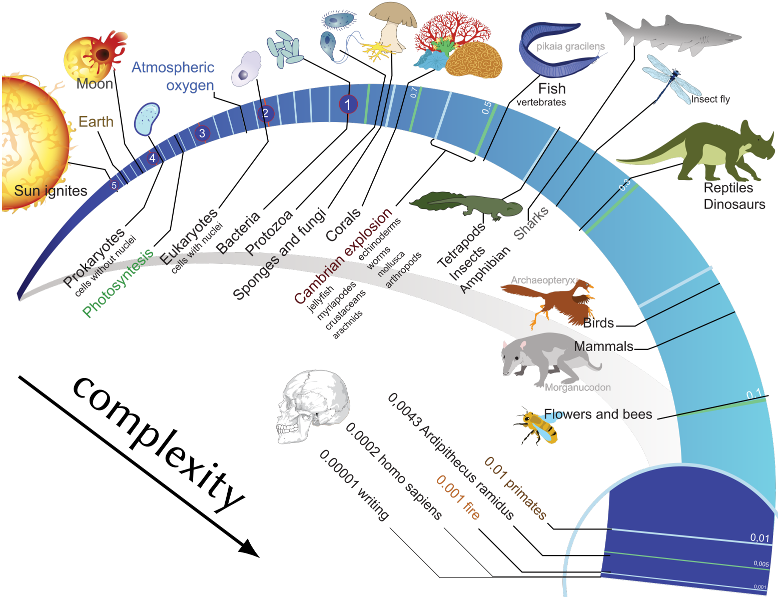
Understanding Evolution:

A Methodology for Evaluating the Extensibility of Boolean Networks' Structure and Function

Rémi Segretain[1]
Sergiu Ivanov[2]
Laurent Trilling[1]
Nicolas Glade[1]

[1]TIMC-IMAG, Univesité Grenoble-Alpes, France
[2]IBISC, Université Paris-Saclay, France

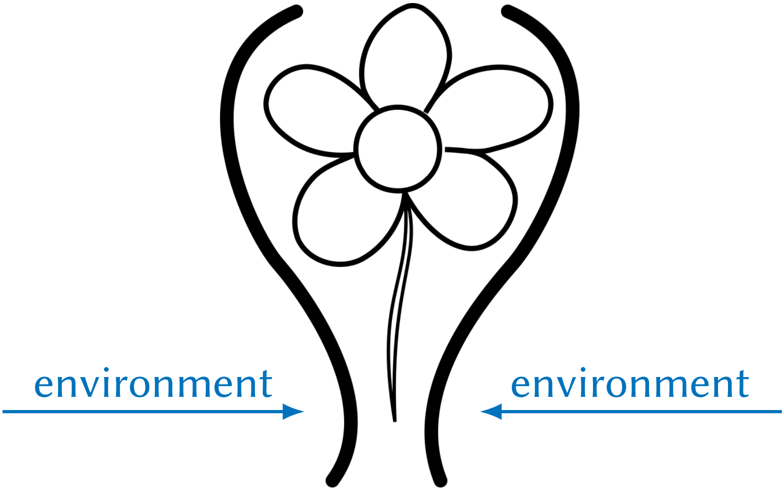
Evolution



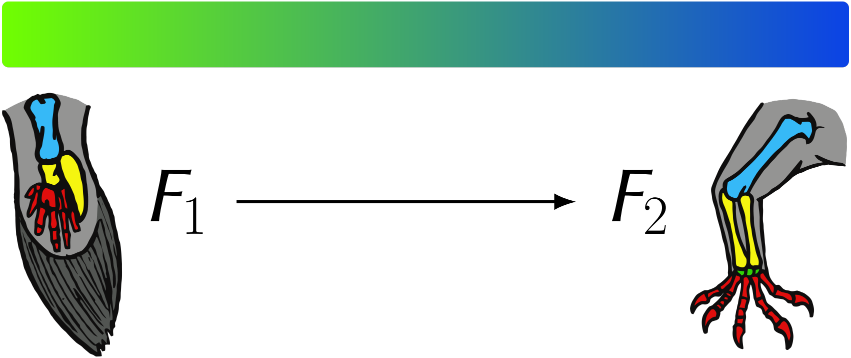
https://openclipart.org/download/320330/timelineevolutionoflife-pd-ladyofhats-wikimedia.svg

Natural selection

reinforces the features best suited to the environment.



Evolution is gradual

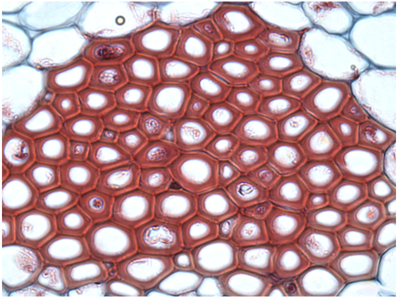


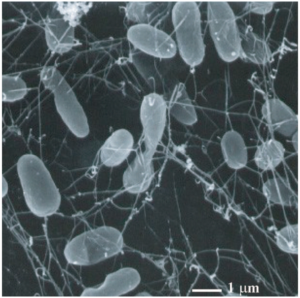
The cards are rarely entirely reshuffled: maintaining the functioning of the whole system is essential for survival.

https://en.wikipedia.org/wiki/File:Crossopterygii_fins_tetrapod_legs.svg

Evolution and cancer

Atavistic theory of cancer: cancerous cells devolve to primitive unicellular mode.


[1]



[2]

[1] https://en.wikipedia.org/wiki/File:Plant_cell_type_sclerenchyma_fibers.png
[2] https://en.wikipedia.org/wiki/File:Thermophile_bacteria2.jpg

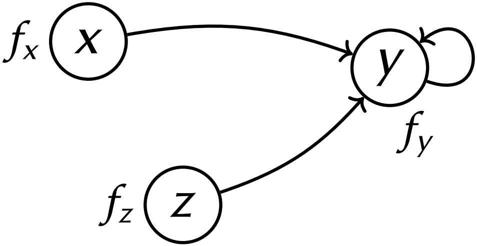
How does an organism grow more complex, while maintaining original structures and functions?

Which structures are the most extensible?

What language to answer this question?

Boolean networks

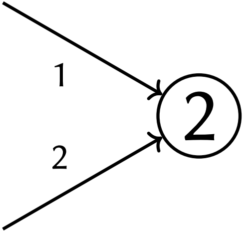
Boolean nodes with Boolean update functions.



$$x, y, z \in \{0, 1\}, \quad \text{a state: } \{x : 0, y : 1, z : 0\} \equiv 010$$

$$f_x, f_y, f_z : \{0, 1\}^3 \to \{0, 1\}$$

TBF: Threshold Boolean Functions

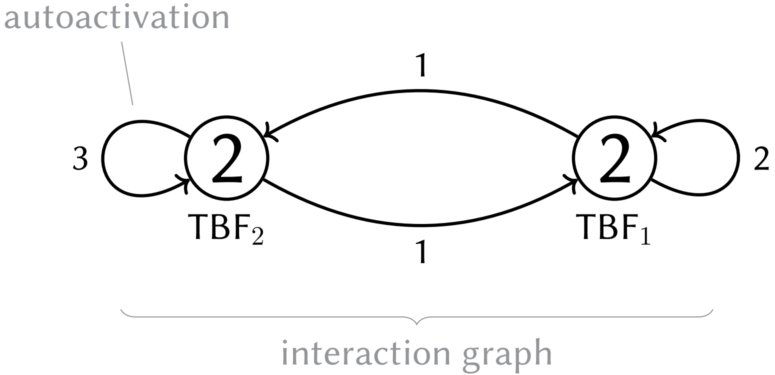


$$f(x_1, x_2) \equiv w_1 x_1 + w_2 x_2 > \theta$$

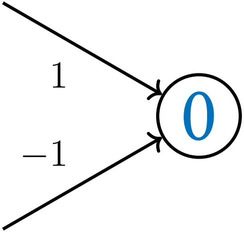
$$\begin{aligned} f &= (w_1, w_2, \theta) \\ w_1 &= 1 \\ w_2 &= 2 \\ \theta &= 2 \end{aligned}$$

w_1 x_1	w_2 x_2	f
0	0	0
0	1	0
1	0	0
1	1	1

TBN: Threshold Boolean networks



SBF/N: Sign Boolean functions/networks

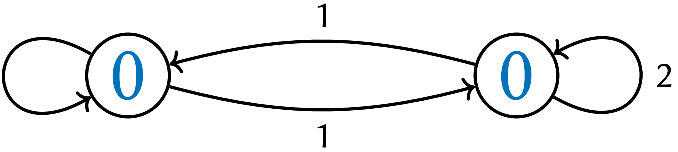


$$f(x_1, x_2) \equiv w_1 x_1 + w_2 x_2 > 0$$

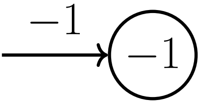
$$f = (w_1, w_2)$$
$$w_1 = 1$$
$$w_2 = -1$$

x_1	x_2	f
0	0	0
0	1	0
1	0	1
1	1	0

SBN:



SBF < TBF

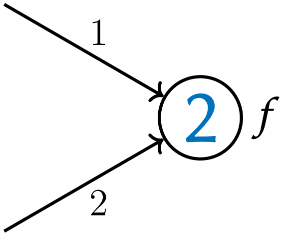


$$f = (w_1 = -1, \theta = -1) \equiv \neg x_1$$

x_1	f
0	1
1	0

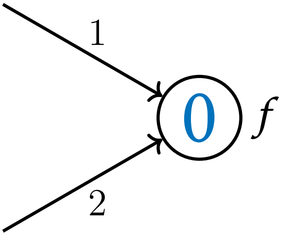
$$\forall f \in \mathsf{SBF} : f(\mathbf{0}) = 0$$

SBN vs. TBF



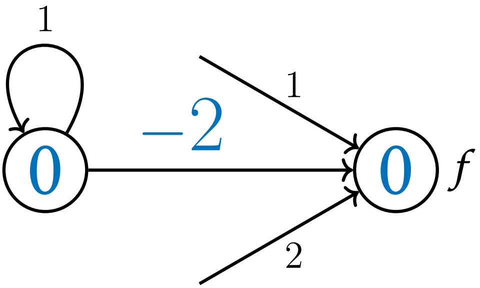
x_1	x_2	f
0	0	0
0	1	0
1	0	0
1	1	1

SBN vs. TBF



x_1	x_2	f
0	0	0
0	1	1
1	0	1
1	1	1

SBN vs. TBF



x_1	x_2	f
0	0	0
0	1	0
1	0	0
1	1	1

TBF = SBF + a hidden SBF

TBF vs. propositional BF

w_1	θ
-1	-1
x_1	$\neg x_1$
0	1
1	0

1	1	0
x_1	x_2	$x_1 \vee x_2$
0	0	0
0	1	1
1	0	1
1	1	1

1	1	1
x_1	x_2	$x_1 \wedge x_2$
0	0	0
0	1	0
1	0	0
1	1	1

$\{\neg, \vee, \wedge\}$ is a basis for BF

$\Longrightarrow$ any BF $=$ a combination of TBF (a TBN)

SBN vs. TBN vs. propositional BN

Propositional BN: much power hidden in the formula

TBN: the power is explicit in the connections

↪ the interaction graph completely specifies the SBN

SBN: a more symmetric and explicit model

↪ all numerical parameters have the same role

syntactic simplicity

---

power(SBN) = power(TBN) = power(propositional BN)

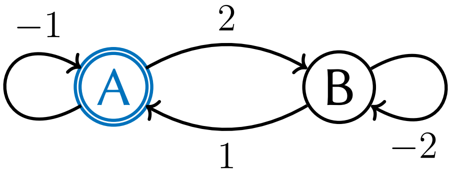
Update modes

We typically consider the synchronous mode

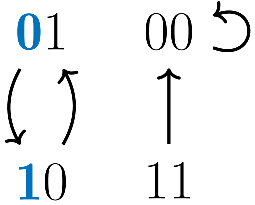
$\hookrightarrow$ all variables are updated at every step.

Other modes can be considered, too.

SBN: Transition graphs and outputs



interaction graph



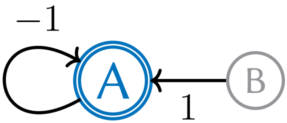
transition graph

Output = the binary sequence observed on 

Synchronous mode $\implies$ outputs $\subseteq uv^*$

$u \in \{0, 1\}^k$ – the pre-period $\quad v \in \{0, 1\}^m$ – the period

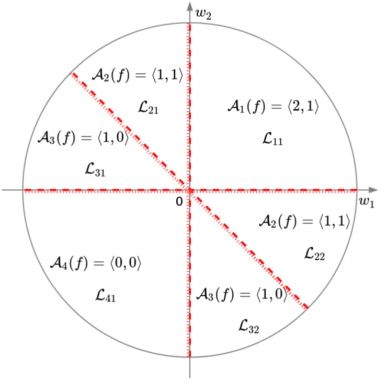
Equivalence by truth tables



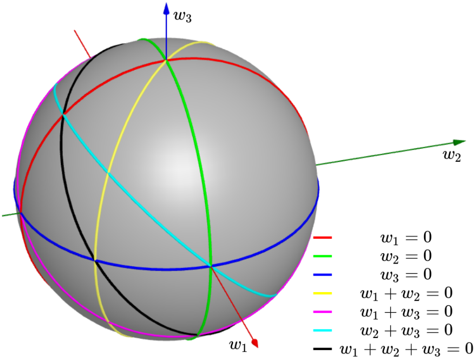
A	B	A	
0	0	0	$0 \leq 0$
0	1	1	$w_B > 0$
1	0	0	$w_A \leq 0$
1	1	0	$w_A + w_B \leq 0$

The inequalities determine subspaces of the parameter space = equivalence classes

Equivalence classes



$d = 2$

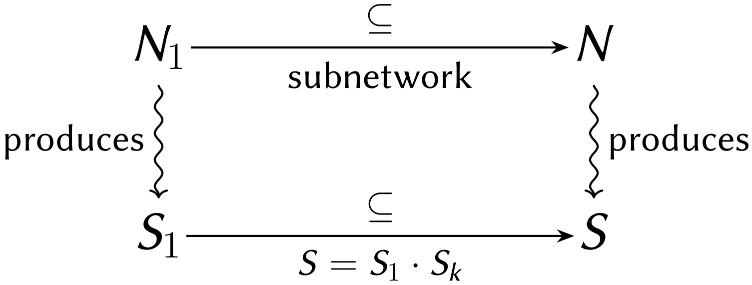


$d = 3$

Definition of extensibility in SBN

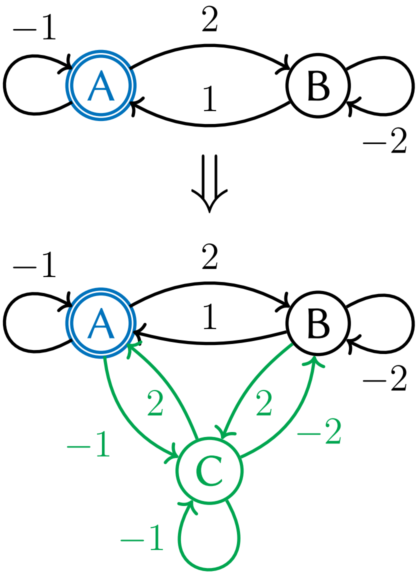
How can an SBN gain a new function, maintaining

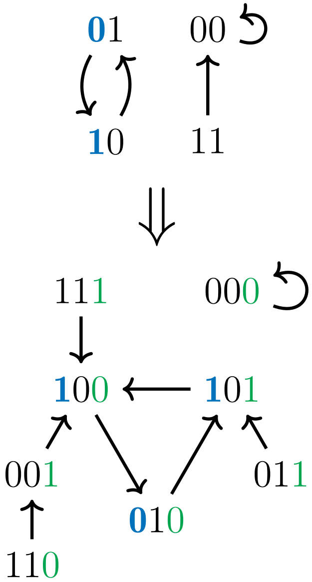
- the prior structures? (interaction graph)
- the prior functions?



commutative diagram

Extensibility in SBN: an example





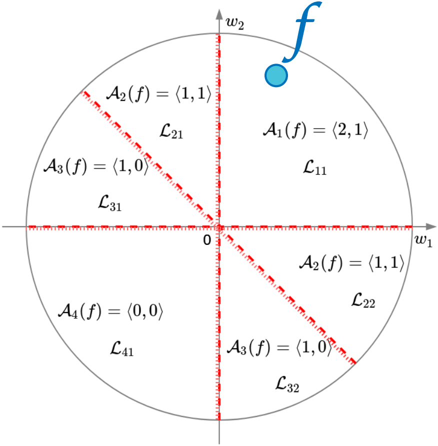
Growth of a network

more robustness? more functions?

Which structures are the most extensible?

Complexity is a good way to discriminate between SBN.

Complexity of an SBF

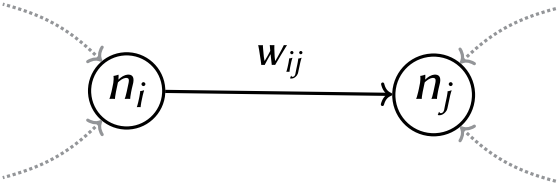


$\mathcal{C}^f$ = complexity(f) = probability of not picking f

$\hookrightarrow$ uniform distribution over the unit ball

Complex function $\implies$ low probability

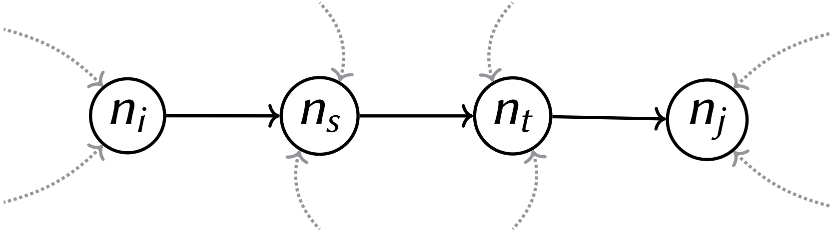
Complexity of an edge E_{ij} in SBN



The probability that n_i influences n_j via E_{ij}:

$$P^I(E_{ij}) = \frac{|w_{ij}|}{\sum_{k=1}^{d} |w_{kj}|}$$

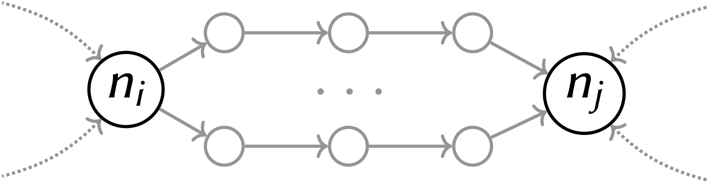
Complexity of a path in SBN



normalize by the length of the path

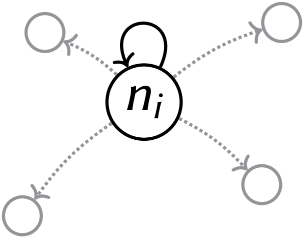
$$P^I(Path_{ab}) = \sqrt[3]{P^I(E_{is}) \cdot P^I(E_{st}) \cdot P^I(E_{tj})}$$

Probability that n_i influences n_j



$$P^I_{ab} = P\left(\bigcup_{Path \in Paths_{ij}} Path\right)$$

A centrality of a node



centrality(n_i) = probability that n_i influences at least one other node, including itself

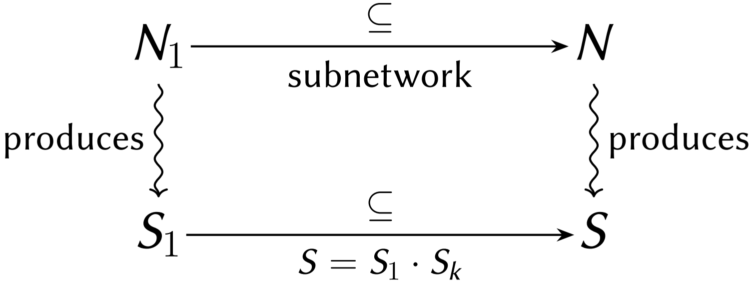
$$\mathcal{C}_i^s = P\left(\bigcup_{k=1}^{d}\left(\bigcup_{Path \in Paths_{ik}} Path\right)\right)$$

Complexity of an SBN

Product of complexities of connected nodes, modulated by their centralities:

$$\mathcal{C}(SBN) = \prod_{i=1}^{d} \begin{cases} \mathcal{C}_i^f \times \mathcal{C}_i^s & \text{if } \mathcal{C}_i^s > 0 \\ 1 & \text{otherwise} \end{cases}$$

Extensibility = function(complexity)



Which N_1 yield the most extensions?

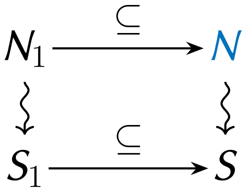
Hypothesis

Which N_1 yield the most extensions?

complexity ↓		
	simple N_1	too simple
	moderate N_1	extensibility sweet spot
	complex N_1	too constrained

Testing the hypothesis for $d = 2$

1. Enumerate all N_1.
2. For each N_1, fix an output node, S_1, and S.
3. Enumerate all N for which the diagram commutes:



4. Compute the complexities of N_1, N, S_1, and S.

a version of Kolmogorov complexity adapted to short strings

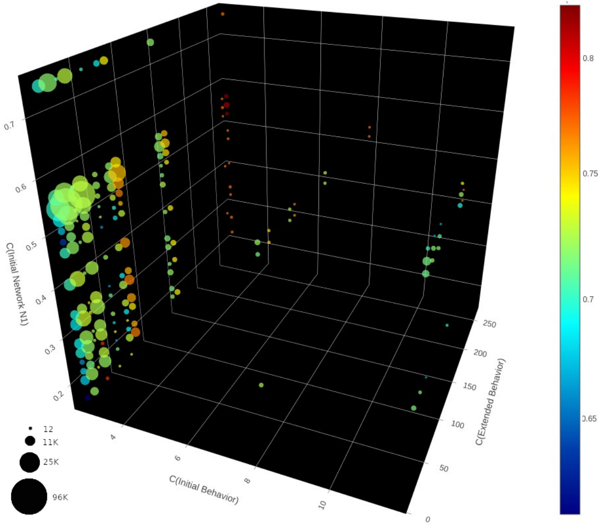
Soler-Toscano F., Zenil H., Delahaye J.-P., Gauvrit N.: Calculating Kolmogorov complexity from the output frequency distributions of small Turing machines; PLoS ONE 9(5): e96223 (2014)

Combinatorics of the method

d	SBF	SBN	$\{N_1, S_1, S\}$	N
2	7	101	$96 \cdot 10^3$	777216
3	17	$206 \cdot 10^3$	$[180 \cdot 10^6 – 32 \cdot 10^9]$	$[1.5 \cdot 10^9 – 280 \cdot 10^9]$
4	47	$76 \cdot 10^9$		

estimations

First results



Color = average complexity of SBN in the class

Point size = number of networks in the class (log-scale)

Discussion

- Behaviours of most $d = 2$ networks are not complex.
- Most extensions have moderate structural and behavioural complexity.
- Complex extended behaviors and networks are mostly obtained from networks of moderate complexity.

Improvement directions

- Refine the analysis.
- Look into higher dimensions.
- Consider other modes of evolution:
  - increase in the number of attractors,
  - increase in the size of basins of attraction.

simple ← ∇ → complex

??

Rémi Segretain  Sergiu Ivanov  Laurent Trilling  Nicolas Glade 

Special thanks to Ibrahim Cheddadi 

UGA Université Grenoble Alpes
Communauté UNIVERSITÉ Grenoble Alpes
RFSI réseau francilien en sciences informatiques
université evry val-d'essonne
université PARIS-SACLAY
îledeFrance
ANR AGENCE NATIONALE DE LA RECHERCHE