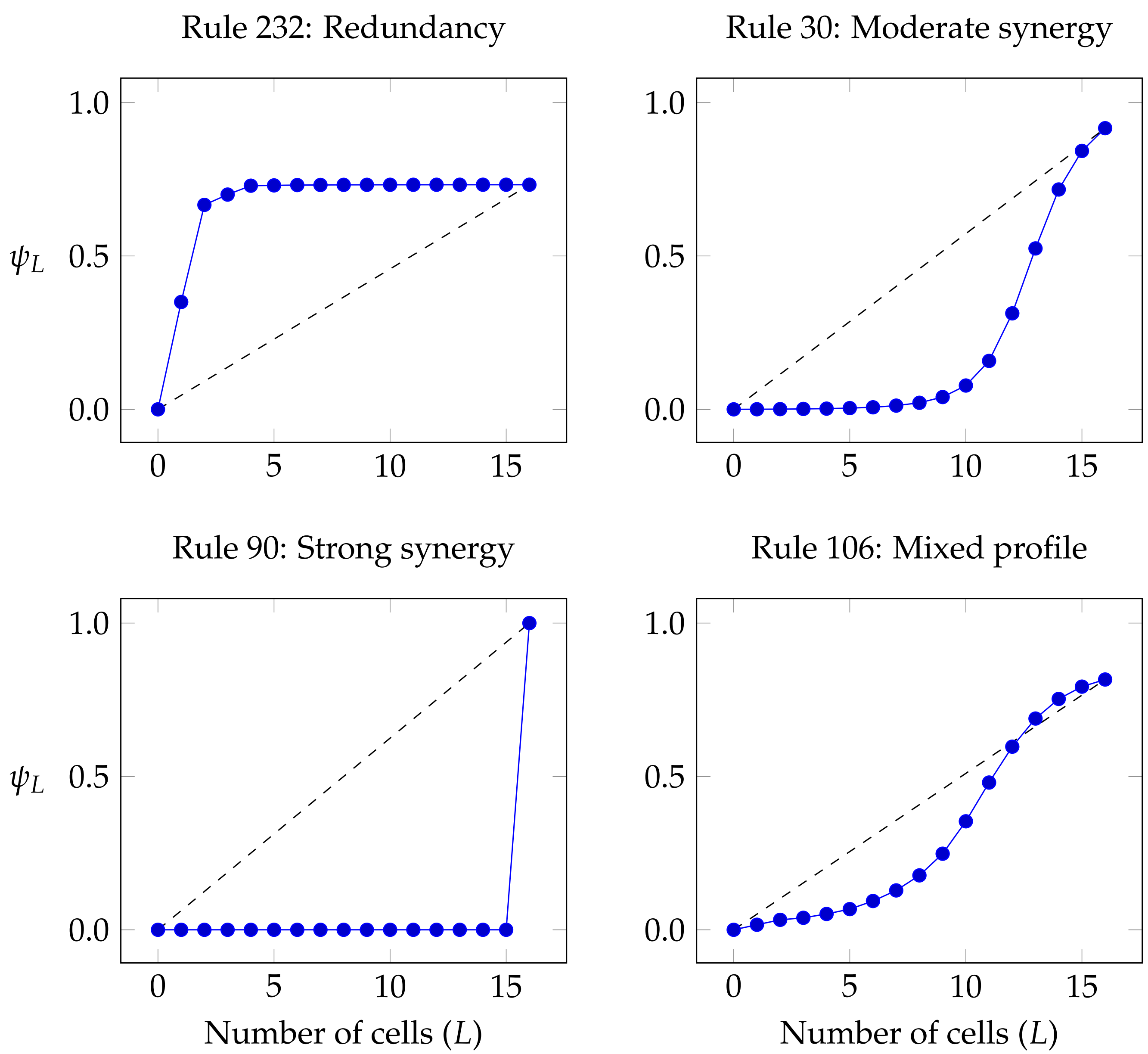


1. Self-organisation: proof of concept

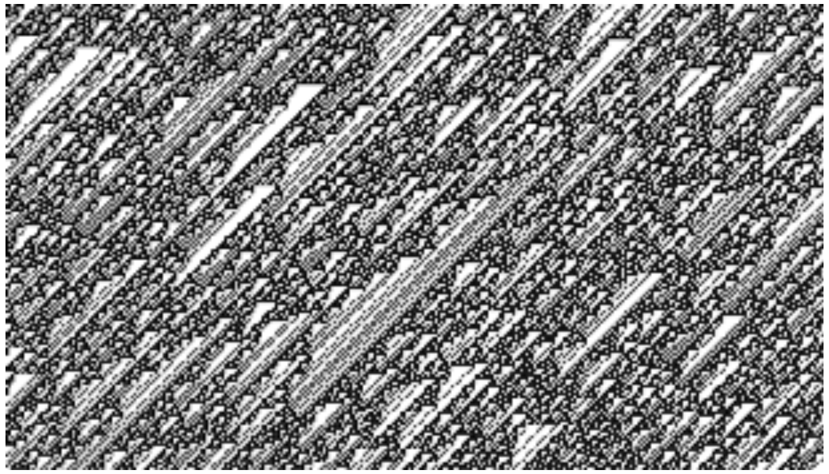
Definition 1. *Let us consider a coupled dynamical system whose evolution is autonomous and whose matrix of interactions has a bound over the number of non-zero elements per row. Then, the system is self-organising if* $B(\mathbf{X}_t)$ *is an increasing function of* t. *Moreover, the value of* $B(\mathbf{X}_t)$ *is used as a metric of organisation strength.*

Definition 2. *A self-organisating process is said to be synergy-dominated if, for a large* t, $\psi_L(t)$ *is convex as function of* L. *If* $\psi_L(t)$ *is concave, then the process is said to be redundancy-dominated.*

1. Self-organisation: proof of concept



Rule 106



Today’s menu
1. Self-organisation
2. O-information
3. Data privacy
4. Summary and current work

2. O-information: fundamentals

1. **Interaction-information:** $I(X_1; X_2; \ldots; X_n) = -\sum_{\gamma \subseteq \{1,\ldots,n\}} (-1)^{|\gamma|} H(\boldsymbol{X}^{\gamma})$

2. **Total correlation:** $$\mathrm{TC}(\boldsymbol{X}^n) = \sum_{j=1}^{n} H(X_j) - H(\boldsymbol{X}^n)$$

3. **Dual total correlation:** $$\mathrm{DTC}(\boldsymbol{X}^n) = H(\boldsymbol{X}^n) - \sum_{j=1}^{n} H(X_j|\boldsymbol{X}^n_{-j})$$

State of affairs:

(+) **TC** and **DTC** are metrics of global correlation strength.

(-) **TC=0** if and only if **DTC=0**. Besides that**, their relationship is unclear**.

(+) $I(X_1; X_2; X_3) = I(X_1; X_3) + I(X_2; X_3) - I(X_1 X_2; X_3)$
is positive for redundant systems, and negative for synergistic ones.

(+) $I(X_1; X_2; X_3)$ **= TC - DTC .**

(-) The meaning of the **interaction-information for *n* > 3 is unclear**.

2. O-information: fundamentals

1. **Interaction-information:** $I(X_1; X_2; \ldots; X_n) = -\sum_{\boldsymbol{\gamma} \subseteq \{1,\ldots,n\}} (-1)^{|\gamma|} H(\boldsymbol{X}^{\gamma})$

2. ***Total correlation:*** $\mathrm{TC}(\boldsymbol{X}^n) = \sum_{j=1}^{n} H(X_j) - H(\boldsymbol{X}^n)$

3. **Dual total correlation:** $\mathrm{DTC}(\boldsymbol{X}^n) = H(\boldsymbol{X}^n) - \sum_{j=1}^{n} H(X_j | \boldsymbol{X}^n_{-j})$

TC *or* **DTC**?
that is the question…

How could I "fix"
the **interaction-info** for
any *n* ??

The information-theoretic Hamlet

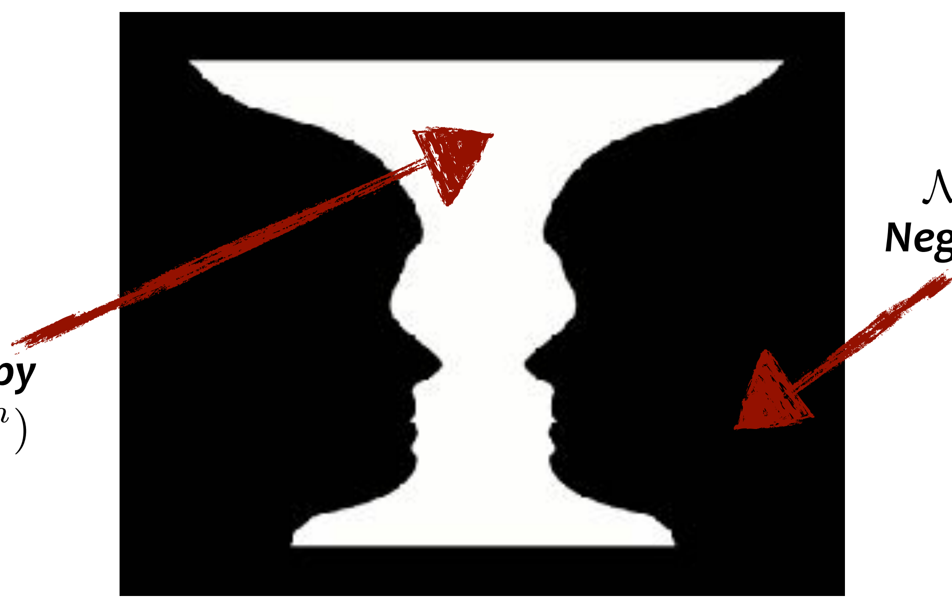
2. O-information: the two faces of interdependency

Consider a system $\boldsymbol{X}^n = (X_1, \dots, X_n)$ where $X_j \in \mathcal{X}_j$, described by a p.d.f. $p_{\boldsymbol{X}^n}(\boldsymbol{x}^n)$.

$$\sum_{j=1}^{n} \log |\mathcal{X}_j|$$

$$\mathcal{N}(\boldsymbol{X}^n) = \sum_j \log |\mathcal{X}_j| - H(\boldsymbol{X}^n)$$

$$H(\boldsymbol{X}^n) = -\sum_{\boldsymbol{x}^n} p_{\boldsymbol{x}^n} \log p_{\boldsymbol{x}^n}$$

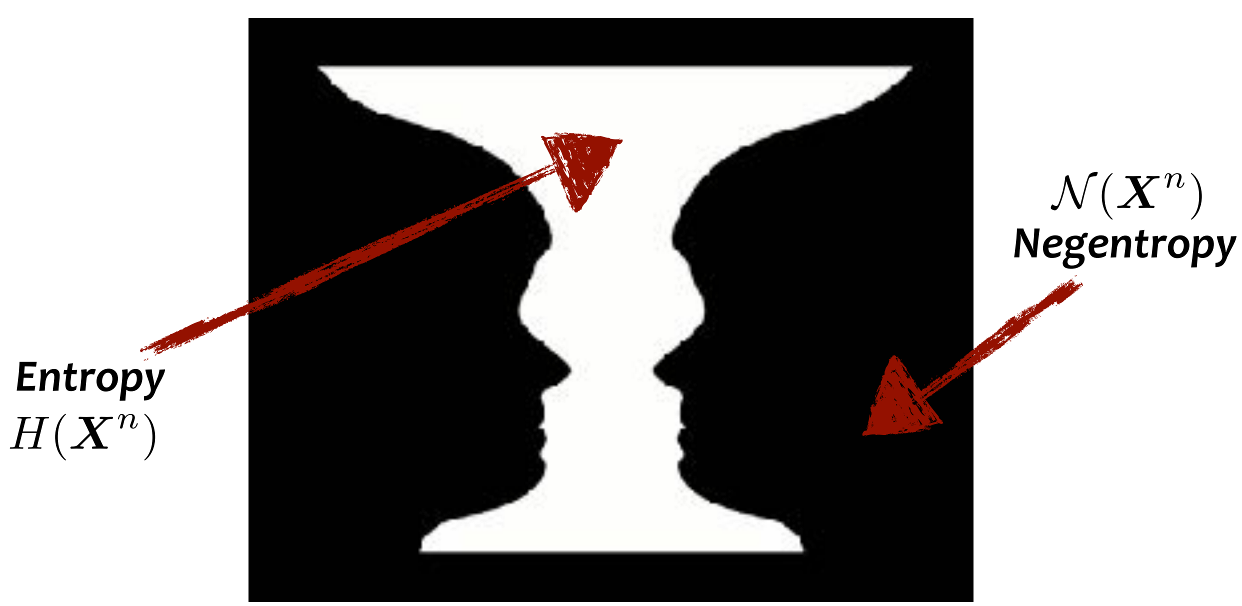


$\mathcal{N}(\boldsymbol{X}^n)$
Negentropy

Entropy
$H(\boldsymbol{X}^n)$

2. O-information: the two faces of interdependency

Consider a system $\boldsymbol{X}^n = (X_1, \dots, X_n)$ where $X_j \in \mathcal{X}_j$, described by a p.d.f. $p_{\boldsymbol{X}^n}(\boldsymbol{x}^n)$.



Example:

$X_1, X_2 \in \{0, 1\}$

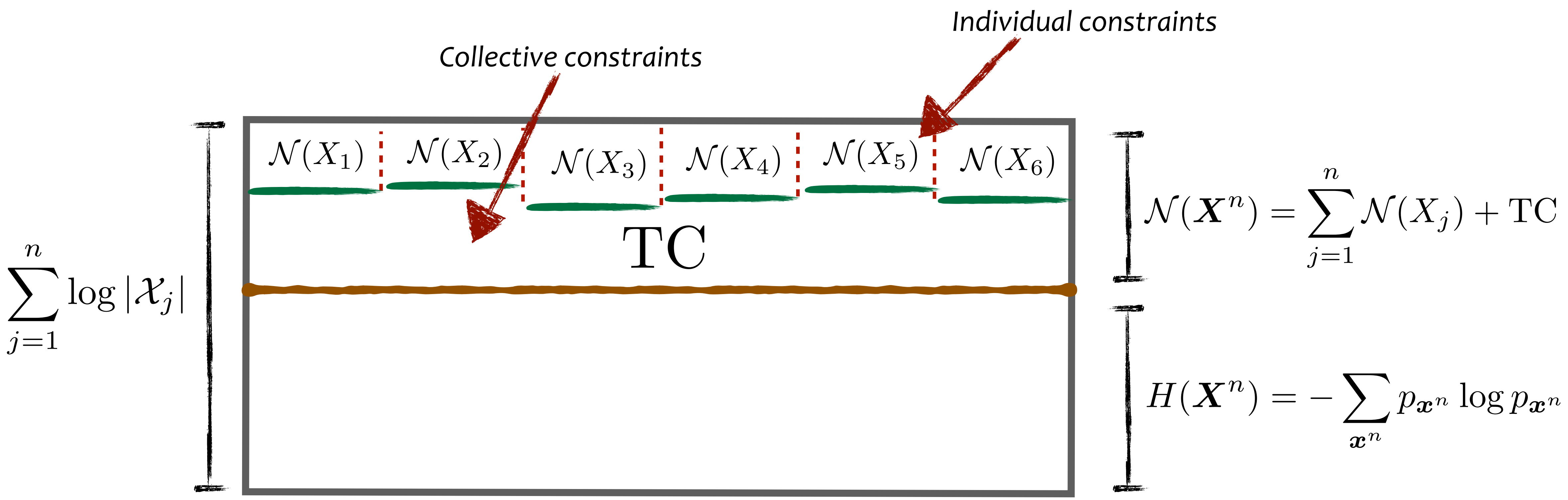
$p_{X_1X_2}(0,1) = p_{X_1X_2}(1,0) = 1/2$

$p_{X_1X_2}(0,0) = p_{X_1X_2}(1,1) = 0$

$H(X_1X_2) = \mathcal{N}(X_1X_2) = 1$

~~$(0,0)$~~	$(0,1)$
$(1,0)$	~~$(1,1)$~~

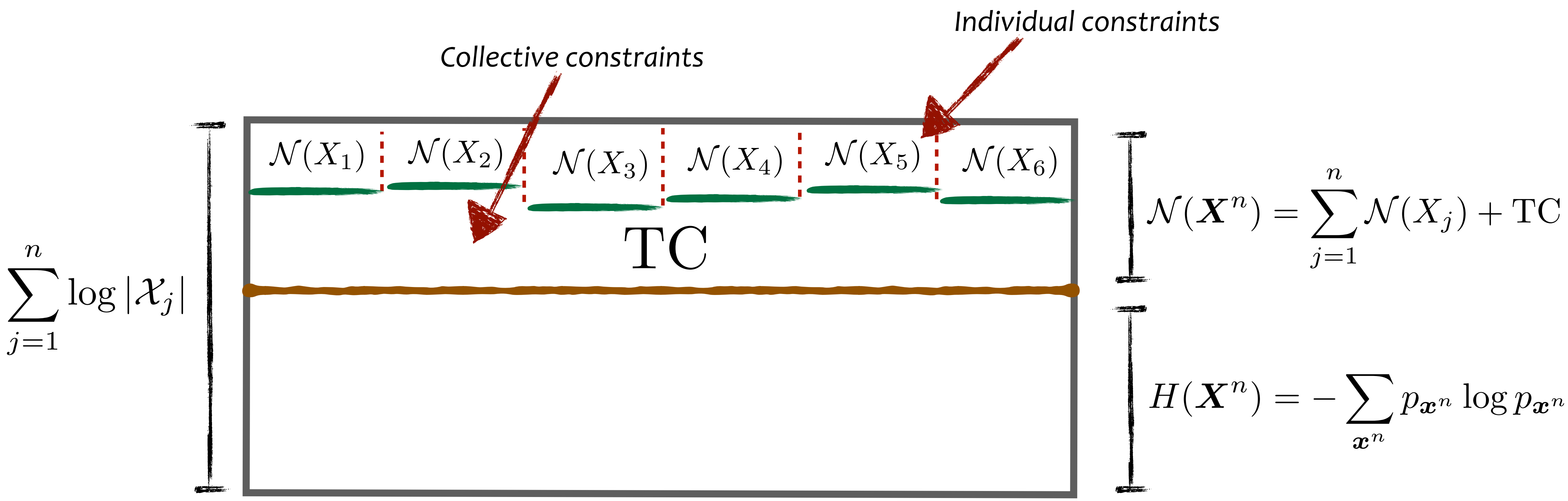
2. O-information: the two faces of interdependency



Marginal negentropy: $\mathcal{N}(X_j) = \log |\mathcal{X}_j| - H(X_j)$

Collective constraints: $\mathrm{TC}(\boldsymbol{X}^n) = \mathcal{N}(\boldsymbol{X}^n) - \sum_{j=1}^{n} \mathcal{N}(X_j) = \sum_{j=1}^{n} H(X_j) - H(\boldsymbol{X}^n)$

2. O-information: the two faces of interdependency



Example:

~~$(0,0)$~~	$(0,1)$
$(1,0)$	~~$(1,1)$~~

$\mathcal{N}(X_1 X_2) = 1$

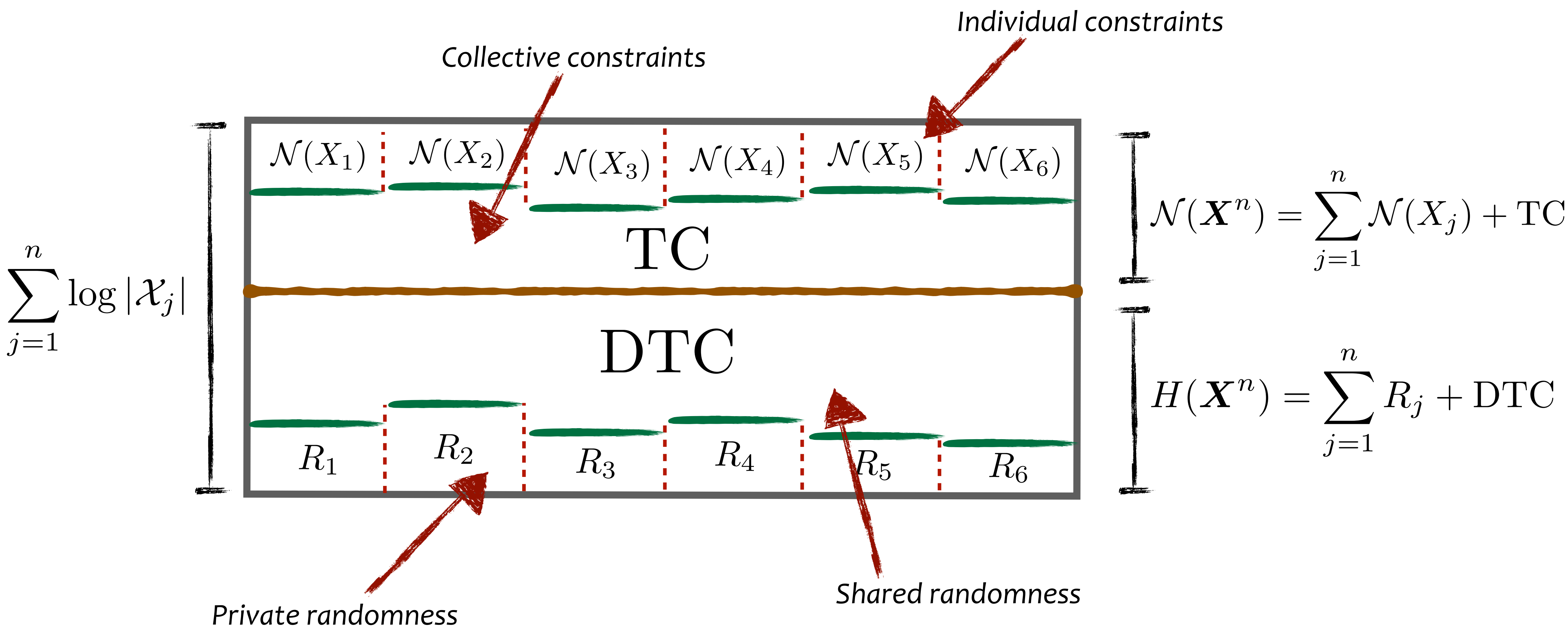
$\mathrm{TC}(X_1 X_2) = 1$

~~$(0,0)$~~	~~$(0,1)$~~
$(1,0)$	$(1,1)$

$\mathcal{N}(X_1 X_2) = 1$

$\mathrm{TC}(X_1 X_2) = 0$

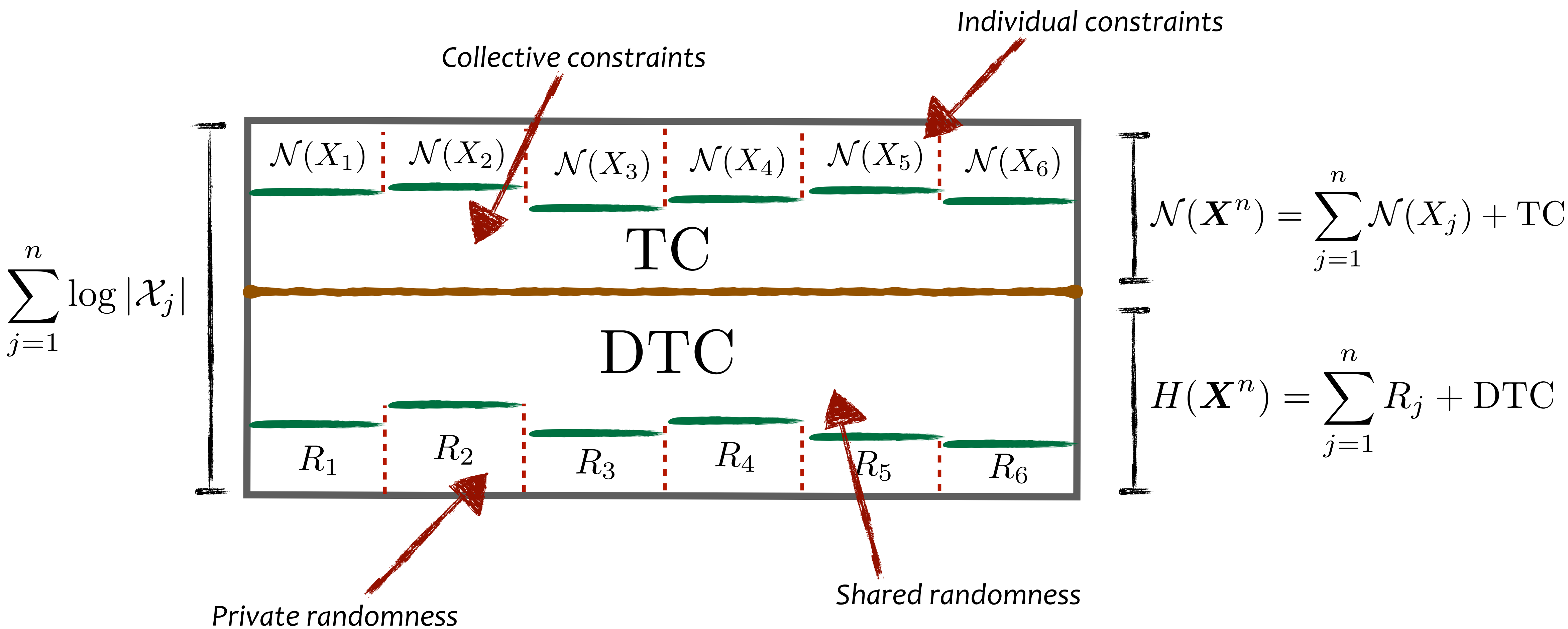
2. O-information: the two faces of interdependency



Residual entropies: $R_j = H(X_j | \boldsymbol{X}^n_{-j})$

Binding entropy: $\mathrm{DTC}(\boldsymbol{X}^n) = H(\boldsymbol{X}^n) - \sum_{j=1}^{n} H(X_j | \boldsymbol{X}^n_{-j})$

2. O-information: the two faces of interdependency



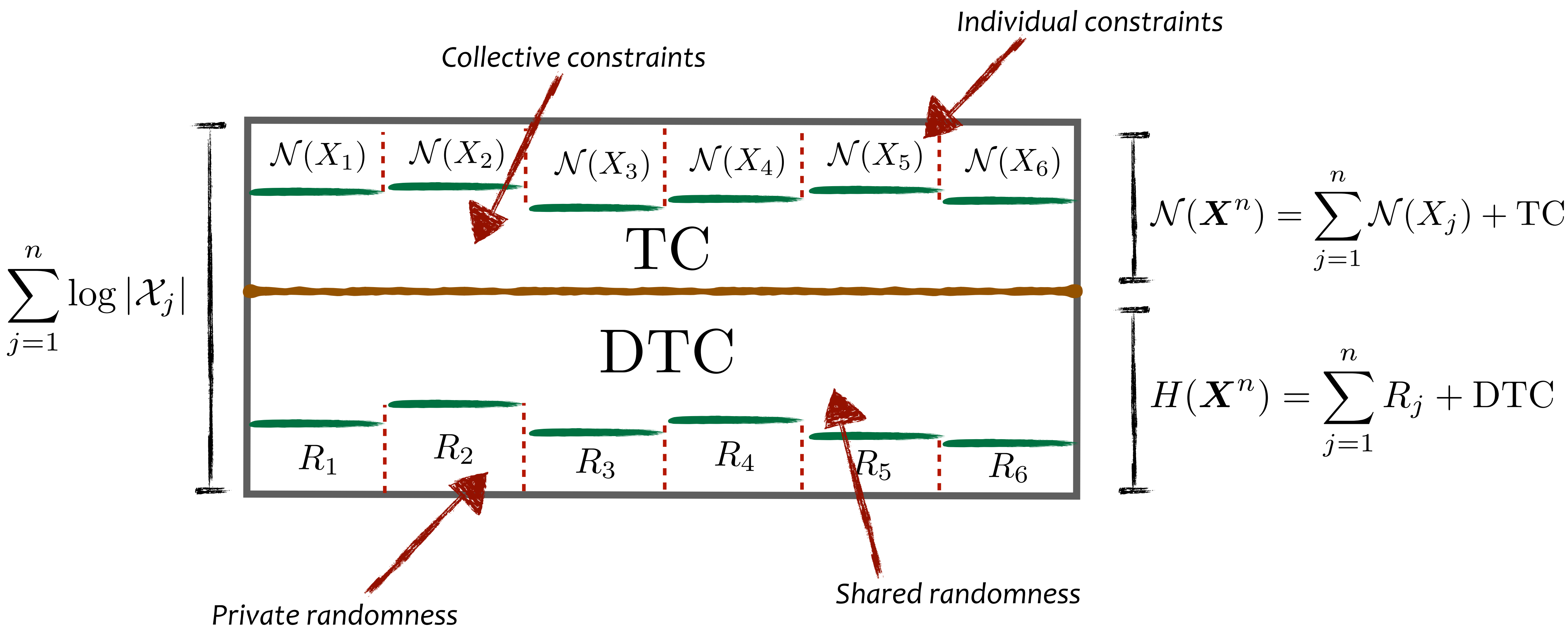
Example:

$$X_1 = X_2 = X_3$$

$$\mathrm{TC} = 2 > \mathrm{DTC} = 1$$

redundancy!

2. O-information: the two faces of interdependency



Example:

$X_1 = X_2 = X_3$	$X_3 = X_1 \oplus X_2$
$\mathrm{TC} = 2 > \mathrm{DTC} = 1$	$\mathrm{TC} = 1 < \mathrm{DTC} = 2$
redundancy!	***synergy!***

2. O-information: definition

Occam's raizor (*lex parsimoniae*): give preference to the simplest description

Definition

O-information: $\Omega(\boldsymbol{X}^n) = \mathrm{TC}(\boldsymbol{X}^n) - \mathrm{DTC}(\boldsymbol{X}^n)$

- If $\Omega(\boldsymbol{X}^n) > 0$ it is shorter to describe the allowed states.
- If $\Omega(\boldsymbol{X}^n) < 0$ it is shorter shorter to describe the constraints.

F. Rosas, P.A. M. Mediano, M. Gastpar, and H. J. Jensen. "Quantifying High-order Interdependencies via Multivariate Extensions of the Mutual Information," accepted, to be published in PRE, 2019

2. O-information: definition

Occam's raizor (*lex parsimoniae*): give preference to the simplest description

Definition

O-information: $\Omega(\boldsymbol{X}^n) = \mathrm{TC}(\boldsymbol{X}^n) - \mathrm{DTC}(\boldsymbol{X}^n)$

- If $\Omega(\boldsymbol{X}^n) > 0$ it is shorter to describe the allowed states.
- If $\Omega(\boldsymbol{X}^n) < 0$ it is shorter shorter to describe the constraints.

Example:

$X_1 = X_2 = \cdots = X_n$	$X_n = X_1 \oplus X_2 \oplus \cdots \oplus X_{n-1}$
$\mathrm{TC} = n - 1 > \mathrm{DTC} = 1$	$\mathrm{TC} = 1 < \mathrm{DTC} = n - 1$
redundancy!	***synergy!***

F. Rosas, P.A. M. Mediano, M. Gastpar, and H. J. Jensen. "Quantifying High-order Interdependencies via Multivariate Extensions of the Mutual Information," accepted, to be published in PRE, 2019

2. O-information: definition

Occam's raizor (*lex parsimoniae*): give preference to the simplest description

Definition

O-*information:* $\Omega(\boldsymbol{X}^n) = \mathrm{TC}(\boldsymbol{X}^n) - \mathrm{DTC}(\boldsymbol{X}^n)$

- If $\Omega(\boldsymbol{X}^n) > 0$ it is shorter to describe the allowed states.
- If $\Omega(\boldsymbol{X}^n) < 0$ it is shorter shorter to describe the constraints.

Definition

1. A system is *redundancy-dominated* if $\Omega(\boldsymbol{X}^n) \geq 0$
2. A system is *synergy-dominated* if $\Omega(\boldsymbol{X}^n) \leq 0$

2. O-information: definition

Occam's raizor (*lex parsimoniae*): give preference to the simplest description!

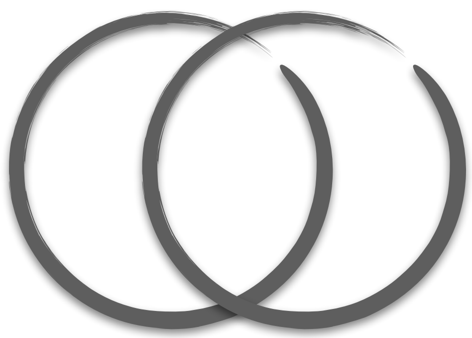
Definition

O-information: $\Omega(\boldsymbol{X}^n) = \mathrm{TC}(\boldsymbol{X}^n) - \mathrm{DTC}(\boldsymbol{X}^n)$

- If $\Omega(\boldsymbol{X}^n) > 0$ it is shorter to describe the allowed states.
- If $\Omega(\boldsymbol{X}^n) < 0$ it is shorter shorter to describe the constraints.

Note:

1. **n=2:** $\Omega(X_1 X_2) = 0$, because $\mathrm{TC}(X_1 X_2) = \mathrm{DTC}(X_1 X_2) = I(X_1; X_2)$
   i.e. shared randomness is equal to predictability



2. O-information: definition

Occam's raizor (*lex parsimoniae*): give preference to the simplest description!

Definition

O-information: $\Omega(\boldsymbol{X}^n) = \mathrm{TC}(\boldsymbol{X}^n) - \mathrm{DTC}(\boldsymbol{X}^n)$

- If $\Omega(\boldsymbol{X}^n) > 0$ it is shorter to describe the allowed states.
- If $\Omega(\boldsymbol{X}^n) < 0$ it is shorter shorter to describe the constraints.

Note:

1. **n=2:** $\Omega(X_1 X_2) = 0$, because $\mathrm{TC}(X_1 X_2) = \mathrm{DTC}(X_1 X_2) = I(X_1; X_2)$
   i.e. shared randomness is equal to predictability

2. **n=3:** $\Omega(X_1 X_2 X_3) = I(X_1; X_2) - I(X_1; X_2 | X_3) =: I(X_1; X_2; X_3)$ —> interaction-information!

2. O-information: (1) it is sum of triple interaction-informations

- Decompositions for the TC, DTC and o-information:

$$C(\boldsymbol{X}^n) = \sum_{i=2}^{n} I(X_i; \boldsymbol{X}^{i-1}) \ ,$$

$$B(\boldsymbol{X}^n) = I(X_n; \boldsymbol{X}^{n-1}) + \sum_{j=2}^{n-1} I(X_j; \boldsymbol{X}^{j-1} | \boldsymbol{X}^n_{j+1}),$$

$$\Omega(\boldsymbol{X}^n) = \sum_{k=2}^{n-1} I(X_k; \boldsymbol{X}^{k-1}; \boldsymbol{X}^n_{k+1}) \ .$$

O-information is an aggregation of triple multi-informations!!

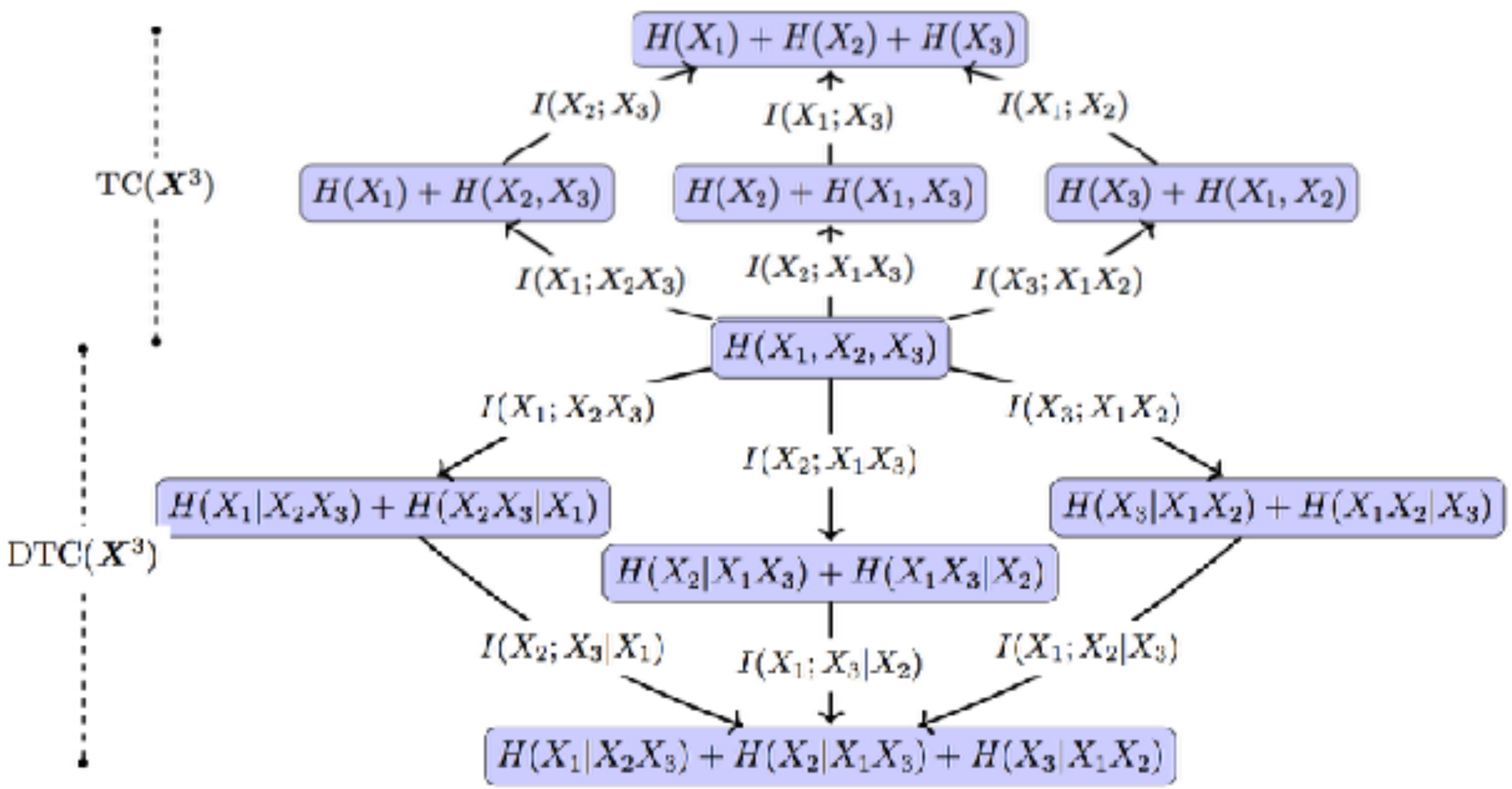
2. O-information: (1) it is sum of triple interaction-informations

- Decompositions for the TC, DTC and o-information:

$$C(\boldsymbol{X}^n) = \sum_{i=2}^{n} I(X_i; \boldsymbol{X}^{i-1}) \ ,$$

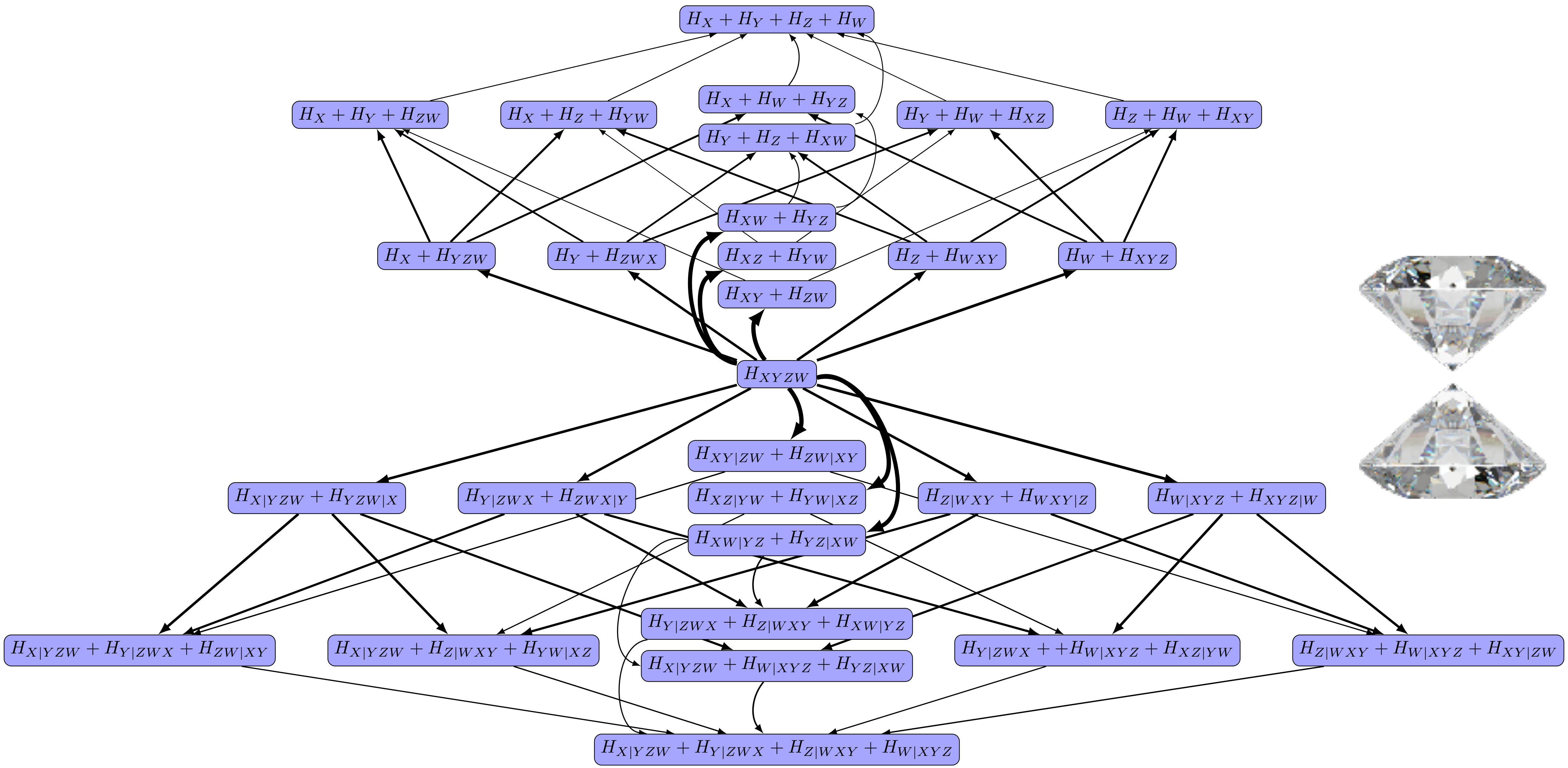
$$B(\boldsymbol{X}^n) = I(X_n; \boldsymbol{X}^{n-1}) + \sum_{j=2}^{n-1} I(X_j; \boldsymbol{X}^{j-1} | \boldsymbol{X}_{j+1}^n),$$

$$\Omega(\boldsymbol{X}^n) = \sum_{k=2}^{n-1} I(X_k; \boldsymbol{X}^{k-1}; \boldsymbol{X}_{k+1}^n) \ .$$



2. O-information: (1) it is sum of triple interaction-informations

- Decompositions for the TC, DTC and o-information:



2. O-information: (2) characterization of extreme values

- Decompositions for the TC, DTC and o-information:

$$C(\boldsymbol{X}^n) = \sum_{i=2}^{n} I(X_i; \boldsymbol{X}^{i-1}) \ ,$$

$$B(\boldsymbol{X}^n) = I(X_n; \boldsymbol{X}^{n-1}) + \sum_{j=2}^{n-1} I(X_j; \boldsymbol{X}^{j-1} | \boldsymbol{X}_{j+1}^n),$$

$$\Omega(\boldsymbol{X}^n) = \sum_{k=2}^{n-1} I(X_k; \boldsymbol{X}^{k-1}; \boldsymbol{X}_{k+1}^n) \ .$$

- Upper and lower bounds:

$$(n-1)\log|\mathcal{X}| \geq C(\boldsymbol{X}^n) \geq 0,$$

$$(n-1)\log|\mathcal{X}| \geq B(\boldsymbol{X}^n) \geq 0,$$

$$(n-2)\log|\mathcal{X}| \geq \Omega(\boldsymbol{X}^n) \geq (2-n)\log|\mathcal{X}|.$$

- Characterization of unique extremes of the o-information:

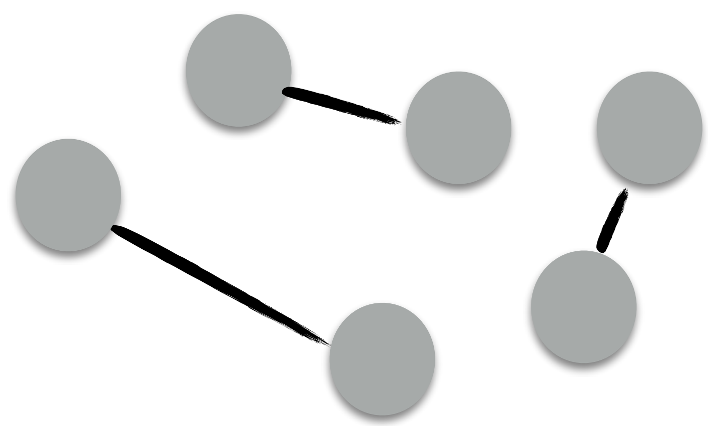
Maximum: $X_1 = X_2 = \cdots = X_n$ *Minimum:* $X_n = \sum_{j=1}^{n-1} X_j \pmod{k}$

2. O-information: (3) null value of non-overlapping pairwise interactions

- O-information of independent subgroups is additive:

$$p_{\boldsymbol{X}^n,\boldsymbol{Y}^m}(\boldsymbol{x}^n,\boldsymbol{y}^m) = p_{\boldsymbol{X}^n}(\boldsymbol{x}^n)p_{\boldsymbol{Y}^m}(\boldsymbol{y}^m) \longrightarrow \Omega(\boldsymbol{X}^n,\boldsymbol{Y}^m) = \Omega(\boldsymbol{X}^n) + \Omega(\boldsymbol{Y}^m)$$

- If a system's is decomposable in disjoint pairwise interactions then $\Omega(\boldsymbol{X}^n) = 0$

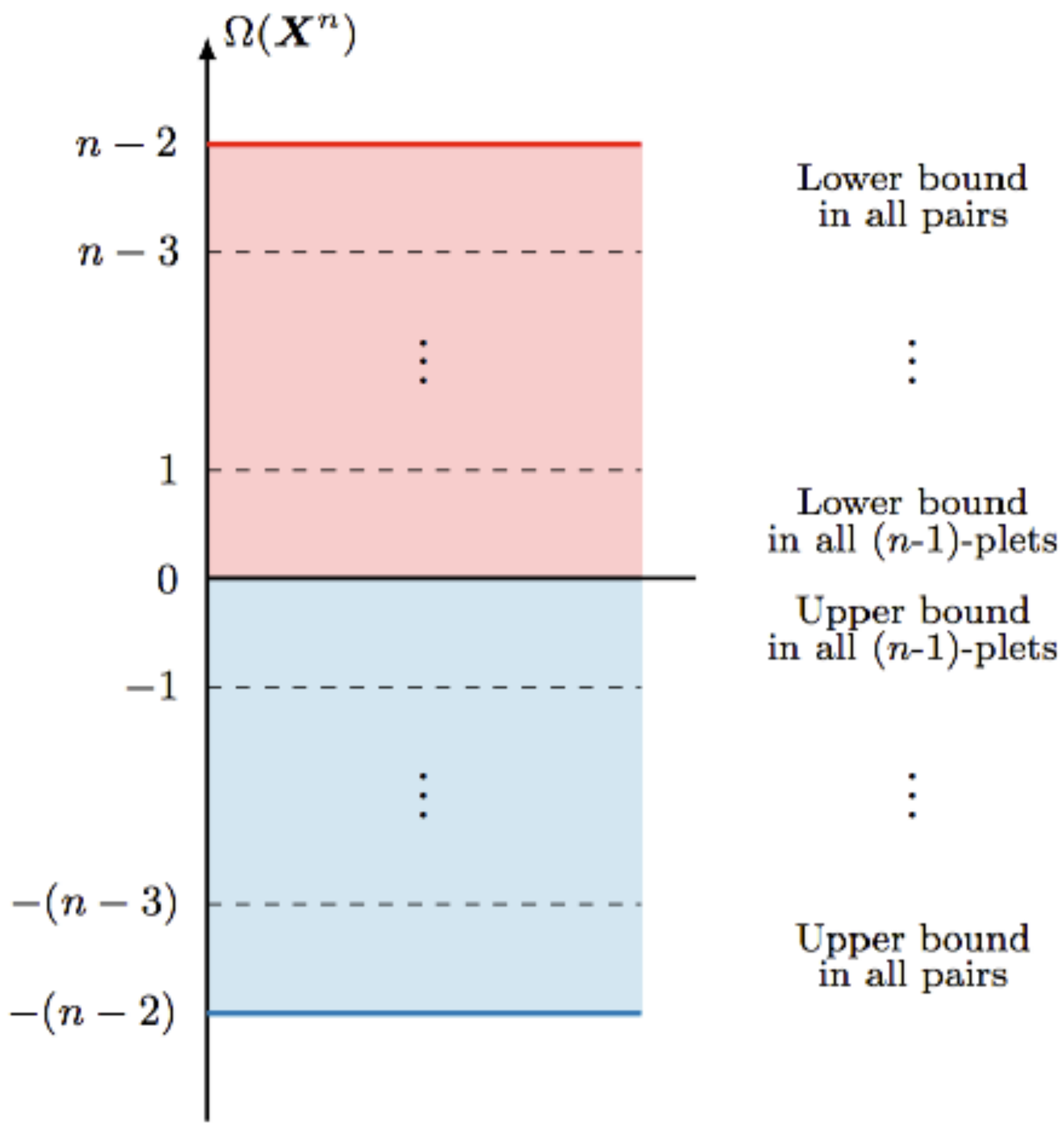


- The converse is not true!! (synergies and redundancies cancel each other)

- Local o-information $\omega_{i,j} = I(X_i;X_j;\boldsymbol{X}^n_{-i-j})$ can give a fine-grained description of the system…

2. O-information: (4) value implies bounds over different scales

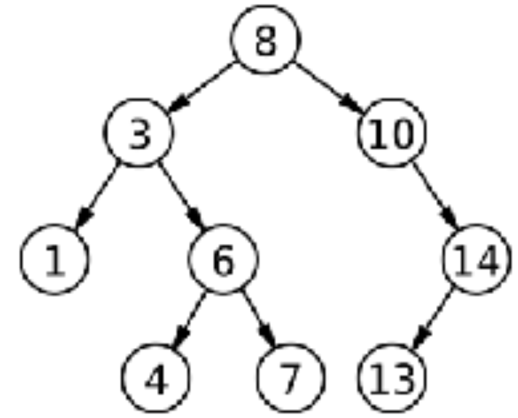
- The value of the O-information provide constraints over the interdependencies of subgroups!



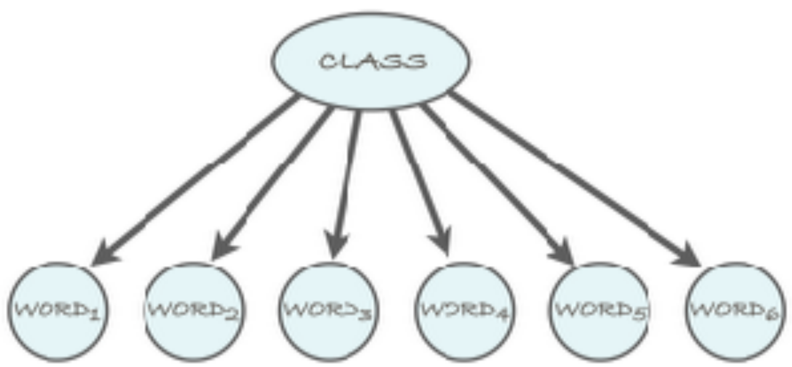
2. O-information: (5) redundancy of trees

- O-information is non-negative in graphical models with three structure!

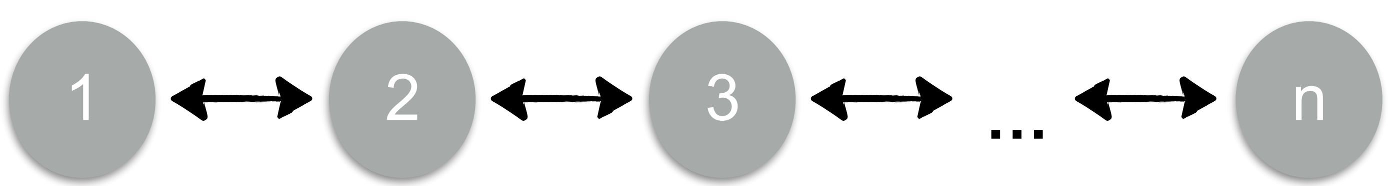
$$p_{\mathbf{X}^n}(x^n) = \prod_{i=1}^{n} p_{X_i|X_{\pi(i)}}(x_i|x_{\pi(i)})$$



- For Naive Bayes, the o-information is given by $\Omega(\boldsymbol{X}^n) = C(\boldsymbol{X}_2^n) \geq 0$

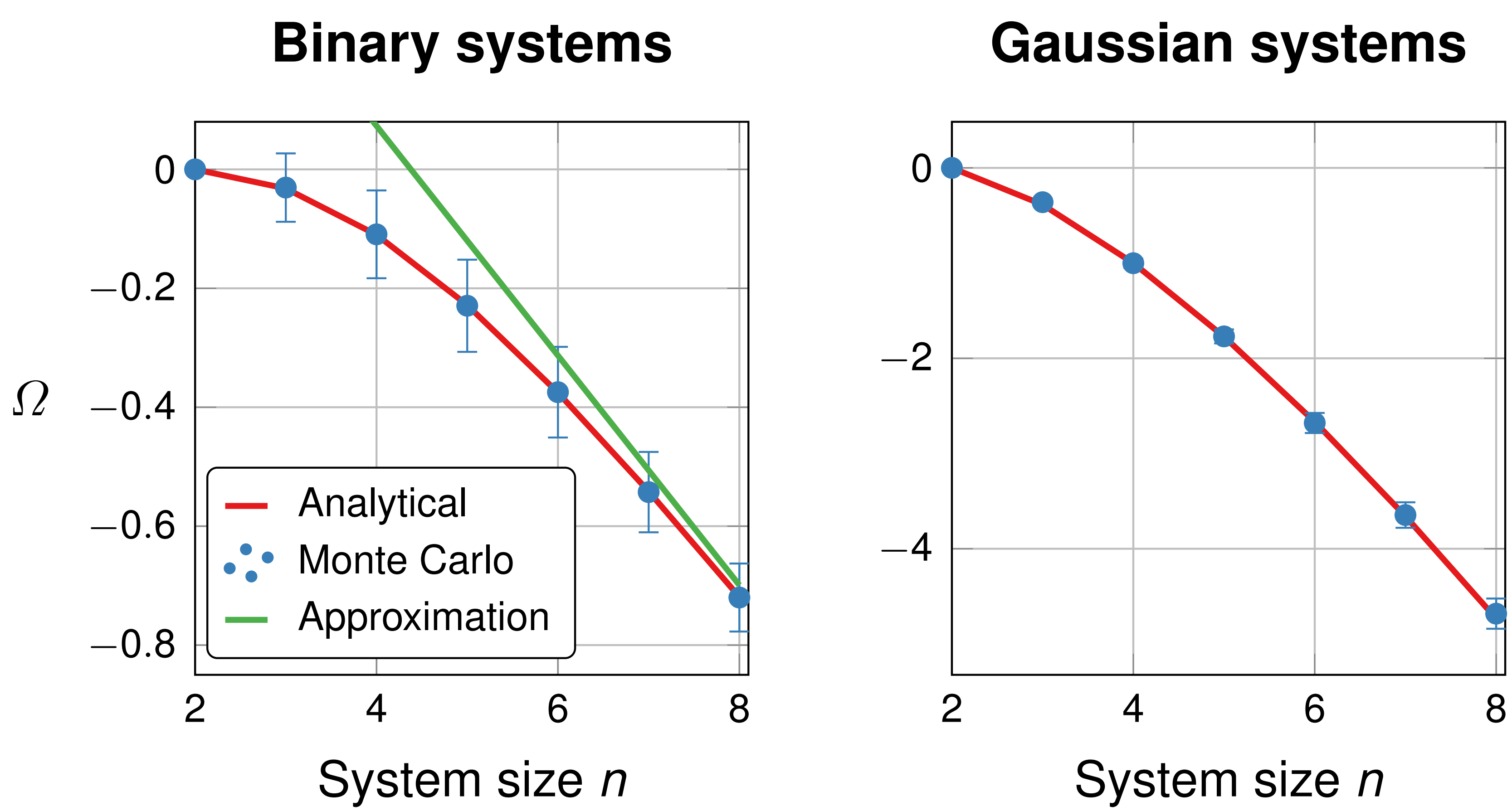


- For Markov chains, the o-information is $\Omega(\boldsymbol{X}^n) = \sum_{j=2}^{n-1} I(X_{j-1}; X_{j+1})$



2. O-information: (6) synergy is pervasive in large systems

- The mean value of the o-information over random distributions grows negative!

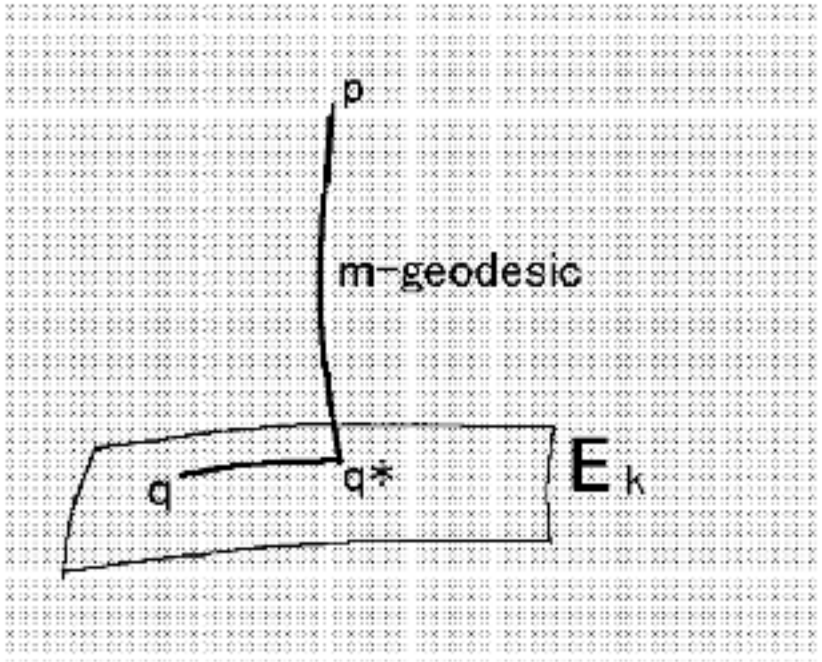


$$\overline{\Omega}(\boldsymbol{X}^n) = (n-2)\left[\psi(2^n\alpha+1) - \psi(\alpha+1)\right] + n\left[\psi(2\alpha+1) - \alpha(2^{n-1}\alpha+1)\right]$$

2. O-information: relationship with statistical mechanics

1. **Connected information** (*Schneidman et al. 2003):*

$$\mathrm{Con}(\boldsymbol{X}^n) = \frac{\sum_{k=1}^{n} H(X_j) - H(\boldsymbol{X}^n_{\mathrm{maxent}})}{\mathrm{TC}(\boldsymbol{X}^n)}$$



State of affairs:

(+) Interesting connection with statistical mechanics.

(+) Intuitive interpretation.

(-) Very hard to compute…

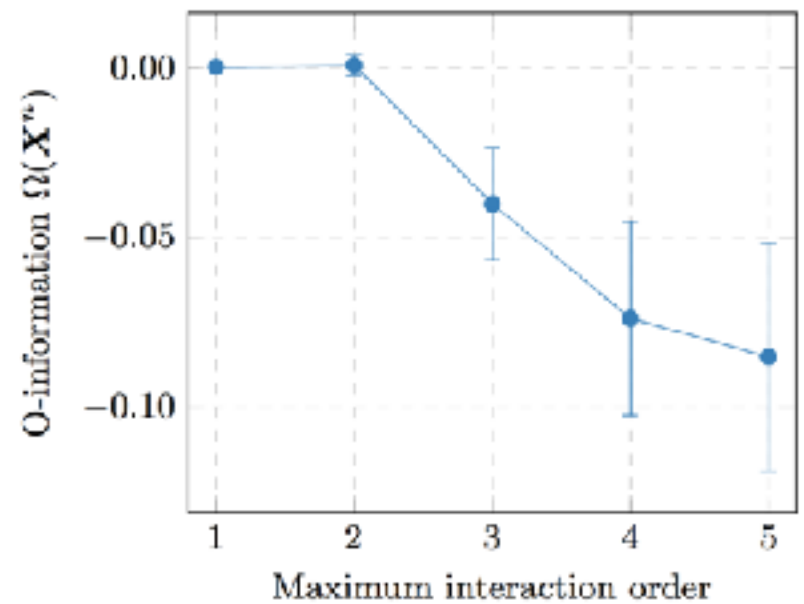
2. O-information: relationship with statistical mechanics

1. **Connected information** (*Schneidman et al. 2003*):

$$\mathrm{Con}(\boldsymbol{X}^n) = \frac{\sum_{k=1}^{n} H(X_j) - H(\boldsymbol{X}^n_{\mathrm{maxent}})}{\mathrm{TC}(\boldsymbol{X}^n)}$$

- Hamiltonian with high-order terms tend to have negative O-information

$$\mathcal{H}_k(\boldsymbol{x}^n) = -\sum_{i=1}^{n} J_i x_i - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} J_{i,j} x_i x_j$$
$$\cdots - \sum_{|\gamma|=k} J_\gamma \prod_{i \in \gamma} x_i \ ,$$



2. O-information: analysis pipeline

Procedure

1. Compute $B(\boldsymbol{X}^n)$ and $C(\boldsymbol{X}^n)$ as metrics of global correlation strength.

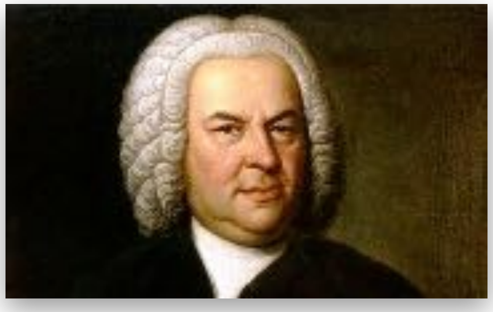
2. Compute $\Omega(\boldsymbol{X}^n)$ to find dominant global behaviour (redundancy or synergy).

3. Study the local o-information terms, i.e. $I(X_i; X_j; \boldsymbol{X}^n_{-i,-j})$ for all *i* and *j* , as a measure of localised behaviour.

2. O-information: Case study on baroque music scores

Data analysis over music scores from the Baroque period (*Python, Music21 package*)

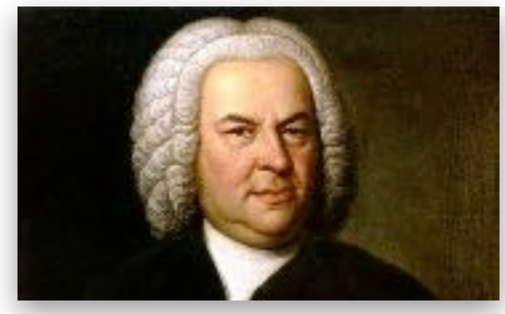
i) chorales for four voices by **J.S. Bach** (1685–1750)



2. O-information: Case study on baroque music scores

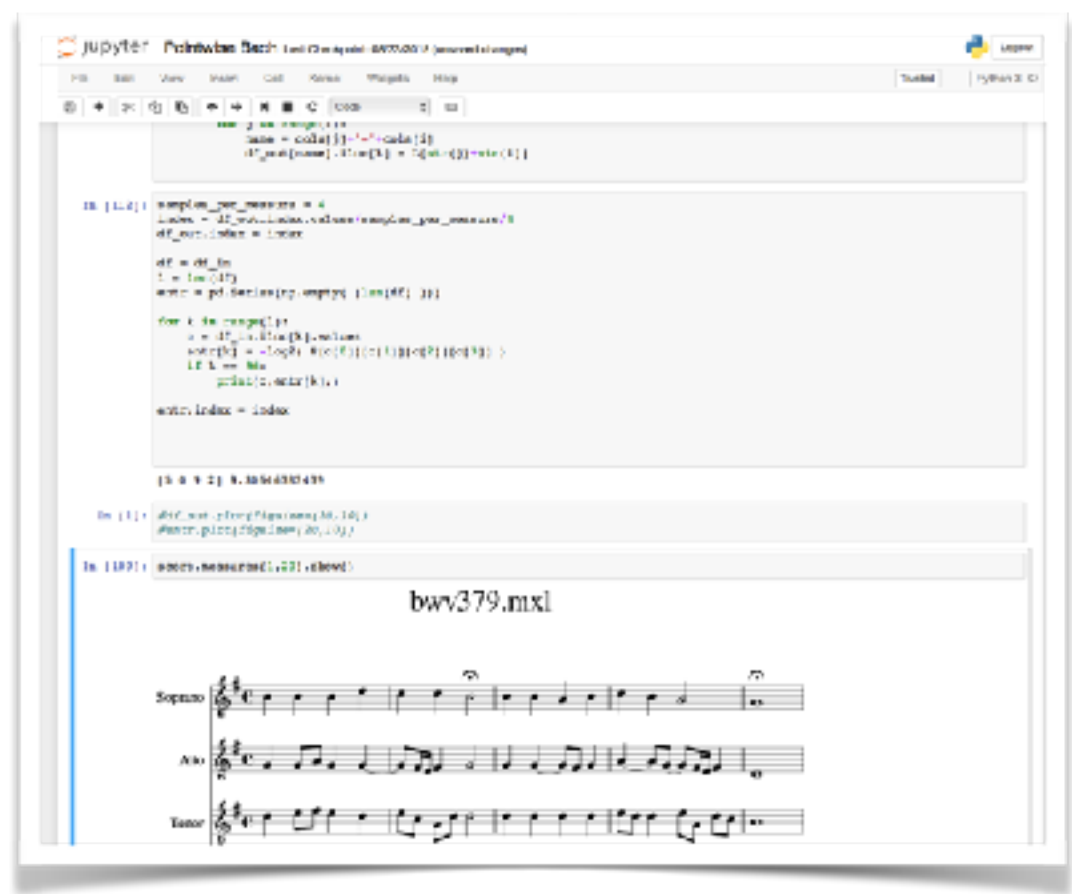
Data analysis over music scores from the Baroque period (*Python, Music21 package*)

i) chorales for four voices by **J.S. Bach** (1685–1750)



Data format: four synchronous time series,
alphabet of 13 values (12 tones+silence)

	Soprano	Alto	Tenor	Bass
0	7	4	0	0
1	7	4	0	0
2	7	4	0	0
3	7	4	0	0

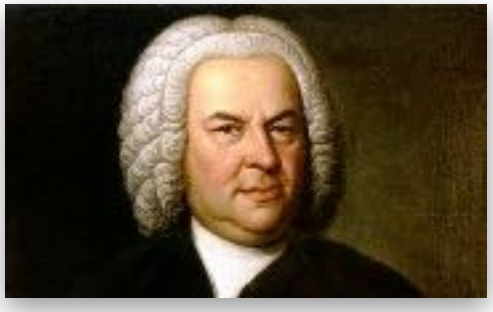


Database: 300~ chorales, 43k four-note chords

2. O-information: Case study on baroque music scores

Data analysis over music scores from the Baroque period (*Python, Music21 package*)

i) chorales for four voices by **J.S. Bach** (1685–1750)
(43k four-note chords)

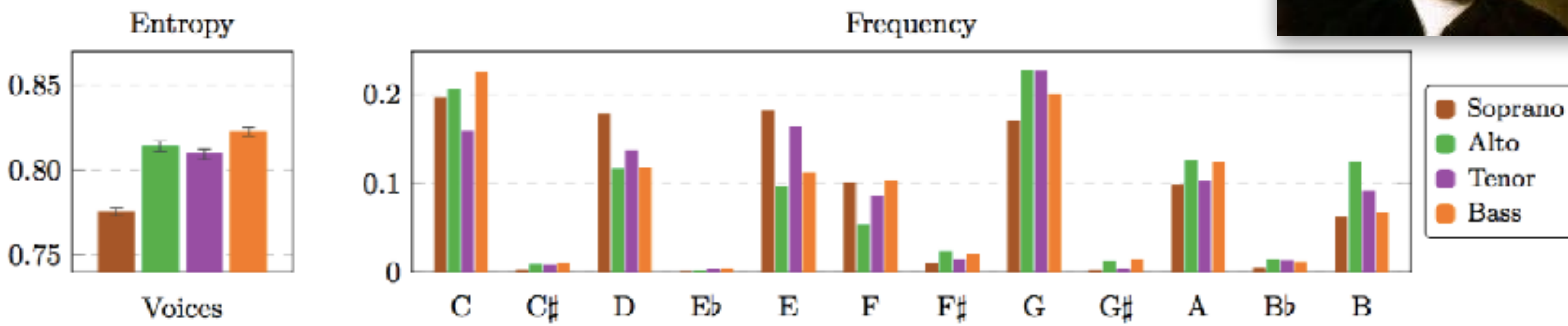


ii) Op. 1, 3, 4, 5 and 6 of **A. Corelli** (1653–1713)
(80k four-note chords)

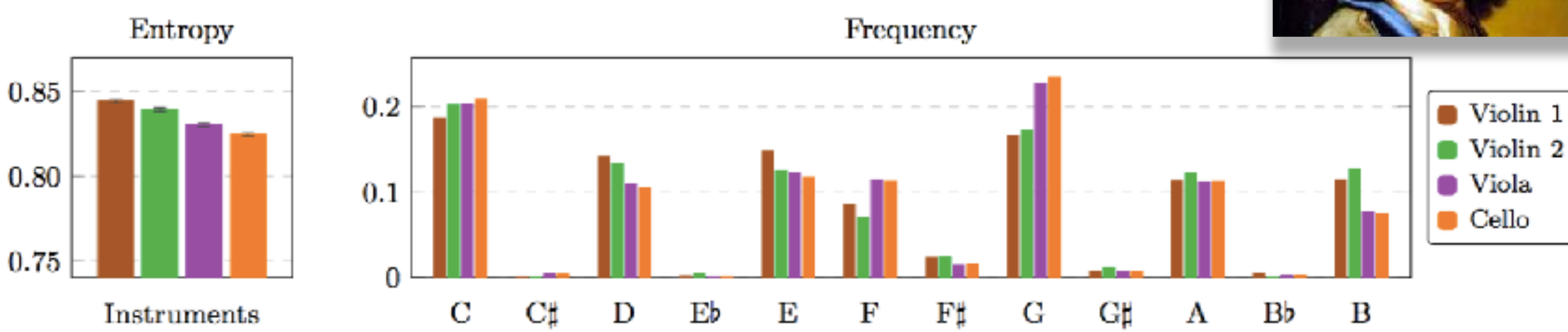
2. O-information: Case study on baroque music scores

Data analysis over music scores from the Baroque period (*Python, Music21 package*)

Bach's chorales

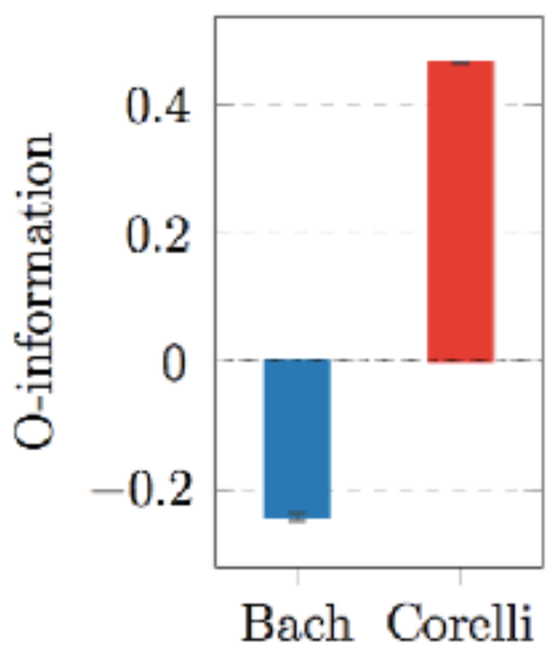


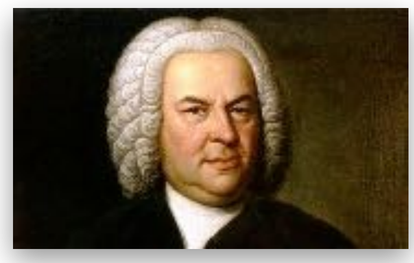
Corelli's op. 1,3-6

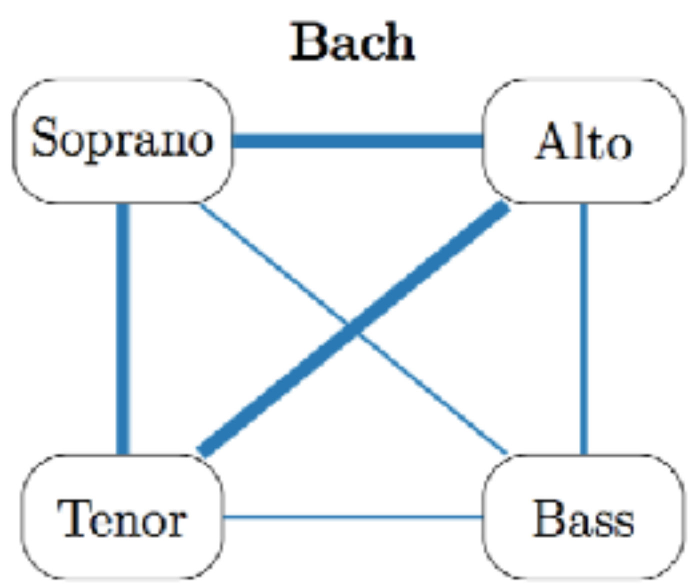


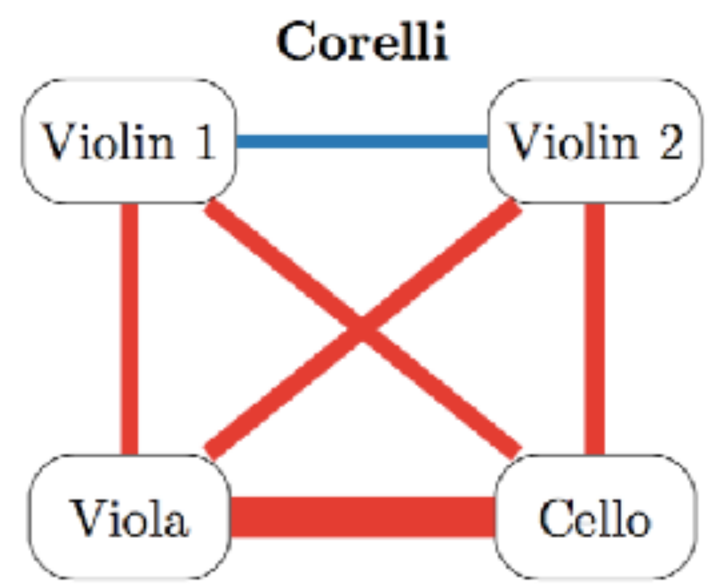
2. O-information: Case study on baroque music scores

Data analysis over music scores from the Baroque period (*Python, Music21 package*)







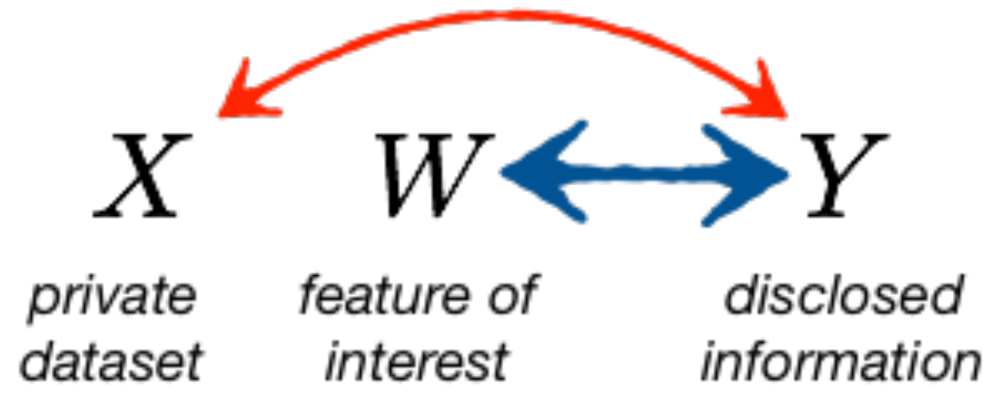


Synergy allows the coexistence of
local independency and global coordination

Today's menu
1. Self-organisation
2. O-information
3. Data privacy
4. Summary and current work

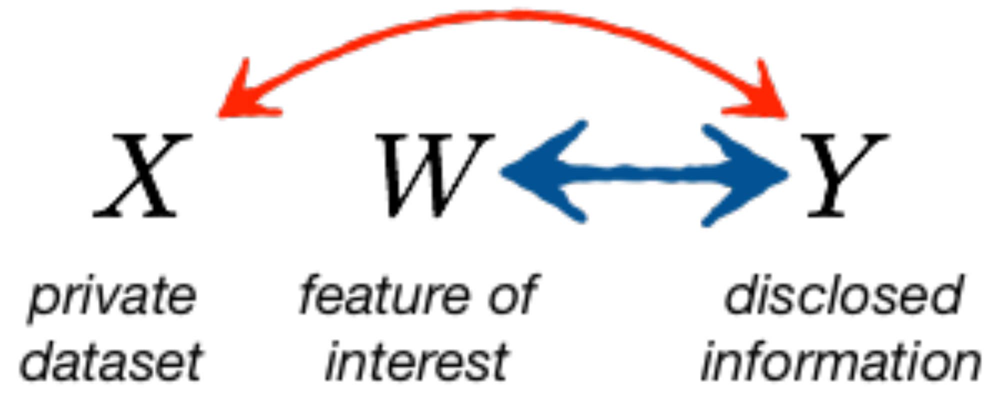
3. Synergy and data privacy: key ideas

Privacy funnel: maximize correlation with feature of interest,
while keeping private stuff secure

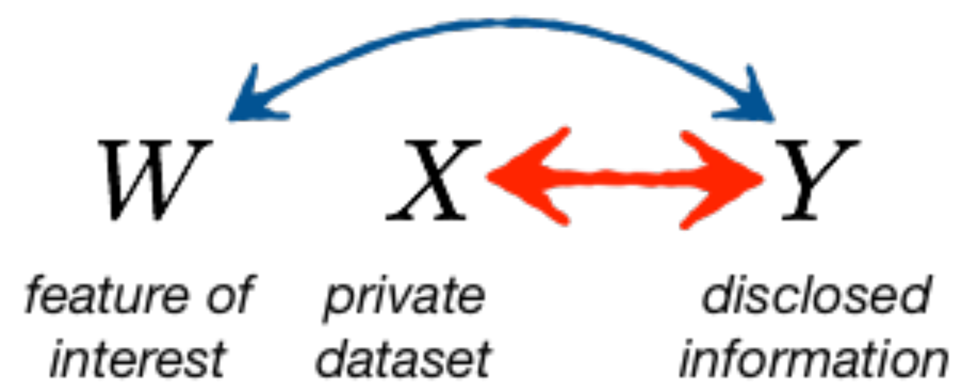


3. Synergy and data privacy: key ideas

Privacy funnel: maximize correlation with feature of interest,
while keeping private stuff secure

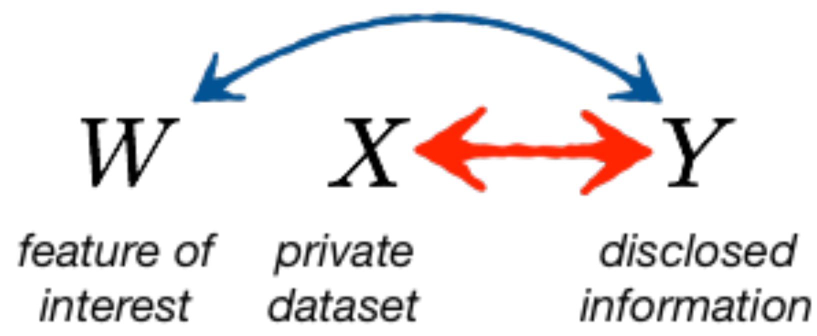


However, it is often the case where the feature of interest is unknown



But the DP inequality make it seem unfeasible…

3. Synergy and data privacy: key ideas



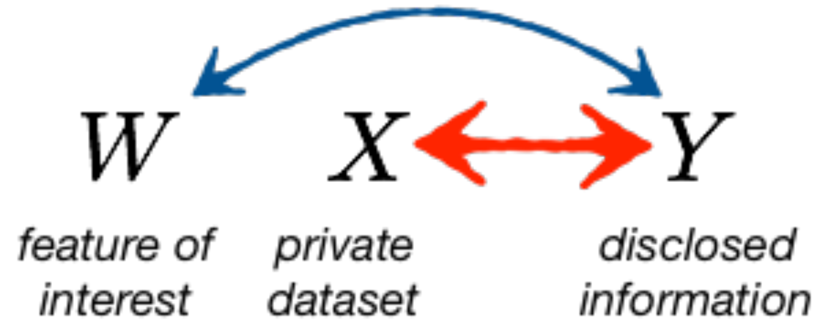
Idea: make Y independent of each coordinate of X, but correlated with the whole!!

Definition

Y generated from X satisfies perfect sample privacy if $p_{Y|X} \in \mathcal{A}_X$, where

$$\mathcal{A}_X = \left\{ p_{Y|X} \,\middle|\, Y \perp X_i, \forall i \in [1:n] \right\}. \quad (1)$$

3. Synergy and data privacy: key ideas



Idea: make Y independent of each coordinate of X, but correlated with the whole!!

Definition

Y generated from X satisfies perfect sample privacy if $p_{Y|X} \in \mathcal{A}_X$, where

$$\mathcal{A}_X = \left\{ p_{Y|X} \,\middle|\, Y \perp X_i, \forall i \in [1:n] \right\}. \tag{1}$$

Definition

The *private disclosure capacity* for a latent variable W under perfect sample privacy is then defined as

$$I_s \triangleq \max_{\substack{p_{Y|X} \in \mathcal{A}_X: \\ W - X - Y}} I(W;Y). \tag{2}$$

3. Synergy and data privacy: main results

Theorem

The optimal disclosure mapping $p^*_{Y|X}$, can be obtained as the solution to a standard LP.

Algorithm to build the optimal mapping $\mathbf{P}_{Y|X^n}$

function FINDOPTIMALMAPPING($\mathbf{P}, \mathbf{P}_{W|X^n}, \mathbf{p}_{X^n}$)
    $\mathbf{p}_1, \ldots, \mathbf{p}_K = \texttt{FindExtremePoints}(\mathbf{P}, \mathbf{p}_{X^n})$
    $c_k = H(\mathbf{P}_{W|X^n}\mathbf{p}_k)$ for $k = 1, \ldots, K$
    Find $\mathbf{u}^* = \text{Argmin} \sum_{k=1}^{K} u_k c_k$ s.t. $[\mathbf{p}_1, \ldots, \mathbf{p}_K]\mathbf{u} = \mathbf{p}_{X^n}$ and $\mathbf{u} \geq 0$
    $\mathbf{p}_Y = [p(Y=1), \ldots, p(Y=K)] = [u_1, \ldots, u_K]$
    $\mathbf{P}_{X^n|Y} = [\mathbf{p}_1, \ldots, \mathbf{p}_K]$
    $\mathbf{P}_{Y|X^n} = \text{diag}(\mathbf{p}_Y) \cdot \mathbf{P}^T_{X^n|Y} \cdot \text{diag}(\mathbf{p}_{X^n})^{-1}$
    return $\mathbf{P}_{Y|X^n}$
end function

3. Synergy and data privacy: main results

Theorem 4. *Consider a stochastic process* $\{X_i\}_{i\geq 1}$, *with* $|\mathcal{X}_i| \leq M < \infty$, $\forall i \geq 1$. *If the entropy rate of this process exists, then*

$$\lim_{n\to\infty} \frac{\hat{I}_s(X^n)}{n} = H(\mathcal{X}), \tag{52}$$

where $H(\mathcal{X})$ *denotes the entropy rate of the stochastic process* $\{X_i\}_{i\geq 1}$. *Furthermore, if* $H(\mathcal{X}) \neq 0$, *we have*

$$\lim_{n\to\infty} \hat{\eta}(X^n) = 1. \tag{53}$$

To learn more about this:

1. B. Rassouli, F. E. Rosas, and D. Gunduz. "Data Disclosure under Perfect Sample Privacy." Submitted to IEEE Transactions in Information Forensics and Security (TIFS), under review. arXiv preprint arXiv:1904.01711 (2019).
2. B. Rassouli*, F. E. Rosas*, and D. Gunduz,"Latent Feature Disclosure under Perfect Sample Privacy." In 2018 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1-7. IEEE, 2018.

Today's menu
1. Self-organisation
2. O-information
3. Data privacy
4. Summary and current work

Further reading:

1. B. Rassouli*, F. E. Rosas*, and D. Gunduz,"*Latent Feature Disclosure under Perfect Sample Privacy*." In 2018 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1-7. IEEE, 2018.

2. B. Rassouli, F. E. Rosas, and D. Gunduz. "*Data Disclosure under Perfect Sample Privacy*." Submitted to IEEE Transactions in Information Forensics and Security (TIFS), under review. arXiv preprint arXiv: 1904.01711 (2019).

3. F. Rosas, P.A. M. Mediano, M. Ugarte, H.J. Jensen, "*An information-theoretic approach to self-organization: emergence of complex interdependencies in coupled dynamical systems*", Entropy 20, no. 10 (2018): 793.

4. F. Rosas, P.A. M. Mediano, M. Gastpar, and H. J. Jensen. "*Quantifying High-order Interdependencies via Multivariate Extensions of the Mutual Information,*" accepted, to be published in PRE, 2019

PD: current work

Driving questions:

1. Is the brain a synergistic multi-agent system?
2 .Is consciousness related with high-order statistics?
3. Can PID principles help us to understand the effect of psychedelic drugs?

The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs

Robin L. Carhart-Harris[1*], Robert Leech[2], Peter J. Hellyer[2], Murray Shanahan[3], Amanda Feilding[4], Enzo Tagliazucchi[5], Dante R. Chialvo[6] and David Nutt[1]

[1] Division of Brain Sciences, Department of Medicine, Centre for Neuropsychopharmacology, Imperial College London, London, UK
[2] C3NL, Division of Brain Sciences, Department of Medicine, Imperial College London, London, UK
[3] Department of Computing, Imperial College London, London, UK
[4] The Beckley Foundation, Beckley Park, Oxford, UK
[5] Neurology Department and Brain Imaging Center, Goethe University, Frankfurt am Main, Germany
[6] Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Buenos Aires, Argentina

Entropy is a dimensionless quantity that is used for measuring uncertainty about the state of a system but it can also imply physical qualities, where high entropy is synonymous with high disorder. Entropy is applied here in the context of states of consciousness and their associated neurodynamics, with a particular focus on the psychedelic state. The psychedelic state is considered an exemplar of a primitive or primary state of consciousness that preceded the development of modern, adult, human, normal waking consciousness. Based on neuroimaging data with psilocybin, a classic psychedelic drug, it is argued

THE CENTRE FOR PSYCHEDELIC RESEARCH

Complexity Science

Data Science Institute

Acknowledges:

This work is the result of the fortunate collaboration with

- Pedro Mediano (*Imperial College London*)
- Henrik J. Jensen (*Imperial College London*)
- Michael Gastpar (*EPFL*)
- Martin Ugarte (*Université libre de Bruxelles*)
- Borzoo Rasouli (*University of Essex*)
- Deniz Gündüz (*Imperial College London*)

Special thanks for the *Marie Slodowska-Curie* program from *H20202* and the *European Commission*.

Thank you!

Contact information:

Fernando Rosas

f.rosas@imperial.ac.uk