DETERMINATIVE POWER AND TOLERANCE TO PERTURBATIONS IN BOOLEAN NETWORKS

Reinhard Heckel[1], Steffen Schober[2] and Martin Bossert[2]

[1]Department of Information Technology and Electrical Engineering, ETH Zurich,
[2]Institute of Telecommunications and Applied Information Theory, University of Ulm
heckel@nari.ee.ethz.ch, steffen.schober@uni-ulm.de, martin.bossert@uni-ulm.de

ABSTRACT

Consider a large Boolean network with a feed forward structure. Given a probability distribution on the inputs, can one find—possibly small—collections of input nodes that determine the states of most other nodes in the network? To answer this question, a notion that quantifies the *determinative power* of an input over the states in the network is needed. We argue that the mutual information (MI) between a given subset of the inputs $\mathbf{X} = \{X_1, ..., X_n\}$ of some node i and its associated function $f_i(\mathbf{X})$ quantifies the determinative power of this subset of inputs over node i. We compare the determinative power of a set of input nodes to the sensitivity to perturbations to this input nodes, and find that, maybe surprisingly, an input that has a large sensitivity to perturbations does not necessarily have large determinative power. However, for *unate* functions, which play an important role in genetic regulatory networks, we find a direct relation between MI and sensitivity to perturbations. As an application of our methods, we analyze the large-scale regulatory network of *E. coli* numerically: We identify the most determinative nodes and show that a small set of those reduces the overall uncertainty of network states significantly.

1. INTRODUCTION

A Boolean network (BN) is a discrete dynamical system, which is often used to study and model a variety of biochemical networks. BNs have been introduced by Kauffman [1] as models of gene regulatory networks. Amongst others, they are used to model large-scale networks such as the *Escherichia coli* regulatory network [2] which is analyzed in Sec. 4. In the analysis of BNs, it is common to consider measures that quantify the effect of perturbations, whereas determinative power has not received much attention, even though there are several settings where such a notion is of interest, e.g., the following. Given a feed-forward network where the states of the nodes are controlled by the states in the input layer, we might ask whether a possibly small set of inputs suffices to determine most states, i.e., reduces the uncertainty about the network's states significantly. This can be addressed by quantifying the determinative power of the input nodes. For example, in the *E. coli* regulatory network it turns out that a small set of metabolites and other inputs determine most genes that account for *E. coli*'s metabolism (see Sec. 4).

In this paper, we view the state of each node in the network as an independent random variable. This modeling assumption applies e.g. for networks with a tree-like topology, and is standard when studying the effect of perturbations. For this setting, determinative power of nodes and measures of perturbations are properties of single functions, hence the analysis of the BN reduces to the analysis of single functions. As the main tool for the analysis, we use Fourier analysis of Boolean functions. Fourier analytic techniques were first applied in the context of Boolean networks by Kesseli et al. [3], and also in [4]. However, the setting and problems addressed in [3, 4] are different to the problem considered here.

The contributions of this paper are as follows. We argue that the MI between a set of nodes and the state of a node is a measure of the determinative power of this set of inputs, as MI is a measure that quantifies the mutual dependence of random variables. If a set of inputs to a node and the state of this node are strongly mutually dependent, then this set can be viewed as having large determinative power over this node. To understand determinative power and mutual dependencies in Boolean networks better, we systematically study the MI of a sets of inputs and the state of a node. We relate the mutual information to measures of perturbations, and prove that—maybe surprisingly—a set of inputs that is highly sensitive to perturbations, might not necessarily have determinative power. Conversely, an input that has determinative power, must be sensitive to perturbations to some extent. These results are proven, using Fourier analytic techniques. Moreover we show, that for the class of *unate* functions, which model functional dependencies in gene regulatory networks well, any input and the function's output are statistically dependent. For unate functions we also prove a direct relation between the mutual information and the influence of a variable. As an application of the theoretical results in this paper, we show that mutual information can be used to identify the determinative nodes in the large-scale model of the control network of *E. coli's* metabolism [2]. Due to limited space, proofs and a more detailed exposition are omitted but can be found in the preprint [5].

2. PRELIMINARIES

We start by shortly stating some standard facts about BNs and Fourier analysis of Boolean functions and introduce

notation. A Boolean network (BN) can be viewed as a collection of n nodes with memory. The state of a node i is described by a binary state $x_i(t) \in \{-1, +1\}$ at discrete time $t \in \mathbb{N}$. Choosing the alphabet to be $\{-1, +1\}$ rather then $\{0, 1\}$ as more common in the literature on BNs, will turn out to be advantageous later. However, both choices are equivalent. In most of the Boolean network models used in biology, $f_i(\mathbf{x})$ does not depend on all arguments $x_1, ..., x_n$, but on a small subset only. Obviously, to study determinative power and tolerance to perturbations, a probabilistic setup is needed. In our analysis, we assume that each state is an independent random variable X_i which follows the distribution $\Pr[X_i = x_i]$, $x_i \in \{-1, +1\}$. The assumption of independence holds for networks with tree-like topology, but is not feasible for networks with strong local dependencies. In many relevant settings a BN has a tree-like topology, for instance the *E. coli* network which is analyzed in Sec. 4.

Notation. We use $[n]$ for the set $\{1, 2, ..., n\}$, and all sets occurring in this paper are subsets of $[n]$. With $\sum_{S \subseteq A}$ we mean the sum over all sets S that are subsets of A. Throughout this paper, we use capital letters for random variables, e.g., X, and lower case letters for their realizations, e.g., x. Boldface letters denote vectors, e.g., $\mathbf{X}$ is a random vector, and $\mathbf{x}$ its realization. For a vector $\mathbf{x}$ and a set $A \subseteq [n]$, $\mathbf{x}_A$ denotes the subvector of $\mathbf{x}$ corresponding to the entries indexed by A.

Fourier Analysis of Boolean Functions. Let $\mathbf{X} = (X_1, ..., X_n)$ be a binary, product distributed random vector, i.e., the entries of $\mathbf{X}$ are independent random variables $X_i, i \in [n]$ with distribution $\Pr[X_i = x_i]$, $x_i \in \{-1, +1\}$. Throughout this paper, probabilities $\Pr[\cdot]$ and expectations $\mathbb{E}[\cdot]$ are with respect to a product distributed $\mathbf{X}$. We denote $p_i \triangleq \Pr[X_i = 1]$, $\mathrm{Var}(X_i)$ as the variance of X_i, $\sigma_i \triangleq \sqrt{\mathrm{Var}(X_i)}$ as its standard deviation and finally, $\mu_i \triangleq \mathbb{E}[X_i]$ as its mean. The inner product of the Boolean functions $f, g \colon \{-1, +1\}^n \to \{-1, +1\}$ with respect to the distribution of $\mathbf{X}$ is defined as $\langle f, g \rangle \triangleq \mathbb{E}[f(\mathbf{X})g(\mathbf{X})]$ which induces the norm $\|f\| = \sqrt{\langle f, f \rangle}$. An orthonormal basis is given by the functions [6]

$$\Phi_S(\mathbf{x}) = \prod_{i \in S} \frac{x_i - \mu_i}{\sigma_i}, \quad S \subseteq [n] \setminus \emptyset$$

and $\Phi_S(\mathbf{x}) = 1$, $S = \emptyset$. Thus, each function f can be uniquely expressed as $f(\mathbf{x}) = \sum_{S \subseteq [n]} \hat{f}(S)\Phi_S(\mathbf{x})$, where $\hat{f}(S) \triangleq \langle f, \Phi_S \rangle$ are the Fourier coefficients. Note that this is a representation of the function f as a multilinear polynomial, and the Fourier coefficients are the coefficients of that polynomial.

Influence and Average Sensitivity. Next, we discuss measures of perturbations and their relation to the Fourier spectrum. We start with the *influence* of variable i, which is defined as [6] $I_i(f) \triangleq \Pr[f(\mathbf{X}) \neq f(\mathbf{X} \oplus e_i)]$, where $\mathbf{x} \oplus e_i$ is the vector obtained from $\mathbf{x}$ by flipping its ith entry. By definition, the influence of variable i is the probability that a perturbation of input i, i.e., flipping input i,

changes the function's output. Hence influence captures the effect of a single perturbation of input i. In Boolean networks it is common to study the sum of all influences, i.e., the *average sensitivity* of function f. The average sensitivity of f to the variables in set A is defined as

$$I_A(f) \triangleq \sum_{i \in A} I_i(f)$$

and captures whether flipping an input, chosen uniformly at random from the set A affects the function's output. Most commonly all inputs are taken into account, i.e., the average sensitivity of f, $as(f) \triangleq I_{\{1,...,n\}}(f)$ is studied. The average sensitivity with respect to A (and hence the influence, by setting $A = \{i\}$) can be expressed in terms of Fourier coefficients as

$$I_A(f) = \sum_{S \subseteq [n]} \hat{f}(S)^2 \sum_{i \in S \cap A} \frac{1}{\sigma_i^2}. \tag{1}$$

From (1) (by setting $A = \{1, .., n\}$) it becomes apparent that the average sensitivity $as(f)$ is large if the sum over the squared Fourier coefficients $\hat{f}(S)^2$ of high degree $d = |S|$, is large. As $\sum_{S \subseteq [n]} \hat{f}(S)^2 = 1$, the terms $\hat{f}(S)^2$ for which the degree $d = |S|$ is small must then be small. Hence for f to be tolerant to single perturbations, i.e., to have a small average sensitivity, the Fourier coefficients must be concentrated on coefficients with low degree. Let's see an example: Suppose $p_1 = p_2 = p_3 = 1/2$ and consider the AND3 function, i.e., $f_{AND3}(x_1, x_2, x_3) = 1$ if and only if $x_1 = x_2 = x_3 = 1$. The average sensitivity of the AND3 function is $as(f_{AND3}) = 0.75$. Hence, f_{AND3} is tolerant to perturbations. The spectrum of f_{AND3} is concentrated on the coefficients of low degree. In contrast, consider the parity of three variables: $f_{PARITY3}(x_1, x_2, x_3) = x_1 x_2 x_3$, for which $as(f_{PARITY3}) = 3$. Hence, PARITY3 is maximal sensitive to perturbations. The spectrum of the PARITY3 function is maximal concentrated on the coefficient of highest degree as $\hat{f}(\{1, 2, 3\}) = 1$.

3. MAIN RESULTS

In this section, we study the mutual information $\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A)$ between $f(\mathbf{X})$ and $\mathbf{X}_A$, where $\mathbf{X}_A$ consists of the entries of $\mathbf{X}$ corresponding to the indices in the set $A \subseteq [n]$. Proofs are omitted due to space limitations; those and further details can be found in [5]. We start by defining the mutual information. Mutual information is the reduction of uncertainty of a random variable Y due to the knowledge of X, hence we define a measure of uncertainty first, which is entropy. As a reference for the following definitions see [7]. The entropy $H(X)$ of a discrete random variable X with alphabet $\mathcal{X}$ is defined as $H(X) \triangleq -\sum_{x \in \mathcal{X}} \Pr[X = x] \log_2 \Pr[X = x]$. The conditional entropy $H(Y|X)$ of a pair of discrete and jointly distributed random variables (Y, X) is defined as $H(Y|X) \triangleq \sum_{x \in \mathcal{X}} \Pr[X = x] H(Y|X = x)$. Finally, the mutual information $\mathrm{MI}(Y; X)$ between Y and X is defined as $\mathrm{MI}(Y; X) \triangleq H(Y) - H(Y|X)$. For a binary random variable X with alphabet $\mathcal{X} = \{x_1, x_2\}$ and

$p \triangleq \Pr\left[X = x_1\right]$, we have $H(X) = h(p)$, where $h(p)$ is the binary entropy function, defined as

$$h(p) \triangleq -p \log_2(p) - (1-p) \log_2(1-p). \quad (2)$$

Mutual information is a measure of determinative power because of the following reasons. Consider a single variable X_i of the argument $\mathbf{X}$: If knowledge of X_i reduces the uncertainty of $f(\mathbf{X})$, then X_i determines the state of $f(\mathbf{X})$ to some extent, because then knowledge about the state of X_i helps in predicting $f(\mathbf{X})$. Furthermore, we require from a measure of determinative power, that not all variables can have large determinative power simultaneously. This is guaranteed for mutual information, as

$$\sum_{i=1}^{n} \mathrm{MI}(f(\mathbf{X}); X_i) \leq \mathrm{MI}(f(\mathbf{X}); \mathbf{X}) \leq 1, \quad (3)$$

which follows from the chain rule [7] of mutual information and independence of the variables $X_i, i \in [n]$. Hence, if $\mathrm{MI}(f(\mathbf{X}); X_i)$ is large, i.e., close to 1, we can be sure that X_i has some determinative power over $f(\mathbf{X})$, since (3) implies that $\mathrm{MI}(f(\mathbf{X}); X_j)$ must be small for $j \neq i$. Influence lacks this property: Each input can have large influence. An example is the parity function, where each input has influence 1. If variable i has large influence, this just implies that input i has power to change the output, but not to determine it.

Our results are based on the following novel characterization of the mutual information in terms of Fourier coefficients: Let $\mathbf{X}$ be product distributed and let $\mathbf{X}_A = \{X_i : i \in A\}$ be a fixed set of arguments, where $A \subseteq [n]$. Then

$$\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A) = h\left(1/2(1 + \hat{f}(\emptyset)\right)$$
$$- \mathbb{E}\left[h\left(\frac{1}{2}\left(1 + \sum_{S \subseteq A} \hat{f}(S)\Phi_S(\mathbf{X}_A)\right)\right)\right] \quad (4)$$

where $h(\cdot)$ is the binary entropy function as defined in (2).

Let us start with discussing $\mathrm{MI}(f(\mathbf{X}); X_i)$, based on (4). As seen by (4), $\mathrm{MI}(f(\mathbf{X}); X_i)$ just depends on $\hat{f}(\{i\})$, $\hat{f}(\emptyset)$ and p_i. In Figure 1 we depict $\mathrm{MI}(f(\mathbf{X}); X_i)$ for $p_i = 0.3$ as a function of $\hat{f}(\{i\})$ and $\hat{f}(\emptyset)$. It can be seen that $\mathrm{MI}(f(\mathbf{X}); X_i) = 0$, i.e., $f(\mathbf{X})$ and X_i are statistically independent, if and only if $\hat{f}(\{i\}) = 0$. Furthermore it is seen that $\mathrm{MI}(f(\mathbf{X}); X_i)$ is increasing in $|\hat{f}(\{i\})|$. Both observations can be proven rigorously. Hence X_i has large determinative power, i.e., $\mathrm{MI}(f(\mathbf{X}); X_i)$ is large, if and only if $|\hat{f}(\{i\})|$ is large (i.e., close to one).

Next, let us consider the (trivial) case where $A = [n]$ and hence $\mathbf{X}_A = \mathbf{X}$. Then $\mathrm{MI}(f(\mathbf{X}); \mathbf{X}) = h(1/2(1 + \hat{f}(\emptyset))$. It follows that $\mathrm{MI}(f(\mathbf{X}); \mathbf{X})$ is maximized for $\hat{f}(\emptyset) = 0$, i.e, $\Pr\left[f(\mathbf{X}) = 1\right] = 1/2$, i.e., if the variance of $f(\mathbf{X})$ is 1. In general, the closer to zero $\hat{f}(\emptyset)$ is, the larger the mutual information between a function's output and all its inputs.

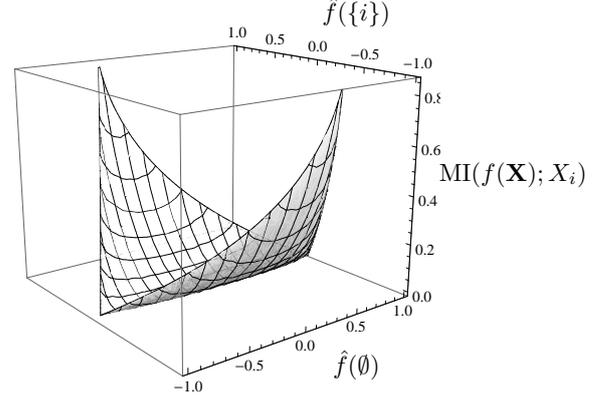We continue with studying the relation of mutual information and average sensitivity.

Figure 1. $\mathrm{MI}(f(\mathbf{X}); X_i)$ as a function of $\hat{f}(\{i\})$ and $\hat{f}(\emptyset)$ for $p_i = 0.3$.

Theorem 1. *For any Boolean function f, for any product distributed $\mathbf{X}$,*

$$I_A(f) \geq \min_{i \in A}\left(\frac{1}{\sigma_i^2}\right)\left(\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A) - \Psi\left(\mathrm{Var}\left(f(\mathbf{X})\right)\right)\right)$$
$$(5)$$

with $\Psi(x) \triangleq (x)^{1/\ln(4)} - x$.

The term $\Psi(\mathrm{Var}\left(f(\mathbf{X})\right))$ should be understood as an error term which satisfies $0 \leq \Psi(\mathrm{Var}\left(f(\mathbf{X})\right)) < 0.12$ and which is close to zero for situations (i.e., functions and distributions of $\mathbf{X}$) of interest. Theorem 1 shows that a large value of $\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A)$ implies that f must be sensitive to perturbations of the entries of $\mathbf{X}_A$. Moreover, if $I_A(f)$ is small, i.e., if f is tolerant to perturbations of the entries of $\mathbf{X}_A$, then $\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A)$ must be small, i.e., the entries of $\mathbf{X}_A$ do not have large determinative power. For the case that $A = [n]$, Theorem 1 states that $as(f)$ is lower-bounded by $\mathrm{MI}(f(\mathbf{X}); \mathbf{X})$ minus some small term.

Again, we discuss the special case that $A = \{i\}$. Theorem 1 evaluated for the case that $A = \{i\}$ yields that

$$I_i(f) \geq 1/\sigma_i^2(\mathrm{MI}(f(\mathbf{X}); X_i) - \Psi(\mathrm{Var}\left(f(\mathbf{X})\right)))$$

which shows that if $\mathrm{MI}(f(\mathbf{X}); X_i)$ is large, then $I_i(f)$ is also large. That proves the intuitive idea that if an input determines $f(\mathbf{X})$ to some extent, this input also has to be sensitive to errors. Conversely, an input i can have large influence and still $\mathrm{MI}(f(\mathbf{X}); X_i) = 0$. An example of such a function is the PARITY function, where $I_i(f) = 1$ and $\mathrm{MI}(f(\mathbf{X}); X_i) = 0$.

Interestingly, the influence also has an information theoretic interpretation:

$$I_i(f) = \frac{H\left(f(\mathbf{X})|\mathbf{X}_{[n]\setminus\{i\}}\right)}{H(X_i)}$$

which shows that the influence of a variable is a measure for the uncertainty of the function's output that remains if all variables except variable i are set.

Finally, we characterize statistical independence of $f(\mathbf{X})$ and a set of its arguments $\mathbf{X}_A$ in terms of Fourier coefficients. This result generalizes a theorem derived by

Xiao and Massey [8] from uniform to product distributed $\mathbf{X}$.

Theorem 2. *Let $A \subseteq [n]$ be fixed, f be a Boolean function and $\mathbf{X}$ be product distributed. Then $f(\mathbf{X})$ and the inputs $\mathbf{X}_A = \{X_i : i \in A\}$ are statistically independent if and only if*

$$\hat{f}(S) = 0 \text{ for all } S \subseteq A \setminus \emptyset.$$

Theorem 2 shows that if a function is concentrated on the coefficients of low degree $d = |S|$, which is the case for functions that are tolerant to perturbations, then small sets of inputs and the function's output are statistically dependent.

Unate Functions. A Boolean function f is said to be unate in variable x_i if for each $\mathbf{x} = (x_1, ..., x_n) \in \{-1, +1\}^n$ and for some fixed $a_i \in \{-1, +1\}$, $f(x_1, ..., x_i = -a_i, ..., x_n) \leq f(x_1, ..., x_i = a_i, ..., x_n)$. The function f is said to be unate, if f is unate in each variable x_i. For example, each linear threshold function and each nested canalizing function is unate and one can suppose that the majority of regulatory interactions in a biological network are unate. The basic argument is that if an element acts either as a repressor or an activator for some gene, but never as both, then the function determining the gene's state is unate by definition. For unate functions, we have that $\hat{f}(\{i\}) = a_i \sigma_i I_i(f), \ \forall i \in [n]$ where $a_i \in \{-1, +1\}$ is the parameter as given in the definition above. The proof goes along the same lines as the proof for monotone functions in [6, Lem. 4.5]. With (4) this yields an explicit relation of $I_i(f)$ and $\mathrm{MI}(f; X_i)$, based on which we find that for unate functions, the mutual information $\mathrm{MI}(f; X_i)$ is increasing in the influence $|I_i(f)|$. Moreover if f is unate, and x_i is a relevant variable, i.e., a variable on which the function actually depends on, then $|\hat{f}(\{i\})| > 0$. We furthermore find that if f is unate, the statement "x_i is a relevant variable" is equivalent to $\mathrm{MI}(f(\mathbf{X}); X_i) \neq 0$. In a Boolean model of a biological regulatory network, this implies that if the functions in the network are unate, then a regulator and the target gene must be statistically dependent.

4. *E. COLI* REGULATORY NETWORK

In [2], the authors presented a complex computational model of the *E. coli* transcriptional regulatory network that controls central parts of the *E. coli* metabolism. The network consists of 798 nodes and 1160 edges and has a layered feed-forward structure, i.e., no feedback-loops exist. The 133 elements in the first layer can be viewed as the inputs of the system and the elements in the following 7 layers are interacting genes representing the internal state of the system. Our investigations showed that all functions are unate, which is a non-typical property of the network. We identified the input-nodes that have large determinative power using the MI. To this end, we define the determinative power of input X_j over the states in the network as

$$D(j) \triangleq \sum_{i=1}^{m} \mathrm{MI}(f_i(\mathbf{X}); X_j)$$

where the sum is over all m nodes that represent genes, and hence are functions of the input node's states. We assumed that $\Pr[X_j = 1] = 1/2$ and computed $D(j)$ for each input variable and found that $D(j)$ is large just for some inputs, such as the variables *o2_xt* (36.9 bit), *leu-l_xt* (20.9 bit) and *glc-d_xt* (19.3 bit), (here we adopted the names from the original dataset), but is small for most other variables. From the previous section, it is clear that this cannot be explained solely from the fact that nodes with large values of $D(j)$ tend to have many outgoing edges, while most other nodes do not. This is also what we observed from analyzing the E. coli network, e.g., the state variable *glc-d_xt* has 99 outgoing edges, but *D(glc-d_xt) = 19.3 bit*, whereas variable *o2_xt* has out degree 72, but *D(o2_xt) = 36.9 bit*. Next, let $X_{\tau(1)}, ..., X_{\tau(l)}$ be the inputs with the l largest determinative powers. To see whether knowledge about a small set of those reduces the entropy of the networks states significantly, we computed $H(\mathbf{Y}|X_{\tau(1)}, ..., X_{\tau(l)})$ as a function of l and found that knowledge of merely the states of the most determinative nodes reduces the uncertainty about the network's states significantly. The quantity $H(\mathbf{Y}|X_{\tau(1)}, ..., X_{\tau(l)})$ can be interpreted as a measure of the size of a subset of the overall state space where the system is likely to be found, given knowledge about the states $X_{\tau(1)}, .., X_{\tau(l)}$ [5].

5. REFERENCES

[1] S. Kauffman, Homeostasis and differentiation in random genetic control networks, Nature 224 (5215) (1969) 177–178.

[2] M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard, B. O. Palsson, Integrating high-throughput and computational data elucidates bacterial networks, Nature 429 (6987) (2004) 92–96.

[3] J. Kesseli, P. Rämö, O. Yli-Harja, On spectral techniques in analysis of Boolean networks, Phys. D: Nonlin. Phenom. 206 (1-2) (2005) 49–61.

[4] A. S. Ribeiro, S. A. Kauffman, J. Lloyd-Price, B. Samuelsson, J. E. S. Socolar, Mutual information in random Boolean models of regulatory networks, Phys. Rev. E 77 (1) (2008) 011901.

[5] R. Heckel, S. Schober, M. Bossert, Harmonic analysis of Boolean networks: Determinative Power and Perturbations, arXiv:1109.0807v1, (2010).

[6] N. H. Bshouty, C. Tamon, On the Fourier spectrum of monotone functions, J. ACM 43 (4) (1996) 747–770.

[7] T. M. Cover, J. A. Thomas, Elements of Information Theory, 2nd Edition, Wiley-Interscience, 2006.

[8] G. Xiao, J. Massey, A spectral characterization of correlation-immune combining functions, IEEE Trans. Inf. Theory 34 (3) (1988) 569–571.

[9] L. Raeymaekers, Dynamics of Boolean networks controlled by biologically meaningful functions, J. Theor. Biol. 218 (3) (2002) 331–341.