Gromov's Ergobrain Program as a Mathematical Promenade

Matilde Marcolli and Doris Tsao

Ma191b Winter 2017
Geometry of Neuroscience

References for this lecture:

- Misha Gromov, *Structures, Learning, and Ergosystems*, http://www.ihes.fr/~gromov/PDF/ergobrain.pdf
- Misha Gromov, *Ergostructures, Ergologic and the Universal Learning Problem* http://www.ihes.fr/~gromov/PDF/ergologic3.1.pdf

Gromov's Ergobrain

• Gromov conjectures the existence of a *mathematical* structure implementing the transformation of incoming signal to representation in the brain: ergobrain

• a dynamic entity continuously built by the brain: (goal free) structure learning

• with constraints from network architecture

• Gromov's proposed terminology:

- neuro-brain: a (mathematical) model of the physiology of the brain (chemical, electrical, connectivity, etc.)
- ergo-brain: a "dynamical reduction" of neuro-brain (quotient)
- ergo-system: like ergobrain but not necessarily derived from a neurobrain
- ego-mind: mental processes, interactions of organisms with outside world
- ergo-learning: spontaneous structure building process

The role of Mathematics

• long history of very successful interactions of mathematics and physics

- "The Unreasonable Effectiveness of Mathematics in the Natural Sciences" (Eugene Wigner, 1960): referring to physical science and how mathematics can drive new advances in physics
- in more recent years we have also seen the "unreasonable effectiveness of physics in the mathematical sciences" with new progress in mathematics driven by physics

• Biology has traditionally not put too much emphasis on theoretical ideas as a guiding principle for progress in the field; interactions between mathematics and biology are recent and still largely underdeveloped

• Gromov's speculation on mathematics and neuroscience: "basic mental processes can be meaningfully described, if at all, only in a broader mathematical context and this mathematics does not exist at the present day"

Example of mathematical viewpoint: symmetry

• several molecules occur with symmetries (icosahedral symmetry of viruses; helix and double helix symmetries in proteins and DNA, etc.)...

• *energy and symmetry*: configuration space $\mathcal{M}$ of molecules with group action G, invariant energy functional E, typically local minima over G-symmetric configurations are local minima for all configuration

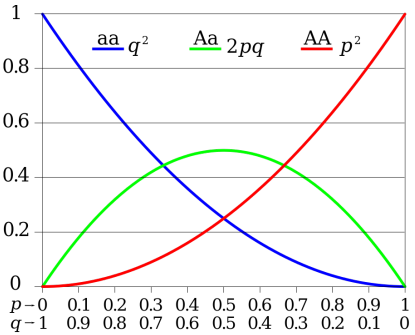
• *information and symmetry*: a symmetric form is specified by fewer parameters, preferable from the Shannon information viewpoint

Other Example: Hardy-Weinberg principle

- allele and genotype frequencies in a population would remain constant through generations in the absence of other evolutionary influences (mutation, selection, etc.)
- two alleles with frequencies $f_0(A) = p$ and $f_0(a) = q$

$f_1(AA) = p^2$, $f_1(aa) = q^2$ homozygotes, $f_1(Aa) = 2pq$ heterozygotes

with $p^2 + q^2 + 2pq = 1$, ok for $q = 1 - p$ (similar for n alleles a_i)



Hardy-Weinberg equilibrium

• linear algebra identity: $M = (m_{ij})$ matrix, $m_{ij} \geq 0$, and $\sum_{ij} m_{ij} = 1$

$$M' = (m'_{ij}) \quad \text{with} \quad m'_{ij} = (\sum_j m_{ij}) \cdot (\sum_i m_{ij})$$

$$\hat{M} = \frac{M'}{\sum_{ij} m'_{ij}}$$

$$\text{then} \quad \hat{\hat{M}} = \hat{M}$$

distribution of phenotype features depending on a single gene changes in the first generation but remains constant in the successive generations

• gene recombination not sufficient to explain evolution, in the absence of additional phenomena like gene mutation

Other Example: Complexity and Patterns

- Kolmogorov Complexity

shortest length of a program required to compute the given output (theory of computation, Turing machines)

- Gell-Mann Effective Complexity

description length of "regularities" (structured patterns) contained in the object

More generally What kind of mathematics?

General framework:

1. Combinatorial objects: from *graphs* to *n-categories* (networks and relations)
2. Transformations and symmetries: *groups* and generalizations (groupoids, Hopf algebras, operads, etc.)
3. Probabilities and information/entropy: algebraic structures with superimposed probabilistic and thermodynamical (statistical physics) structures

- a more detailed list of the mathematical tookbox

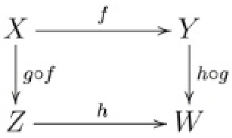
Categories

• traditional setting of mathematics based on Set Theory, first example of categories: Sets (or Finite Sets)

• Category $\mathcal{C}$: Objects $X, Y, \dots \in \mathcal{O}(\mathcal{C})$, Morphisms $\mathrm{Hom}_{\mathcal{C}}(X, Y)$,

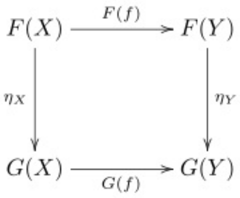
$$h \circ (g \circ f) = (h \circ g) \circ f$$

associative composiiton $X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} W$ and identity $1_X \in \mathrm{Hom}_{\mathcal{C}}(X, X)$ unit for composition



• Functors $F : \mathcal{C} \to \mathcal{C}'$, on objects $\mathcal{O}(\mathcal{C}) \ni X \mapsto F(X) \in \mathcal{O}(\mathcal{C}')$, on morphisms $F(f) : F(X) \to F(Y)$ (covariant; also contravariant) with $F(g \circ f) = F(g) \circ F(f)$ and $F(1_X) = 1_Y$

• Natural Transformations: $\eta : F \to G$, to every object a morphism $\eta_X : F(X) \to G(X)$ with $\eta_Y \circ F(f) = G(f) \circ \eta_X$



• Categories of Vector Spaces, Topological Spaces, Smooth Manifolds, Groups, Rings, etc.

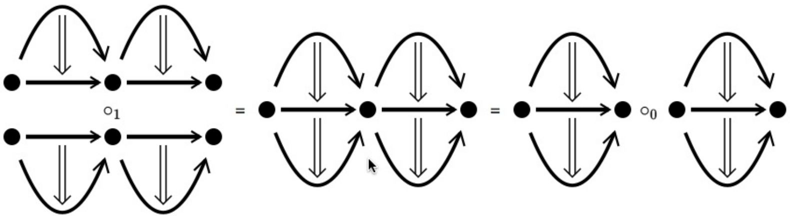
... concept of mathematical structure

Higher Categorical Structures

- 2-Categories: a category $\mathcal{C}$ where the sets $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ are themselves the objects of a category (ie there are morphisms between morphisms: 2-morphisms)

- vertical and horizontal composition of 2-morphisms with exchange relation

$$(\alpha \circ_0 \beta) \circ_1 (\gamma \circ_0 \delta) = (\alpha \circ_1 \gamma) \circ_0 (\beta \circ_1 \delta)$$



- objects = vertices; morphisms = 1-cells; 2-morphisms = 2-cells

Homotopy and Higher Category Theory

• Note: 2-category composition of 1-morphisms associative; bicategory associative up to 2-isomorphism

• Example: Objects = points in an open set in the plane; Morphisms = oriented paths connecting points; 2-morphisms = homotopies of paths; composition up to homotopy... can also compose homotopies up to homotopy etc. ... higher *n*-categories (with *n* different levels of "isomorphism": 0-isomorphism equality, 1-isomorphism realized by invertible 1-morphisms; 2-isomorphism by 2-morphisms etc.)... ∞-categories

• Contemporary mathematical viewpoint is shifting more and more towards these higher (or ∞) structures and homotopy

Groups, Semigroups, Groupoids, Algebras, and Categories

• Group: small category with one object and all morphisms invertible (product, associativity, unit)
... symmetries (action by automorphisms)

• Group algebra $\mathbb{C}[G]$ (discrete group G) finitely supported functions $f : G \to \mathbb{C}$ with convolution

$$(f_1 \star f_2)(g) = \sum_{g=g_1 g_2} f_1(g_1) f_2(g_2)$$

involution $f^*(g) \equiv \overline{f(g^{-1})}$

• Semigroup: small category with one object (not always inverses)
... actions by endomorphisms

• Semigroup algebra: $f : S \to \mathbb{C}$ with convolution

$$(f_1 \star f_2)(s) = \sum_{s=s_1 s_2} f_1(s_1) f_2(s_2)$$

no longer necessarily involutive

• Groupoid: small category where all morphisms are invertible (product is defined only when target of first arrow is source of second)

... another type of symmetry

• Groupoid algebra $\mathcal{G} = (\mathcal{G}^{(0)}, \mathcal{G}^{(1)}, s, t)$ (objects and morphisms, source and target); algebra of functions $f : \mathcal{G}^{(1)} \to \mathbb{C}$ with convolution

$$(f_1 \star f_2)(\gamma) = \sum_{\gamma=\gamma_1\circ\gamma_2} f_1(\gamma_1) f_2(\gamma_2)$$

and involution $f^*(\gamma) = \overline{f(\gamma^{-1})}$

• Semigroupoid: a small category (associative composition of morphisms)

• Semigroupoid algebra: functions of morphisms with convolution

$$(f_1 \star f_2)(\phi) = \sum_{\phi=\phi_1\circ\phi_2} f_1(\phi_1) f_2(\phi_2)$$

Topology and invariants

• Topological spaces and continuous functions (the study of shapes up to continuous deformations)

• Invariants: ways of distinguishing between topological spaces

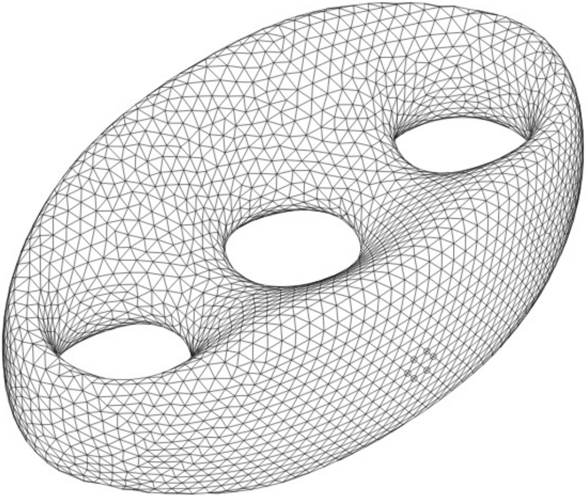
• Invariants are functors from the category $\mathcal{T}$ of Topological Spaces to another category (vector spaces, groups, rings, etc.)

• Homology: functor $H_* : \mathcal{T} \to \mathcal{V}_{\mathbb{Z}}$ to $\mathbb{Z}$-graded vector spaces

$$H_n(X, \mathbb{Z}) = \frac{\mathrm{Ker}(\partial_n : C_n(X, \mathbb{Z}) \to C_{n-1}(X, \mathbb{Z}))}{\mathrm{Image}(\partial_{n+1} : C_{n+1}(X, \mathbb{Z}) \to C_n(X, \mathbb{Z}))}$$

$C_n(X, \mathbb{Z})$ abelian group spanned by n-simplexes of a triangulation of X; ∂_n (oriented) boundary map

Homology is independent of the choice of triangulation; it measures "holes" and "connectivity" in various dimensions



a triangulation of a surface of genus 3

Metric Spaces, Riemannian Manifolds

• Topological spaces with a quantitative measure of proximity: topology is induced by a metric (open sets generated by open balls in the metric)

• distance function (metric) $d : X \times X \to \mathbb{R}_+$ with

- $d(x,y) = d(y,x)$ for all $x, y \in X$
- $d(x,y) = 0$ iff $x = y$
- triangle inequality $d(x,y) \leq d(x,z) + d(z,y)$ for all $x, y, z \in X$

• smooth manifold M (covered by local charts homeomorphic to $\mathbb{R}^n$ with $\mathcal{C}^\infty$ changes of coordinates in charts overlap); tangent spaces T_xM at all points (tangent bundle TM)

• Riemannian manifold: smooth manifold, metric structure determined by a metric tensor $g = (g_{\mu\nu})$ symmetric positive, section of $T^*M \otimes T^*M$

• length of curves $L(\gamma) = \int g(\gamma'(t), \gamma'(t))^{1/2}\, dt$, geodesic distance:

$$d(x,y) = \inf_\gamma L(\gamma)$$

Ambiguity, Galois Symmetry, Category Theory

• Symmetries describe ambiguities (up to isomorphism, up to invertible transformations, up to homotopy, etc.)

• Categorical viewpoint on Symmetry: need categories with good properties... very much like the category of vectors spaces: abelian (kernels and cokernels), tensor, rigid (duality, internal Hom)

• fiber functor $\omega : \mathcal{C} \to \mathcal{V}$ to vector spaces preserving all properties (tensor, etc.) and symmetries

$$G = \mathrm{Aut}(\omega)$$

the invertible natural transformations of the fiber functor

• $(\mathcal{C}, \omega)$ as above: Tannakian category with Galois group G

• this includes: usual Galois theory; Motives; Regular-singular differential systems; symmetries of Quantum Field Theory, etc.

Formal Languages and Grammars

- A very general abstract setting to describe languages (natural or artificial: human languages, codes, programming languages, ...)
- Alphabet: a (finite) set $\mathfrak{A}$; elements are *letters* or *symbols*
- Words (or strings): $\mathfrak{A}^m$ = set of all sequences $a_1 \dots a_m$ of length m of letters in $\mathfrak{A}$
- Empty word: $\mathfrak{A}^0 = \{\epsilon\}$ (an additional symbol)

$$\mathfrak{A}^+ = \cup_{m\geq 1}\mathfrak{A}^m, \qquad \mathfrak{A}^\star = \cup_{m\geq 0}\mathfrak{A}^m$$

- concatenation: $\alpha = a_1 \dots a_m \in \mathfrak{A}^m$, $\beta = b_1 \dots b_k \in \mathfrak{A}^k$

$$\alpha\beta = a_1 \dots a_m b_1 \dots b_k \in \mathfrak{A}^{m+k}$$

associative $(\alpha\beta)\gamma = \alpha(\beta\gamma)$ with $\epsilon\alpha = \alpha\epsilon = \alpha$

- semigroup $\mathfrak{A}^+$; monoid $\mathfrak{A}^\star$
- Length $\ell(\alpha) = m$ for $\alpha \in \mathfrak{A}^m$

- subword: $\gamma \subset \alpha$ if $\alpha = \beta\gamma\delta$ for some other words $\beta, \delta \in \mathfrak{A}^\star$:
- prefix β and suffix δ
- Language: a subset of $\mathfrak{A}^\star$
- Question: how is the subset constructed?
- Rewriting system on $\mathfrak{A}$: a subset $\mathcal{R}$ of $\mathfrak{A}^\star \times \mathfrak{A}^\star$
- $(\alpha, \beta) \in \mathcal{R}$ means that for any $u, v \in \mathfrak{A}^\star$ the word $u\alpha v$ *rewrites* to $u\beta v$
- Notation: write $\alpha \to_{\mathcal{R}} \beta$ for $(\alpha, \beta) \in \mathcal{R}$
- $\mathcal{R}$-derivation: for $u, v \in \mathfrak{A}^\star$ write $u \stackrel{\bullet}{\to}_{\mathcal{R}} v$ if $\exists$ sequence $u = u_1, \ldots, u_n = v$ of elements in $\mathfrak{A}^\star$ such that $u_i \to_{\mathcal{R}} u_{i+1}$

Grammar: a quadruple $\mathcal{G} = (V_N, V_T, P, S)$

- V_N and V_T disjoint finite sets: *non-terminal* and *terminal* symbols
- $S \in V_N$ *start symbol*
- P finite rewriting system on $V_N \cup V_T$

$P =$ *production rules*

Language produced by a grammar $\mathcal{G}$:

$$\mathcal{L}_{\mathcal{G}} = \{ w \in V_T^\star \,|\, S \stackrel{\bullet}{\to}_P w \}$$

language with alphabet V_T

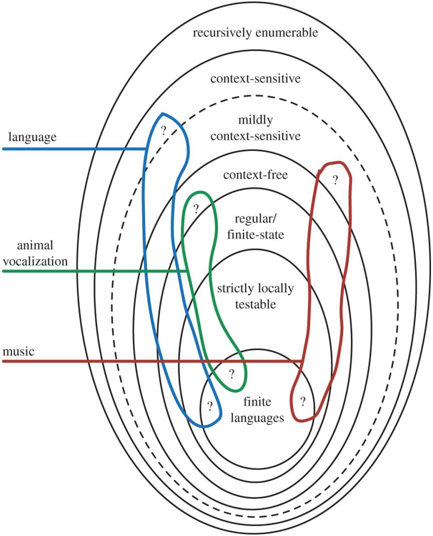
The Chomsky hierarchy of formal languages

Types:

- Type 0:*unrestricted grammars*
- Type 1: *context-sensitive grammars*
- Type 2: *context-free grammars*
- Type 3: *regular grammars*

Language of type n if produced by a grammar of type n

(formal languages will be discussed in more detail later in the class)



from: M.Rohrmeier, W.Zuidema, G.A.Wiggins, C.Scharff, "Principles of structure building in music, language and animal song" Phil. Trans. Royal Soc. B, 370 (2015) N.1664

Turing machine $T = (Q, F, \mathfrak{A}, I, \tau, q_0)$

- Q finite set of possible states
- F subset of Q: the final states
- $\mathfrak{A}$ finite set: alphabet (with a distinguished element B *blank symbol*)
- $I \subset \mathfrak{A} \smallsetminus \{B\}$ input alphabet
- $\tau \subset Q \times \mathfrak{A} \times Q \times \mathfrak{A} \times \{L, R\}$ transitions with $\{L, R\}$ a 2-element set
- $q_0 \in Q$ initial state

$qaq'a'L \in \tau$ means T is in state q, reads a on next square in the tape, changes to state q', overwrites the square with new letter a' and moves one square to the left

- *tape description* for T: triple (a, α, β) with $a \in \mathfrak{A}$, $\alpha : \mathbb{N} \to \mathfrak{A}$, $\beta : \mathbb{N} \to \mathfrak{A}$ such that $\alpha(n) = B$ and $\beta(n) = B$ for all but finitely many $n \in \mathbb{N}$ (sequences of letters on tape right and left of a)

- *configuration* of T: (q, a, α, β) with $q \in Q$ and (a, α, β) a tape description

- configuration c' from c in a single move if either
  - $c = (q, a, \alpha, \beta)$, $qaq'a'L \in \tau$ and $c' = (q', \beta(0), \alpha', \beta')$ with $\alpha'(0) = a'$ and $\alpha'(n) = \alpha(n-1)$, and $\beta'(n) = \beta(n+1)$
  - $c = (q, a, \alpha, \beta)$, $qaq'a'R \in \tau$ and $c' = (q', \alpha(0), \alpha', \beta')$ with $\alpha'(n) = \alpha(n+1)$, and $\beta'(0) = a'$, $\beta'(n) = \beta(n-1)$

- *computation* $c \to c'$ in T starting at c and ending at c': finite sequence $c = c_1, \ldots, c_n = c'$ with c_{i+1} from c_i by a single move

- computation *halts* if c' *terminal configuration*, $c' = (q, a, \alpha, \beta)$ with no element in τ starting with qa

• word $w = a_1 \cdots a_n \in \mathfrak{A}^\star$ *accepted* by T if for $c_w = (q_0, a_1 \cdots a_n)$ there is a computation in T of the form $c_w \to c' = (q, a, \alpha, \beta)$ with $q \in F$

• Language recognized by T

$$\mathcal{L}_T = \{ w \in \mathfrak{A}^\star \,|\, w \text{ is accepted by } T \}$$

• Turing machine T *deterministic* if for given $(q, a) \in Q \times \mathfrak{A}$ there is at most one element of τ starting with qa

Automata and Formal Languages

- Types and Machine Recognition:

The different types of formal languages in the Chomsky hierarchy are recognized by:

- Type 0: Turing machine
- Type 1: linear bounded automaton
- Type 2: non-deterministic pushdown stack automaton
- Type 3: finite state automaton

(automata and formal languages will be discussed in more detail later in the class)

- A Key Idea: languages are a type of mathematical structure

Probabilities and Entropy

• mathematical structures (especially algebraic structures) endowed with probabilities

• a successful approach in formal languages and generative grammars: probabilistic grammars (generative rules applied with certain probabilities)

• other algebraic structures can be made probabilistic: groups, semigroups, groupoids, semigroupoids... all like directed graphs: assign probabilities at each successive choice of next oriented edge in a path... like Markov processes

• also attach information measures (entropy) to algebraic structures (operations weighted by entropy functionals): information algebras

The Key Idea: in applications to Biology all mathematical structures should be endowed with probabilistic weights and entropy/information weights

Kolmogorov complexity

• Let $T_{\mathcal{U}}$ be a universal Turing machine (a Turing machine that can simulate any other arbitrary Turing machine: reads on tape both the input and the description of the Turing machine it should simulate)

• Given a string w in an alphabet $\mathfrak{A}$, the Kolmogorov complexity

$$\mathcal{K}_{T_{\mathcal{U}}}(w) = \min_{P:\, T_{\mathcal{U}}(P)=w} \ell(P),$$

minimal length of a program that outputs w

• universality: given any other Turing machine T

$$\mathcal{K}_T(w) = \mathcal{K}_{T_{\mathcal{U}}}(w) + c_T$$

shift by a bounded constant, independent of w; c_T is the Kolmogorov complexity of the program needed to describe T for $T_{\mathcal{U}}$ to simulate it

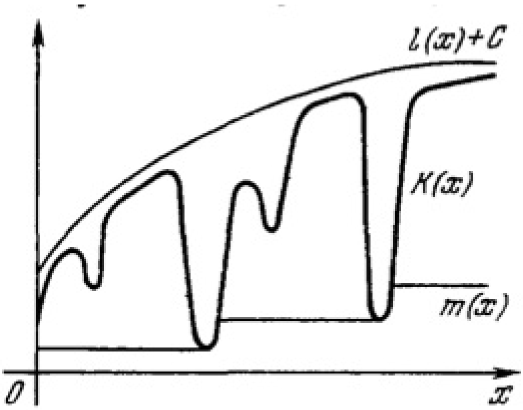
• any program that produces a description of w is an upper bound on Kolmogorov complexity $\mathcal{K}_{T_{\mathcal{U}}}(w)$

• think of Kolmogorov complexity in terms of data compression

• shortest description of w is also its most compressed form

• can obtain upper bounds on Kolmogorov complexity using data compression algorithms

• finding upper bounds is easy... but NOT lower bounds



with $m(x) = \min_{y \geq x} \mathcal{K}(y)$

Problem

Kolmogorov complexity is NOT a computable function

• suppose list programs P_k (increasing lengths) and run through $T_{\mathcal{U}}$: if machine halts on P_k with output w then $\ell(P_k)$ is an upper bound on $\mathcal{K}_{T_{\mathcal{U}}}(w)$

• but... there can be an earlier P_j in the list such that $T_{\mathcal{U}}$ has not yet halted on P_j

• if eventually halts and outputs w then $\ell(P_j)$ is a better approximation to $\mathcal{K}_{T_{\mathcal{U}}}(w)$

• would be able to compute $\mathcal{K}_{T_{\mathcal{U}}}(w)$ if can tell exactly on which programs P_k the machine $T_{\mathcal{U}}$ halts

• but... halting problem is unsolvable

Kolmogorov Complexity and Entropy

- Independent random variables X_k distributed according to Bernoulli measure $\mathbb{P} = \{p_a\}_{a\in\mathfrak{A}}$ with $p_a = \mathbb{P}(X = a)$
- Shannon entropy $S(X) = -\sum_{a\in\mathfrak{A}} \mathbb{P}(X = a) \log \mathbb{P}(X = a)$
- $\exists c > 0$ such that for all $n \in \mathbb{N}$

$$S(X) \leq \frac{1}{n} \sum_{w\in\mathcal{W}^n} \mathbb{P}(w)\, \mathcal{K}(w \,|\, \ell(w) = n) \leq S(X) + \frac{\#\mathfrak{A} \log n}{n} + \frac{c}{n}$$

- expectation value

$$\lim_{n\to\infty} \mathbb{E}(\frac{1}{n}\mathcal{K}(X_1 \cdots X_n \,|\, n)) = S(X)$$

average expected Kolmogorov complexity for length n descriptions approaches Shannon entropy

Gell-Mann Effective Complexity

• unlike Kolmogorov complexity does not measure description length of whole object

• based on description length of "regularities" (structured patterns) contained in the object

• a completely random sequence has maximal Kolmogorov complexity but zero effective complexity (it contains no structured patterns)

• based on measuring Kolmogorov complexity of subsequences

• criticized because it requires a criterion for separating subsequences into regularities and random