

On two information-theoretic measures of random fuzzy networks

Octavio Zapata¹, Hyobin Kim¹, Carlos Gershenson^{1,2}

¹Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, México

²Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, México
octavio.zapata@c3.unam.mx hyobin.kim@c3.unam.mx cgg@unam.mx

Suppose we have to our disposal a **genome** of size n , that is to say, a set of genetic material sequenced into n genes. Further assume that (i) we can perform a measurement which assigns a non-negative real value called **gene expression** to every gene in the genome, (ii) we can perform a joint measurement of the gene expression of all the genes at once, and (iii) we can repeat these joint measurements and keep record of our observations.

Let $I := \{1, \dots, n\}$ be a set of n indices. We identify each index $i \in I$ with a single gene in the genome.

Definition 1 (State). A **state** is a tuple $x = (x_i)_{i \in I}$ with $x_i \in [0, 1]$. We say that a state $x \in [0, 1]^n$ is **Boolean** if $x_i \in \{0, 1\}$ for all $i \in I$.

A countable collection of states represents the time evolution of the gene expression of the whole genome, measured at discrete instants. In the real world, the expression of each gene is affected by the expression of other genes in a process called gene regulation. This gives rise to a network-like structure called gene regulatory network. Random Boolean networks were introduced in (Kauffman, 1969) as a model for gene regulatory networks.

Definition 2 (Random Boolean network). A **random Boolean network** with parameters n and k , $1 \leq k \leq n$, is a function $F : \{0, 1\}^n \rightarrow \{0, 1\}^n$ mapping Boolean states into themselves, specified as follows:

- for each gene $i \in I$, we choose independently and uniformly a subset of genes $\text{re}(i) \subseteq I$ such that $|\text{re}(i)| = k$; the genes in $\text{re}(i)$ are called the **regulators** of i ;
- for each gene $i \in I$, we choose independently and uniformly a Boolean function $f_i : \{0, 1\}^k \rightarrow \{0, 1\}$ called the **regulating function** of i ;
- for each Boolean state $x \in \{0, 1\}^n$, we let $F(x) := (f_i(\check{x}_i))_{i \in I}$ where $\check{x}_i := (x_j)_{j \in \text{re}(i)}$ denotes the tuple in $\{0, 1\}^k$ obtained from x by selecting only the entries whose indices correspond to the regulators of gene i .

Every random Boolean network can be extended to a function defined on arbitrary (not only Boolean) states. Furthermore, this process can be done in such a way that the

value of the Boolean network coincides with the value of the extended function for all Boolean states.

Theorem 1 (Random fuzzy network). *For each random Boolean network $F : \{0, 1\}^n \rightarrow \{0, 1\}^n$ there exists a function $\hat{F} : [0, 1]^n \rightarrow [0, 1]^n$ satisfying $\hat{F}(x) = F(x)$ for all $x \in \{0, 1\}^n$.*

The proof of this elementary statement is our main contribution. We provide 3 non-trivial ways of constructing such extensions. Let \wedge, \vee , and \neg denote the **Boolean operations** of conjunction (AND), disjunction (OR), and negation (NOT), defined by the following truth-tables:

x, y	$x \wedge y$	$x \vee y$
0, 0	0	0
0, 1	0	1
1, 0	0	1
1, 1	1	1

x	$\neg x$
0	1
1	0

Lemma 1 (Disjunctive normal form). *For every Boolean function $f : \{0, 1\}^k \rightarrow \{0, 1\}$ and $x \in \{0, 1\}^k$ the following functional equation holds:*

$$f(x) = \bigvee_{y: f(y)=1} \bigwedge_{1 \leq j \leq k} t_j^{x,y} \quad (1)$$

with

$$t_j^{x,y} := \begin{cases} x_j & \text{if } y_j = 1, \\ \neg x_j & \text{if } y_j = 0. \end{cases}$$

The right-hand side of equation (1) is called the **disjunctive normal form** of f .

Definition 3 (Triangular norm). A **triangular norm** is a binary operation $\odot : [0, 1]^2 \rightarrow [0, 1]$, such that for all $x, y, z \in [0, 1]$ the following axioms hold:

1. (comutativity) $x \odot y = y \odot x$;
2. (associativity) $x \odot (y \odot z) = (x \odot y) \odot z$;
3. (neutral element) $x \odot 1 = x$;
4. (monotonicity) $y \leq z \implies x \odot y \leq x \odot z$.

	$x \otimes y$	$x \odot y$	$\dot{\neg}x$
Gödel	$\min\{x, y\}$	$\max\{x, y\}$	$1 - x$
Łukasiewicz	$\max\{0, x + y - 1\}$	$\min\{1, x + y\}$	$1 - x$
Product	xy	$x + y - xy$	$1 - x$

Table 1: Fuzzy logic operations provided by different triangular (co)norms and with standard negation.

Triangular norms were introduced in (Menger, 1942), in order to study metric spaces where the distance between two elements is represented with a probability distribution rather than a non-negative real number. A **triangular conorm** is a binary operation $\odot : [0, 1]^2 \rightarrow [0, 1]$ satisfying commutativity, associativity and monotonicity, and with 0 as neutral element, i.e. for all $x \in [0, 1]$, $x \odot 0 = x$. The Boolean operations of conjunction and disjunction satisfy the axioms of triangular norms and conorms, respectively. That is, for all $x, y \in \{0, 1\}$ we have $x \wedge y = x \otimes y$ and $x \vee y = x \odot y$.

Definition 4 (Strong complement). A **strong complement** is a unary operation $\dot{\neg} : [0, 1] \rightarrow [0, 1]$, such that for all $x, y \in [0, 1]$ the following axioms are satisfied:

1. (order-reversing) $x \leq y \implies \dot{\neg}x \geq \dot{\neg}y$;
2. (boundary conditions) $\dot{\neg}0 = 1$ and $\dot{\neg}1 = 0$;
3. (idempotence) $\dot{\neg}(\dot{\neg}x) = x$.

The strong negation defined as $\dot{\neg}x := 1 - x$ is called **standard negation**. Boolean conjunction (disjunction) can be suitably extended to the unit interval $[0, 1]$ using triangular (co)norms. Boolean negation is a strong complement and is easy to see that it coincides with standard negation.

Definition 5 (Fuzzyfication). Let $f : \{0, 1\}^k \rightarrow \{0, 1\}$ be a Boolean function. A **fuzzyfication** of f is a function $\hat{f} : [0, 1]^k \rightarrow [0, 1]$ obtained by interpreting the Boolean operations \wedge, \vee , and \neg in the disjunctive normal form of f with a triangular norm \otimes , a triangular conorm \odot , and a strong complement $\dot{\neg}$ according to Table 1.

Proof of Theorem 1. Given a random Boolean network F , let \hat{F} be the function obtained by fuzzyfication of the regulating functions of F . That is, for all $x \in [0, 1]^n$ we let

$$\hat{F}(x) := \left(\hat{f}_i(\check{x}_i) \right)_{i \in I}, \quad \hat{f}_i(\check{x}_i) := \bigodot_{y: f_i(y)=1} \bigotimes_{1 \leq j \leq k} u_j^{\check{x}_i, y}$$

with $u_j^{\check{x}_i, y} = \check{x}_{i,j}$ if $y_j = \dot{\neg}\check{x}_{i,j}$ and $u_j^{\check{x}_i, y} = 1$ if $y_j = 0$, where $\check{x}_{i,j} \in [0, 1]$ denotes the state of gene expression of the j -th regulator of gene i . \square

The particular choice of triangular (co)norms (Gödel, Łukasiewicz and product) provided by our model is justified by the next result. For more on triangular norms, see e.g. the monograph (Klement et al., 2013).

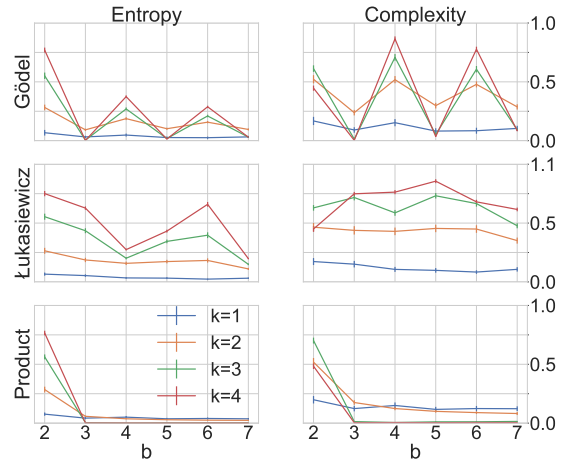


Figure 1: Information-theoretic measures of random fuzzy networks with $n = 15$ and different choice of logic operations (Gödel, Łukasiewicz, and product), depending on $b \in \{2, \dots, 7\}$ (ranging on the x -axis) and $k \in \{1, \dots, 4\}$ (colour-coded).

Theorem 2 (Mostert–Shields). A triangular norm is continuous if and only if it is isomorphic to an ordinal sum of the Gödel, Łukasiewicz, and product triangular norm.

Back into our experimental scenario, let b be the number of gene expression values that our measurement apparatus is able to distinguish. If $b = 2$ then we can see Boolean states only. More generally, for $b \geq 2$ the state of expression of gene $i \in I$ shall be represented with a variable x_i taking values in the set $\{0, \frac{1}{b-1}, \dots, \frac{b-2}{b-1}, 1\}$.

Following (Santamaría-Bonfil et al., 2017), we wrote a program that simulates random fuzzy networks and computes their **entropy** H and **complexity** $C := 4H(1 - H)$. The dynamics of gene expression for cell functions is characterised by these two measures. From a biological perspective, H measures how flexibly the information of gene expression is changed, and $1 - H$ measures how stably the information is maintained. Therefore, complexity C can be regarded as a metric to assess a balance between the flexibility and stability of cell functions.

Figure 1 shows the entropy and complexity of random fuzzy networks depending on different logic operations. For all the logics, the entropy is highest with $b = 2$, which means that gene expression values change most frequently at each time step. However, Gödel has cyclic ups and downs at even and odd values of b , while Łukasiewicz has more stable entropy across all values of b , especially for low k . For product operations, entropy and complexity are very low for $b > 2$. As a result, Łukasiewicz has more stable complexity across the range of b , and thus the balance between flexibility and stability is obtained more robustly. These observations are in accordance with the statistics of attractors and families for Gödel operations reported in (Zapata and Gershenson, 2014). We shall revisit this statistics with respect to Łukasiewicz and product operations in future work.

References

- Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of theoretical biology*, 22(3):437–467.
- Klement, E. P., Mesiar, R., and Pap, E. (2013). *Triangular norms*, volume 8. Springer Science & Business Media.
- Menger, K. (1942). Statistical metrics. *Proceedings of the National Academy of Sciences of the United States of America*, 28(12):535.
- Santamaría-Bonfil, G., Gershenson, C., and Fernández, N. (2017). A package for measuring emergence, self-organization, and complexity based on Shannon entropy. *Frontiers in Robotics and AI*, 4:10.
- Zapata, O. and Gershenson, C. (2014). Random fuzzy networks. In *Artificial Life Conference Proceedings 14*, pages 427–428. MIT Press.