

RESEARCH

Open Access

Properties of Boolean networks and methods for their tests

Johannes Georg Klotz^{1*}, Ronny Feuer², Oliver Sawodny², Martin Bossert¹, Michael Ederer²
and Steffen Schober¹

Abstract

Transcriptional regulation networks are often modeled as Boolean networks. We discuss certain properties of Boolean functions (BFs), which are considered as important in such networks, namely, membership to the classes of unate or canalizing functions. Of further interest is the average sensitivity (AS) of functions. In this article, we discuss several algorithms to test the properties of interest. To test canalizing properties of functions, we apply spectral techniques, which can also be used to characterize the AS of functions as well as the influences of variables in unate BFs. Further, we provide and review upper and lower bounds on the AS of unate BFs based on the spectral representation. Finally, we apply these methods to a transcriptional regulation network of *Escherichia coli*, which controls central parts of the *E. coli* metabolism. We find that all functions are unate. Also the analysis of the AS of the network reveals an exceptional robustness against transient fluctuations of the binary variables.^a

Keywords: Regulatory Boolean networks, Boolean networks, Linear threshold functions, Unate functions, Canalizing function, Sensitivity, Average sensitivity, Restricted functions, *Escherichia coli*

1 Introduction

Boolean modeling is often used to describe signal transduction and regulatory networks [1-3]. Over the last years random Boolean models received much attention to find some generic properties that characterize regulatory networks. In addition to the study of topological features (e.g., [4]), the choice of Boolean functions in such networks is an important question to consider. Many results indicate the importance of functions with a low average sensitivity. For example, it is well known that a low expected average sensitivity is a prerequisite for non-chaotic behavior of random Boolean networks, e.g., [5,6]. Further, so called *canalizing* functions have been conjectured to be characteristic for biological networks [7]. These functions have a stabilizing effect on the network dynamics [1] and many functions occurring in (non-random) regulative networks are canalizing [7].

In this work we follow a non-random approach to find properties characterizing regulatory networks. Namely, we focus on the properties of Boolean functions in a large

scale Boolean regulatory network model. Our goal is also to provide efficient algorithms to test these properties.

First, we consider the membership of the regulatory functions to certain classes of functions. We first consider unate functions, which are monotone in each of their variables and were shown to be implied by a biochemical model [2].

Next, we present a test using Fourier analysis to test canalizing properties of functions. Canalizing functions are used in signal processing for certain classes of filters [8] and play an important role in random and regulatory Boolean networks, as already mentioned. Interestingly, it has been shown in [9] that a subclass of canalizing functions, namely the nested canalizing functions, is identical to the class of unate-cascade functions, a subclass of the unate functions. The test presented in this work is inspired by [10], where the so-called forcing transform was introduced to test the membership of a function to the class of canalizing functions. Here, we generalize this approach to the Fourier transform, which is a more intuitive and natural approach and furthermore some spectral properties of canalizing functions have already been investigated in [11].

*Correspondence: johannes.klotz@uni-ulm.de

¹Institute of Communications Engineering, Ulm University, Albert-Einstein-Allee 43, 89081 Ulm, Germany

Full list of author information is available at the end of the article

It is well known that the average sensitivity can be directly obtained from the Fourier spectral coefficients. Further, the Fourier transform turns out to be useful to prove bounds on the average sensitivity. We derive an upper bound for unate functions similar to known results for monotone functions and recall a well-known lower bound on the average sensitivity.

Finally, we apply our tests to a large-scale Boolean model of the transcriptional network of *Escherichia coli*. We extended the network model of the transcriptional network of *E. coli* (Covert et al. [3]) by mapping genes to their corresponding fluxes in the flux-balance model presented by [12]. The network has a layered feed-forward structure and shows characteristic topological features, such as a long-tail like out-degree distribution.

Throughout this article we use Fourier analysis to investigate the mentioned properties. In particular we use the concept of restricted functions. Therefore we derived both-way relations between the Fourier coefficients of a Boolean function and its restriction. A very general one-way approach of this relation can be found in [13].

The remainder of this article is organized as follows: In Section 2 we give a short introduction to Boolean functions and networks, discuss some fundamentals of Fourier analysis and investigate the spectra of restricted functions. In Section 3 we discuss certain classes and properties of Boolean functions and show efficient ways to check these properties. We also introduce the average sensitivity and prove an upper bound on it for unate functions. In Section 4 we finally introduce Boolean networks and apply our methods and tests to the regulatory network of *E.coli*. Some final remarks are given in Section 5.

2 BFs

A BF $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ maps n -ary binary *input* tuples to a binary *output*. In general, not all variables of a function f are *relevant*. A variable i is called relevant, if there exists at least one argument $\mathbf{x} \in \{-1, 1\}^n$ such that $f(\mathbf{x}) \neq f(\mathbf{x} \oplus e_i)$, where the argument $\mathbf{x} \oplus e_i$ is obtained from \mathbf{x} by changing its i -th entry. In the following, we denote the number of relevant variables by k .

For the sake of simplicity we assume throughout this article, that $k = n$, i.e., all variables are relevant, but note that the expositions in Section 2.1 are valid in general. The assignment of $+1$ and -1 chosen to represent the binary in and outputs is somewhat arbitrary. One can interpret the value -1 as “ON” or “TRUE” and $+1$ as “OFF” or “FALSE”.

2.1 Fourier analysis

Here we will give a short introduction to the concepts of Fourier analysis so far used in this article. Let us consider $\mathbf{x} = (x_1, x_2, \dots, x_N)$ as an instance of a product distributed

random vector $\mathbf{X} = (X_1, X_2, \dots, X_N)$ with probability density function

$$P_{\mathbf{X}}(\mathbf{x}) = \prod_i P_{X_i}(x_i).$$

Furthermore, let μ_i be the expected value of X_i , i.e., $\mu_i = \mathbb{E}[X_i]$ and let $\sigma_i = \sqrt{1 - \mu_i^2}$ be the standard deviation of X_i . It can easily be seen that

$$P_{X_i}(a_i) = \frac{1 + a_i \cdot \mu_i}{2}. \quad (1)$$

It is well known that any BF f can be expressed by the following sum, called Fourier-expansion [14,15],

$$f(\mathbf{x}) = \sum_{\mathbf{U} \subseteq [n]} \hat{f}(\mathbf{U}) \cdot \Phi_{\mathbf{U}}(\mathbf{x}), \quad (2)$$

where $[n] = \{1, 2, \dots, n\}$ and

$$\Phi_{\mathbf{U}}(\mathbf{x}) = \prod_{i \in \mathbf{U}} \frac{x_i - \mu_i}{\sigma_i}. \quad (3)$$

For $\mathbf{U} = \emptyset$ we define $\Phi_{\emptyset}(\mathbf{x}) = 1$. The *Fourier coefficients* $\hat{f}(\mathbf{U})$ can be recovered by

$$\hat{f}(\mathbf{U}) = \sum_{\mathbf{x}} P_{\mathbf{X}}(\mathbf{x}) \cdot f(\mathbf{x}) \cdot \Phi_{\mathbf{U}}(\mathbf{x}). \quad (4)$$

Further, let $\mathbf{A} \subset \mathbf{U}$ and $\bar{\mathbf{A}} = \mathbf{U} \setminus \mathbf{A}$, then

$$\Phi_{\mathbf{U}}(\mathbf{x}) = \Phi_{\mathbf{A}}(\mathbf{x}) \cdot \Phi_{\bar{\mathbf{A}}}(\mathbf{x}), \quad (5)$$

which directly follows from the definition of $\Phi_{\mathbf{U}}$ (Equation 3).

If the input variables X_i are uniformly distributed, i.e., $\mu_i = 0$ and $\sigma_i = 1$, Equation (3) reduces to

$$\Phi_{\mathbf{U}}(\mathbf{x}) = \prod_{i \in \mathbf{U}} x_i,$$

and consequently, as $P_{\mathbf{X}}(\mathbf{x}) = 2^{-n}$ for all \mathbf{x} , Equation (4), reduces to

$$\hat{f}(\mathbf{U}) = 2^{-n} \sum_{\mathbf{x}} f(\mathbf{x}) \cdot \prod_{i \in \mathbf{U}} x_i.$$

2.2 Restricted functions

A function is called restricted, if some of the input variables are set to constants, i.e., variables $i \in \mathbf{K}$ are set to a constant $x_i = a_i$. Hence, the number of relevant variables is reduced by $|\mathbf{K}|$. First, we consider the case that only one variable is restricted ($\mathbf{K} = \{i\}$). The function obtained in this way is denoted as

$$f|_{x_i=a_i} : \{-1, 1\}^{n-1} \rightarrow \{-1, 1\}. \quad (6)$$

The following lemma gives a relation between the Fourier coefficients of the original function and its restriction.

Proposition 1. Let the function $f(\mathbf{x})$ be a function in n variables. Consider the restricted function obtained by setting $x_i = a_i$, further, let $f|_{x_i=a_i}$ be denoted as \hat{f}_{a_i} then

$$\hat{f}_{a_i}(\mathbf{U}) = \hat{f}(\mathbf{U}) + \Phi_{\{i\}}(a_i) \cdot \hat{f}(\mathbf{U} \cup \{i\}) \quad (7)$$

where $\mathbf{U} \subseteq [n] \setminus \{i\}$ and $\Phi_{\{i\}}(a_i) = \frac{a_i - \mu_i}{\sigma_i}$.

Proof. Using Equation (4) we can rewrite (7) as

$$\begin{aligned} \hat{f}_{a_i}(\mathbf{U}) &= \sum_{\mathbf{x}} P_{\mathbf{X}}(\mathbf{x}) f(\mathbf{x}) \cdot \Phi_{\mathbf{U}}(\mathbf{x}) \\ &+ \Phi_{\{i\}}(a_i) \cdot \sum_{\mathbf{x}} P_{\mathbf{X}}(\mathbf{x}) f(\mathbf{x}) \cdot \Phi_{\mathbf{U} \cup \{i\}}(\mathbf{x}). \end{aligned} \quad (8)$$

By applying (5) and (3) we get

$$\Phi_{\mathbf{U} \cup \{i\}}(\mathbf{x}) = \Phi_{\mathbf{U}}(\mathbf{x}) \cdot \Phi_{\{i\}}(\mathbf{x}) = \Phi_{\mathbf{U}}(\mathbf{x}) \cdot \frac{x_i - \mu_i}{\sigma_i}.$$

Hence, we can combine the two sums in (8) and obtain:

$$\hat{f}_{a_i}(\mathbf{U}) = \sum_{\mathbf{x}} (P_{\mathbf{X}}(\mathbf{x}) f(\mathbf{x}) \cdot \Phi_{\mathbf{U}}(\mathbf{x}) \cdot \Xi_i), \quad (9)$$

where

$$\Xi_i = \left(1 + \frac{x_i - \mu_i}{a_i + \mu_i}\right) = \frac{a_i + x_i}{a_i + \mu_i},$$

due to $\Phi_{\{i\}}(a_i) = \frac{a_i - \mu_i}{\sigma_i} = \frac{\sigma_i}{a_i + \mu_i}$.

Further, with $a_i = \frac{1}{a_i}$ and Equation (1) we get

$$\Xi_i = \begin{cases} \frac{2}{1+a_i \cdot \mu_i} = \frac{1}{P_{X_i}(a_i)} & , \text{ if } x_i = a_i \\ 0 & , \text{ if } x_i = -a_i \end{cases}.$$

Thus, the sum in Equation (9) can be simplified to

$$\hat{f}_{a_i}(\mathbf{U}) = \sum_{\mathbf{x}|x_i=a_i} \frac{P_{\mathbf{X}}(\mathbf{x})}{P_{X_i}(a_i)} \cdot f(\mathbf{x}) \cdot \Phi_{\mathbf{U}}(\mathbf{x})$$

and finally

$$\hat{f}_{a_i}(\mathbf{U}) = \sum_{\mathbf{x}|x_i=a_i} P_{\mathbf{X}|x_i}(a_i) \cdot f(\mathbf{x}) \cdot \Phi_{\mathbf{U}}(\mathbf{x}),$$

which is the definition of the Fourier coefficients from Equation (4) and concludes the proof. \square

A closely related property is given by the following proposition. Please note that this result for uniform distributed input variables can also be retrieved using ([13], Lemma 2.17).

Proposition 2. Let $i \in [n]$ be fixed and denote $f|_{x_i=a_i}$ with \hat{f}_a . For any n -ary BF f ,

$$\begin{aligned} \hat{f}(\mathbf{U}) &= \frac{1 + \mu_i}{2} \left((\Phi_{\{i\}}(+1))^{|U \cap \{i\}|} \hat{f}_{+1}(\mathbf{U} \setminus \{i\}) \right) \\ &+ \frac{1 - \mu_i}{2} \left((\Phi_{\{i\}}(-1))^{|U \cap \{i\}|} \hat{f}_{-1}(\mathbf{U} \setminus \{i\}) \right). \end{aligned}$$

Proof. Starting from the definition we obtain

$$\begin{aligned} \hat{f}(\mathbf{U}) &= \mathbb{E} [f(\mathbf{X}) \Phi_{\mathbf{U}}(\mathbf{X})] \\ &= P_{X_i}(+1) \mathbb{E} [f(\mathbf{X}) \Phi_{\mathbf{U}}(\mathbf{X}) | X_i = +1] \\ &+ P_{X_i}(-1) \mathbb{E} [f(\mathbf{X}) \Phi_{\mathbf{U}}(\mathbf{X}) | X_i = -1] \\ &= P_{X_i}(+1) \Phi_{\{i\}}(+1) \mathbb{E} [f_{+1}(\mathbf{X}) \Phi_{\mathbf{U} \setminus \{i\}}(\mathbf{X})] \\ &+ P_{X_i}(-1) \Phi_{\{i\}}(-1) \mathbb{E} [f_{-1}(\mathbf{X}) \Phi_{\mathbf{U} \setminus \{i\}}(\mathbf{X})]. \end{aligned}$$

Note that for $a = +1$ or $a = -1$

$$\mathbb{E} [f_a(\mathbf{X}) \Phi_{\mathbf{U} \setminus \{i\}}(\mathbf{X})] = \hat{f}_a(\mathbf{U} \setminus \{i\})$$

by definition, hence, the proposition follows from Equation (1). \square

For the general case, that a BF is restricted to more than one input, the following Corollary to Proposition 1 applies:

Corollary 1. Let $f(\mathbf{x})$ be a BF and $\hat{f}(\mathbf{U})$ its Fourier coefficients. Furthermore, let \mathbf{K} be a set containing the indices i of the input variables x_i , which are fixed to certain values a_i . The Fourier coefficients of the restricted function are then given as

$$\hat{f}|_{\mathbf{K}}(\mathbf{U}) = \sum_{\mathbf{T} \subseteq \mathbf{K}} (\Phi_{\mathbf{T}}(\mathbf{a}) \cdot \hat{f}(\mathbf{U} \cup \mathbf{T})),$$

where \mathbf{U} contains the indices for the Fourier coefficients of the restricted functions, i.e., $\mathbf{U} \subseteq [n] \setminus \mathbf{K}$ and \mathbf{a} is a vector containing all $a_i, i \in \mathbf{K}$.

3 Classes and properties of functions

In this section, we will present and discuss some classes of BFs, namely unate and canalizing functions. Further, we will discuss properties of functions characterizing their robustness, like for example the AS.

3.1 Unate functions

A BF is unate if it is monotone (either increasing or decreasing) in each of its variables, a precise definition will be given below. The class of unate functions is a simple extension of the class of monotone functions defined as follows

Definition 1. A BF $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ is called monotone, if for each $i \in \{1, \dots, n\}$ it holds that $f(x_1, \dots, x_i = -1, \dots, x_n) \leq f(x_1, \dots, x_i = 1, \dots, x_n)$.

Now unate functions can be defined as follows.

Definition 2. A BF f is unate, if there exists a vector $\mathbf{a} \in \{-1, 1\}^n$ such that the function $f(a_1 \cdot x_1, \dots, a_n \cdot x_n)$ is monotone.

The class of unate functions is closed with respect to restriction, since every restriction of a locally monotone function yields again in a locally monotone function.

To test whether a function is unate or not it is sufficient to use the definition, however, a necessary condition for a function to be unate is given by the following proposition:

Proposition 3 (for example [16]). *If f is a unate function, then for each relevant variable i*

$$\hat{f}(\{i\}) \neq 0.$$

3.2 Canalizing functions

A BF is called canalizing, if there exists a canalizing variable x_i and a Boolean value $a_i \in \{-1, 1\}$ such that the function

$$f|_{x_i=a_i}(\mathbf{x}) = b_i, \tag{10}$$

for all $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$, where $b_i \in \{0, 1\}$ is a constant. If the restricted function, which is obtained by setting $x_i = 1 - a_i$, is again canalizing and so on, the function is called nested canalizing.

The following propositions give a relation between the Fourier coefficients and the canalizing property.

Proposition 4. *A BF f is canalizing in variable i , if for any constants $a_i, b_i \in \{-1, 1\}$ the Fourier coefficients $\hat{f}(\mathbf{U})$ fulfill the following condition.*

$$\hat{f}(\emptyset) + \Phi_{\{i\}}(a_i) \cdot \hat{f}(\{i\}) = b_i, \tag{11}$$

where μ_i is the expected value of x_i and σ_i the corresponding standard derivation.

Proof. Obviously, if a function is canalizing, $\mathbb{E}[f|_{x_i=a_i}(\mathbf{x})] = b_i$ holds. Since the expected value of a BF can be expressed as $\mathbb{E}[f(x)] = \hat{f}(\emptyset)$ we obtain

$$\mathbb{E}[f|_{x_i=a_i}(\mathbf{x})] = \hat{f}|_{x_i=a_i}(\emptyset).$$

Using Proposition 1, we get

$$\mathbb{E}[f|_{x_i=a_i}(\mathbf{x})] = \hat{f}(\emptyset) + \Phi_{\{i\}}(a_i) \cdot \hat{f}(\{i\}),$$

and the proposition follows from Equation (11). □

A similar result namely the calculation of the Fourier coefficients of a canalizing BF from the coefficients of the restricted functions $\hat{f}|_{x_i=a_i}(\mathbf{U})$ is addressed in [11]. These results can also be achieved using Proposition 2.

Proposition 4 can easily be extended for nested canalizing functions:

Proposition 5. *Assume $f(x)$ is canalizing for variables $x_i = -a_i, i \in \mathbf{K}$, then $f(x)$ is canalizing for $x_j = a_j, j \notin \mathbf{K}$, i.e., $\mathbb{E}[f_{\mathbf{K} \cup \{j\}}(x)] = b_j$, if*

$$\sum_{\mathbf{T} \subseteq \mathbf{K} \cup \{j\}} \left(\prod_{i \in \mathbf{T}} (\Phi_{\{i\}}(a_i)) \cdot \hat{f}(\mathbf{T}) \right) = b_j.$$

Proof. The proof follows from Corollary 1 and Proposition 4. □

From Proposition 4 it is clear that the canalizing property can be tested by considering all Fourier coefficients of order one. Using the Fast Walsh Transform [17] this test is as fast as the one presented in [10], however, once we have retrieved the spectra of a function, we can easily compute other properties, such as the AS (see next section).

3.3 AS of functions

The AS [18] gives the influence of random disturbance at the input on the output of a BF. This can be interpreted as an indicator for the robustness of this BF and finally for the whole Boolean network.

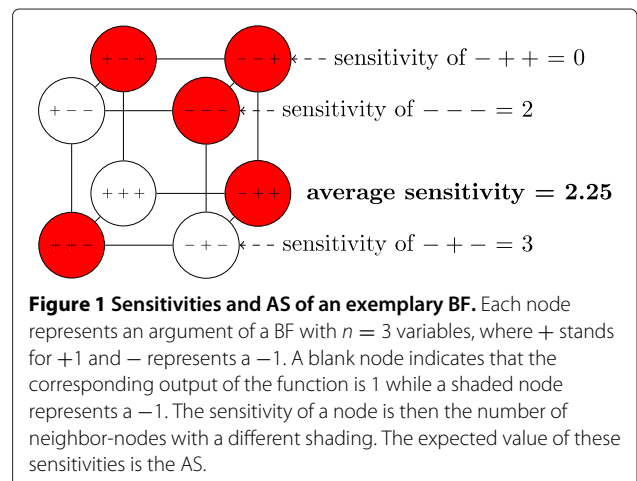
To define the as we first have to look at the sensitivity $s_{\mathbf{x}}(f)$ of an input argument $\mathbf{x} \in \{0, 1\}^n$. It is defined as the number of single bit-flips in \mathbf{x} so that the output of the function will change, i.e., $s_{\mathbf{x}}(f)$ is number of variables i for which $f(\mathbf{x}) \neq f(\mathbf{x} \oplus e_i)$. The AS $as(f)$ is the expected value over all arguments \mathbf{x} :

$$as(f) = \mathbb{E}_{\mathbf{x}}[s_{\mathbf{x}}(f)]. \tag{12}$$

It is worth noting that the as depends on the distribution of the input vector. For example, a function having a low AS for the uniform distribution may have a large AS for other distributions. In general, the AS can be as large as the number of relevant variables k , i.e.,

$$0 \leq as(f) \leq k.$$

Figure 1 explains the concepts defined above at an example.



Alternatively, the AS can be defined using the notion of *influence*. The influence $I_i(f)$ of a single input variable i on the functions f is defined as

$$I_i(f) = \text{Prob}[f(\mathbf{X}) \neq f(\mathbf{X} \oplus e_i)]. \quad (13)$$

The AS can then be defined as the sum of all influences [19]

$$as(f) = \mathbb{E}_x[s_x(f)] = \sum_i I_i(f). \quad (14)$$

The influence $I_i(f)$ for a unate function f is directly related to the corresponding Fourier coefficient:

$$I_i(f) = \frac{|\hat{f}(\{i\})|}{\sigma_i}, \quad (15)$$

as it was shown for monotone functions in ([16], Lemma 4.5) and can easily be extended to unate functions. Note that Equation (15) directly gives a proof for Proposition 3. Hence, for unate functions we can write

$$as(f) = \sum_{i=1}^n \frac{|\hat{f}(\{i\})|}{\sigma_i} = \sqrt{\left(\sum_{i=1}^n \frac{|\hat{f}(\{i\})|}{\sigma_i}\right)^2}, \quad (16)$$

and from the Cauchy-Schwarz inequality it follows that

$$\begin{aligned} as(f) &\leq \sqrt{\sum_{i=1}^n \left(|\hat{f}(\{i\})|\right)^2} \sqrt{\sum_{i=1}^n \left(\frac{1}{\sigma_i}\right)^2} \\ &\leq \sqrt{(1 - \hat{f}(\emptyset)^2)} \sqrt{\sum_{i=1}^n \left(\frac{1}{\sigma_i}\right)^2}. \end{aligned} \quad (17)$$

Together with a lower bound as presented in [19,20] and since $1 - \hat{f}(\emptyset)^2 = 1 - \mathbb{E}[f]^2 = \text{Var}(f)$ we obtain the following proposition.

Proposition 6. *Let f be an unate BF with in-degree n , further let σ_i be the standard derivation of the i -th input, then the AS of f ($as(f)$) is bounded by*

$$\text{Var}(f) \leq as(f) \leq \sqrt{\text{Var}(f)} \sqrt{\sum_{i=1}^n \left(\frac{1}{\sigma_i}\right)^2}, \quad (18)$$

where $\text{Var}(f)$ denotes the variance of f .

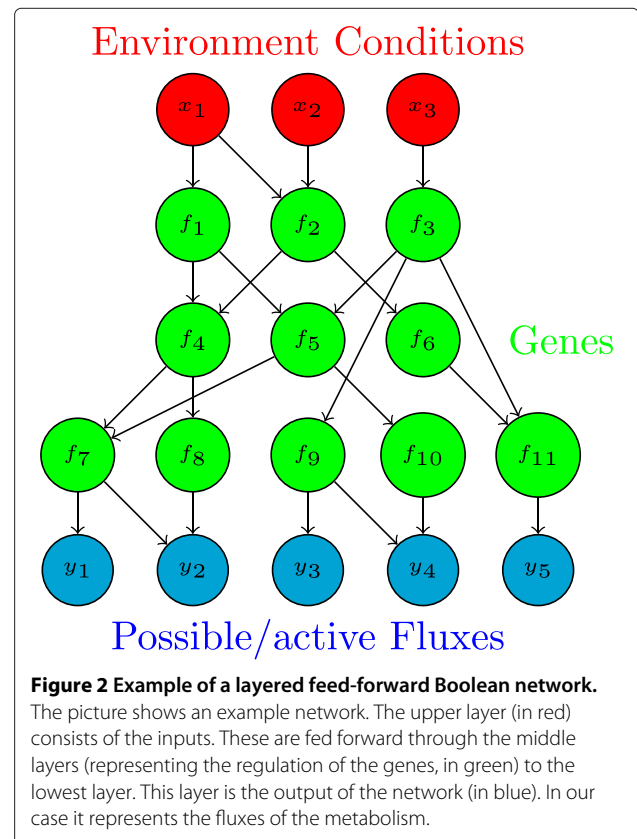
It can be shown that some functions get close to the upper bound. Assuming uniform distribution the upper bound in Equation (18) is smaller than \sqrt{n} . But it is well known that the AS of the majority function behaves like $O(\sqrt{n})$ (see for example [21]).

4 Application to a regulatory network of *E. coli*

In the previous sections, we only considered BFs. Now we will focus on BNs. A synchronous BN of N nodes can be

described by a graph $G = G(V, E)$ with nodes $V \subseteq [N]$, $|V| = N$, and edges $E \subseteq V \times V$, and a set of ordered BFs $F = (f_1, f_2, \dots, f_N)$, where we also allow a *dummy* function (see below). Each f_i has $n_i = k_i = \text{in-deg}(i)$ relevant variables where $\text{in-deg}(i)$ is the in-degree of node i , i.e., the number of edges (j, i) with $j \in V$. In this case a node j is called a *controlling* node of i . If a node i has in-degree zero, the dummy function is attached and we call it an in-node. Consequently, the number of edges emerging from i is called the out-degree of node i . Usually to each node a binary state variable is assigned, i.e., for node i we assign $x_i(t) \in \{-1, +1\}$. For in-nodes the state can be set by some external process at some time t_0 . The state of all other nodes at time t depend on its BF and the states of all controlling nodes at time instant $t - 1$.

In this article, we are only considering feed-forward networks, i.e., networks without *feedback loops*. In such feed-forward BNs, the set of nodes is partitioned in layers L_1, L_2, \dots, L_l . If a node i is an element of layer L_h all controlling nodes are element of layers L_m with $m < h$. The first (highest) layer L_1 consists of the input nodes (in-nodes), while the lowest layer L_l consists of the output nodes (out-nodes). In Figure 2 a sample network is depicted.



4.1 Structural properties

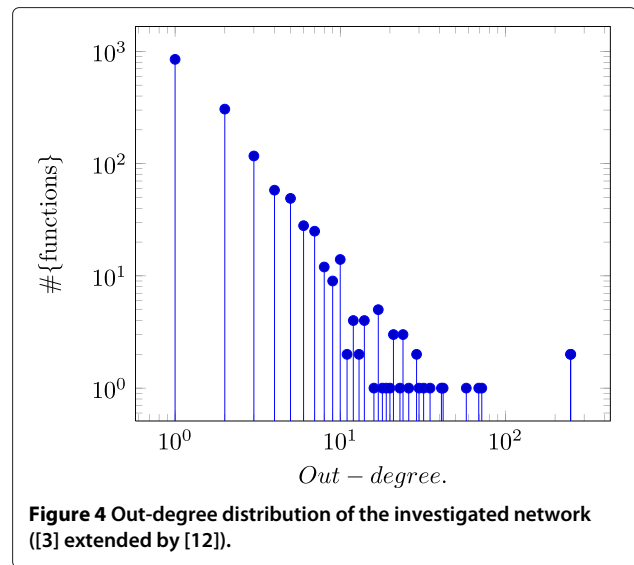
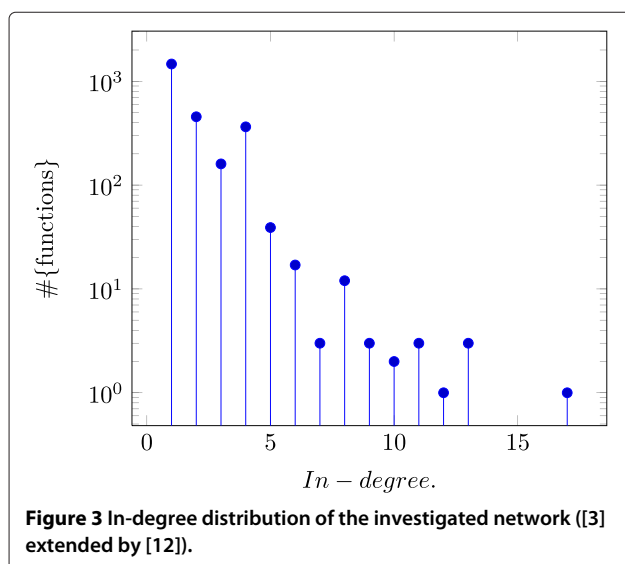
We applied the tests described in the previous sections to the regulatory network of *E. coli* [3]. The model provides Boolean formulas that describe how environmental conditions act on gene expression via a transcriptional regulatory network. We extended this network by the mapping of the genes to their corresponding fluxes in the flux-balance model [12]. The network as described in the literature contains functions with irrelevant variables, respectively, redundant edges, which are removed. A list of the affected nodes and the removed edges can be found in the Additional file 1.

The resulting network has a total of $N = 3915$ nodes and $|E| = 4874$ edges, where 1,386 of these nodes are in layer L_1 , i.e., are inputs, hence, 2,529 nodes have a non-dummy function attached. The in-degree and out-degree distributions can be found in Figures 3 and 4. The average in-degree is 1.92724. The out-degree distribution shows a typical long tail behavior [4].

We found that all functions attached to the nodes are unate. Furthermore 2499 functions (98.8%) are canalizing. An overview of the functions, which are not canalizing, can be found in the Additional file 1.

4.2 Robustness

To evaluate the robustness of the network we assume in general that the state of nodes can be described by binary random variables. In a first step we assumed that each random variable of each nodes is uniformly distributed. This implies that we consider each node independently, i.e., the topology of the network is ignored. We calculated the AS for all functions in the network. In Figure 5, the resulting AS is plotted versus the *bias*, which is the probability that the output of the function equals one (a similar analysis

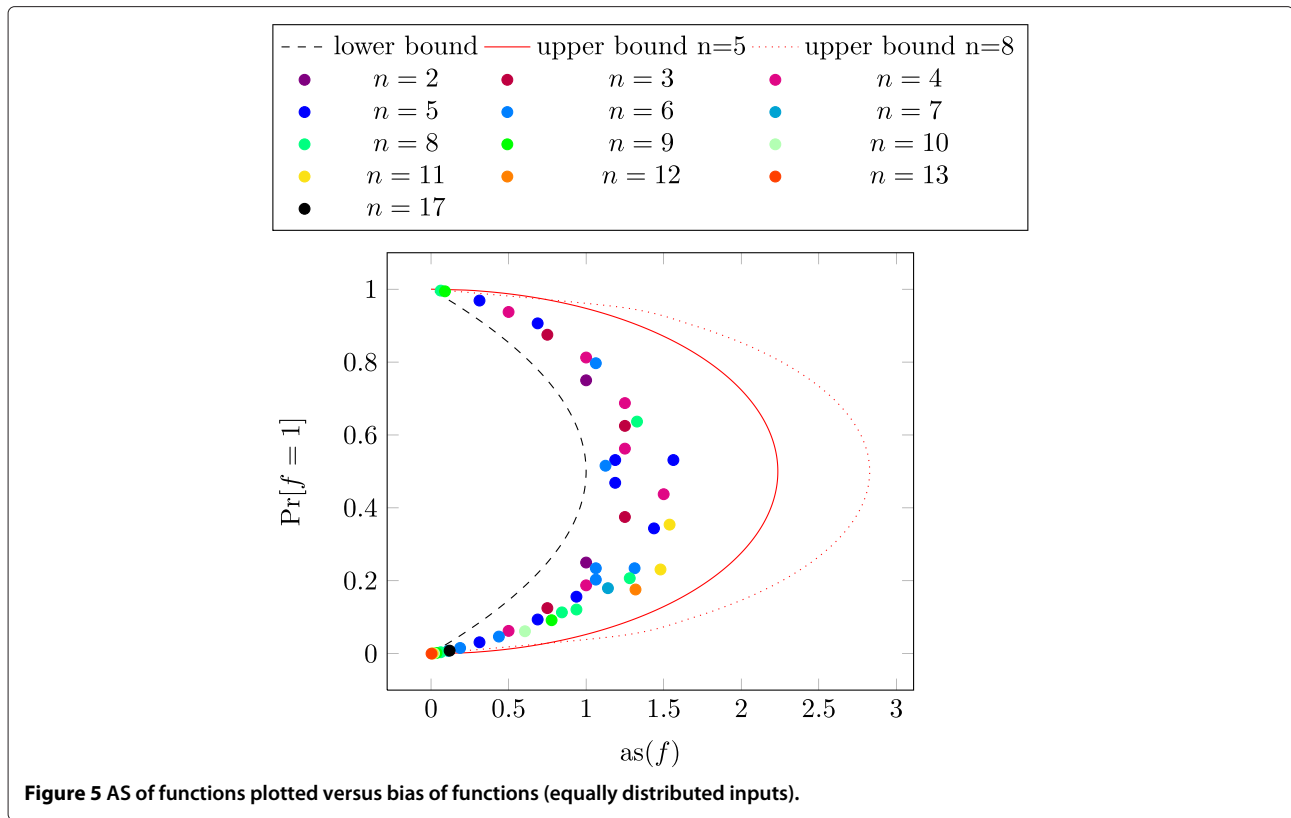


appears in [22]). Each color represents a BF with a certain in-degree n . We also included the lower bound and two exemplary upper bounds for $n = 5$ and $n = 8$ (Equation (18)). For increasing n the upper bounds will grow, i.e., the bound will move further to the right.

Obviously, functions with a strong bias, i.e., with a high probability to be either -1 or 1 , have a low AS. Further it can be seen that the average sensitivities of all functions are very close to the lower bound. The mean value of the AS is 0.918874. Hence, it can be stated that the AS of this network is rather low. Similar results can be obtained considering the network without the extension as originally defined by Covert et al. [3] and Samal and Jain [23].

In a second step we want to take the topology of the network into account. Therefore, we now assume that only the in-nodes of the network are equally distributed. However, the output of these functions will most certainly not be uniform, i.e., the functions have a bias unequal zero. Since the outputs of these functions serve as inputs of the functions of the next layer, we assume that their input distributions follow the output distribution of the first-layer functions. The output distributions of the second-layer functions serve then as input distributions of the third layer and so on. Obviously this has an impact on the *as* of the functions.

The results are shown in Figure 6. We did not include any upper bounds in this figure since these now depend on each input distribution (see Proposition 6). It can be seen that the AS is still very close to the lower bound. However, a few functions have a rather large AS, e.g., it can be seen in Figure 6 that two types of functions with in-degree $K = 2$ are very close to their upper bound (which is in this case at $as(f) = 2$). These functions have an argument with a sensitivity of 2. Due to the input distribution



of these functions this argument has a very large probability ($> 98\%$) which leads to a very high AS close to 2. Such high AS are normally observed for XOR and related functions. The average value of the AS is 0.908445, hence the AS of the network further decreases when applying product distributions at the inputs of the functions.

4.3 Comparison with random ensembles

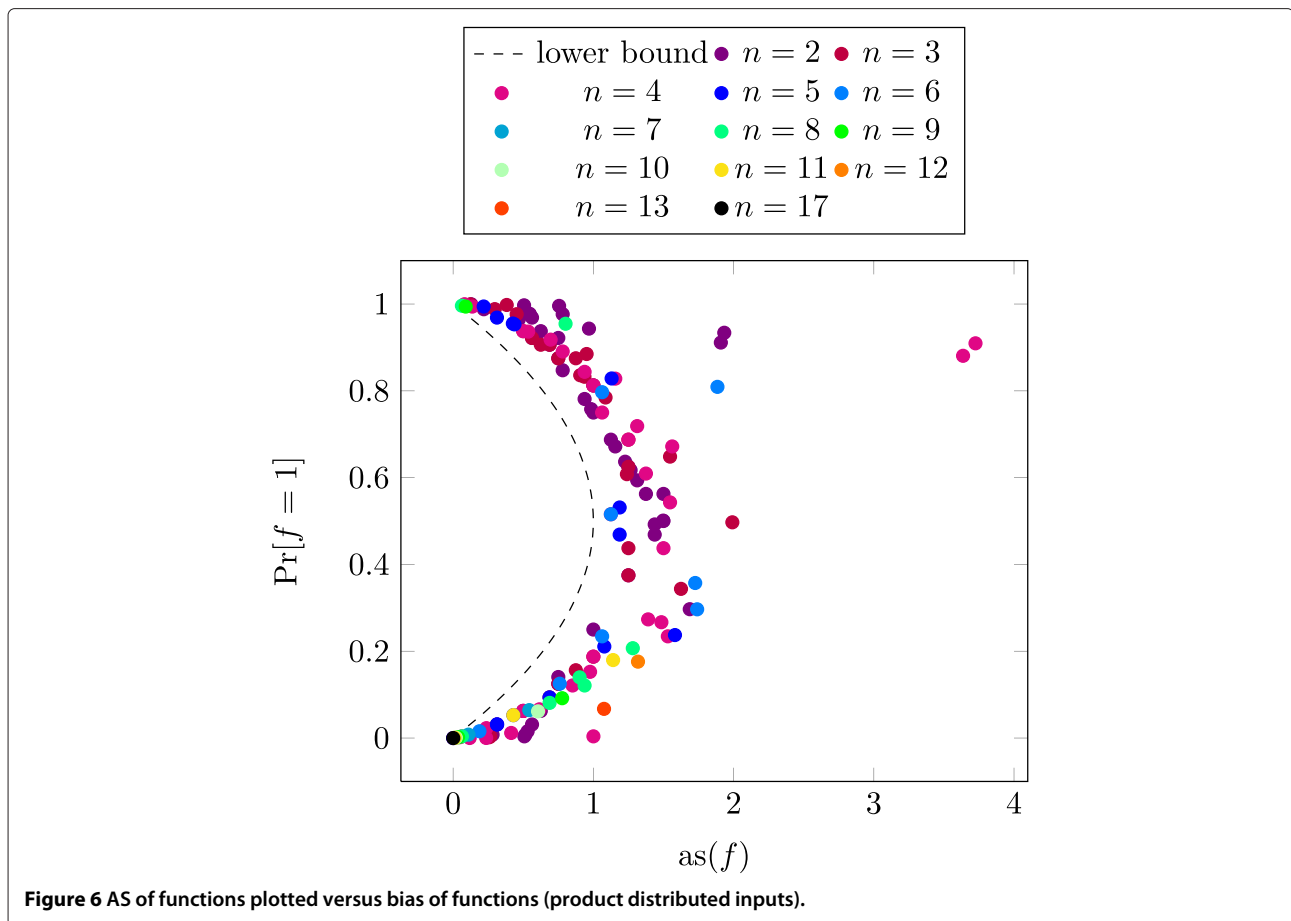
The network appears to be more *robust* against transient errors as for example certain randomly constructed networks. The in-degree distributions of all controlled nodes (in-degree larger zero) is shown in Figure 3. For all nodes with in-degree k we choose a random function out of the set of functions with k relevant variables. For $k = 1$ this results in $\mathbb{E}[as(f)] = 1$, for $k > 1$ we can at least state that $\mathbb{E}[as(f)] > \frac{k}{2}$, as it is well known that if we choose randomly from functions, we expect an AS of $\frac{k}{2}$. Taking the in-degree distribution into account this implies that the expectation of the AS of all BFs chosen in this way is larger than 1.25.

It is well known that random function ensembles with lower expected AS can be constructed, if functions with a higher bias have higher probability to be chosen [24]. To test if the observed robustness can be explained due to the bias of the functions, we proceeded as follows. Again, a random function is chosen for each node with in-degree

k . We determine the frequency distribution for the bias $b = P[f = 1]$ for all functions of the original network model with a certain in-degree k . The random network is generated by replacing the original functions of the network with functions drawn from an ensemble of functions with the same distribution. For example, if $k = 2$, roughly 32% of all functions have $b = 0.25$, while all others have $b = 0.75$. Hence, with probability 32% we choose a function with $b = 0.25$, and $b = 0.75$ otherwise. The data can be found in the Additional file 1. As shown in [25,26], the expectation of the AS is then given by

$$\mathbb{E}(as(f)) = 2kb(1 - b) \frac{2^k}{2^k - 1}. \quad (19)$$

The results obtained are shown in Table 1 sorted according to the in-degree k . For $k = 1$ and $k = 2$ the observed mean of $as(f)$ and the expectation of the random function coincides as only identity functions, respectively, AND or OR functions are chosen. For larger k , the observed mean is always smaller as the expectation of the random function. For some values of k , for example $k = 9$, both values are close to each other. This is due to the fact that the corresponding functions are highly biased, which means that there are three existing functions with the values for b being 0.00195312, 0.0917969 and 0.994141. In contrast, for $k = 11$ the mean of the observed values and the



expectation are far from each other. Indeed out of three functions, there is one function with $b = 0.354004$ for which, according to Equation (19), the expectation of the AS is 5.03353147.

It should be noted that in random BNs the expectation of the AS is an *order parameter* [5,6]. That is, if the expectation is less or equal to one many random networks show the so-called ordered behavior. Namely, single transient errors introduced in network nodes (by flipping their state) do not spread through the network with high probability. This ordered behavior is in sharp contrast to the so-called disordered behavior of random networks which is characterized by an expectation of the AS larger one. Indeed, it has been conjectured that biological relevant networks should be ordered (or critical) but not disordered [27]. A further investigation on how canalizing and nested canalizing functions influence the average sensitivity can be found in [7,11].

4.4 Impact of mutations on the metabolism

When investigating a regulatory network, the impact of the network on the metabolism is of major interest. Hence, only the stability of nodes in the bottom layer, i.e.,

Table 1 Fraction of functions with in-degree k , the mean of the AS of all functions with in-degree k , and the expectation of an accordingly chosen random function with same in-degree and same bias distribution (see text and Equation 19)

k	Fraction of functions	$av(f)$	$\mathbb{E}(av(f))$
1	0.579905	1.000000	1.000000
2	0.179984	1.000000	1.000000
3	0.063291	0.887500	0.985714
4	0.143987	0.572115	0.623077
5	0.015427	0.491987	0.659895
6	0.006725	0.933824	1.737920
7	0.001187	0.796872	1.423026
8	0.004747	0.760416	1.641421
9	0.001187	0.300781	0.547935
10	0.000791	0.312500	0.587713
11	0.001187	1.009441	2.984577
12	0.000396	1.318360	3.481815
13	0.001187	0.003174	0.003174

the output of the network, is relevant. In regulatory networks, mutations are a source for errors. We consider two possible types of mutations. First we assume that a part of promoter region of a gene is mutated or deleted. In terms of our network this means that an edge is removed and the corresponding input is set to false (+1). The gene may still be transcribed, hence, the node itself remains functional. The second type of mutation is the deletion of a gene or a mutation which leads to a dysfunctional gene. In this case, the node is constantly set to false. In both cases, the value of one node may change (error). This error is now fed through the outgoing edges of this node to other nodes. However, due to the low sensitivity of all functions in the network, the error has no impact on many nodes and, therefore, will in most cases not reach the bottom layer, which is, as mentioned above, the only part of the network, whose stability is crucial. From that point of view it can be stated that these permanent errors behave similar to the transient errors described above and that networks with a low mean AS are robust against such errors.

5 Summary

It is an important problem to characterize BFs that appear in Boolean models of regulatory networks. This will help to understand the constraints underlying such networks, but can, for example, also help to improve network inference algorithms (see for examples [28,29] for algorithms that utilize the membership to the class of unate functions). In this study, we focused on several properties that have been shown to be of interest in the context of Boolean regulatory networks. Namely, we discussed different classes of BFs such as unate and canalizing functions. Further, sensitivity measures of BFs, like the influence of variables, or the AS are considered. We devised simple algorithms to test these properties. To test canalizing properties of BFs we applied the Fourier representation of BFs where functions are represented as multivariate, multilinear real polynomials. To this end, we introduced two spectral relationships between the so-called *restricted* BFs and their unrestricted counterpart. The Fourier representation is further useful as many interesting properties such as the influence of unate functions or the AS of BFs can easily be characterized in the spectral domain. For example, we show how to obtain theoretical upper bounds on the AS for unate functions using spectral techniques.

As an application of our results, we analyzed an extended [30] regulatory Boolean network model of the central metabolism of *E. coli*. It turned out that most functions are within the classes of unate functions. Further, the AS of most functions is close to a theoretical lower bound and far from the new upper bound. Especially, functions with large in-degree have low AS even if their so-called

bias is close to 0.5 (see Figure 5). We compared our findings to random BNs with similar parameters and find that the investigated networks have an even lower AS. From that we conclude that the whole network is stable, and robust to small changes, e.g., mutations.

6 Endnote

^aPreliminary results of this study have been presented at the 8th International Workshop on Computational Systems Biology (WCSB 2011) and the 3rd International Conference on Bioinformatics and Computational Biology (BICoB 2011).

Additional file

Additional file 1: This .xls file contains 3 sheets with listings of:

- all nodes having pseudoedges
- all canalizing functions
- average sensitivities of all functions

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors would like to thank Georg Sprenger and Katrin Gottlieb from the Institute for Microbiology at the University of Stuttgart for fruitful collaboration and discussions. Further we thank Reinhard Heckel for creating large parts of the software. This study was supported by the German Research Foundation "Deutsche Forschungsgemeinschaft" (DFG) under Grants Bo 867/25-1 and Sa 847/11-1.

Author details

¹Institute of Communications Engineering, Ulm University, Albert-Einstein-Allee 43, 89081 Ulm, Germany. ²Institute for System Dynamics, University of Stuttgart, 70569 Stuttgart, Germany.

Received: 6 January 2012 Accepted: 26 November 2012

Published: 11 January 2013

References

1. S Kauffman, C Peterson, B Samuelsson, C Troein, Genetic networks with canalizing Boolean rules are always stable. *Proc. Natl Acad. Sci. USA*. **101**(49), 17102–17107 (2004)
2. J Grefenstette, S Kim, S Kauffman, An analysis of the class of gene regulatory functions implied by a biochemical model. *Biosystems*. **84**(2), 81–90 (2006). <http://www.sciencedirect.com/science/article/B6T2K-4HWP4R-3/2/61a3092f98470e99a2c33786416697d0>
3. MW Covert, EM Knight, JL Reed, MJ Herrgard, BO Palsson, Integrating high-throughput and computational data elucidates bacterial networks. *Nature*. **429**(6987), 92–96 (2004). <http://dx.doi.org/10.1038/nature02456>
4. M Aldana, Boolean dynamics of networks with scale-free topology. *Physica D*. **185**, 45–66 (2003)
5. I Shmulevich, SA Kauffman, Activities and sensitivities in Boolean network models. *Phys. Rev. Lett.* **93**(4), 048701 (2004)
6. JF Lynch, *Dynamics of random Boolean networks*. (byK Mahdavi, R Culshaw, J Boucher, eds.) (World Scientific Publishing Co, Singapore, 2007), pp. 15–38
7. SE Harris, BK Sawhill, A Wuensche, S Kauffman, A model of transcriptional regulatory networks based on biases in the observed regulation rules. *Complexity*. **7**(4), 23–40 (2002)
8. M Gabbouj, PT Yu, EJ Coyle, Convergence behavior and root signal sets of stack filters. *Circuits Syst. Signal Process.* **11**, 171–193 (1992)
9. AS Jarrach, B Raposa, R Laubenbacher, Nested Canalizing, Unate Cascade, and Polynomial Functions. *Physica D*. **233**(2), 167–174 (2007)

10. I Shmulevich, H Lahdesmaki, K Egiazarian, Spectral methods for testing membership in certain post classes and the class of forcing functions. *Signal Process. Lett. IEEE.* **11**(2), 289–292 (2004)
11. J Kesseli, P Rämö, O Yli-Harja, in *Proceedings of the 2005 International TICSP Workshop on Spectral Methods and Multirate Signal Processing (SMMSP 2005)*. Analyzing dynamics of Boolean networks with canalizing functions using spectral methods, (Riga, Latvia, 20–22 June 2005), pp. 151–158
12. AM Feist, CS Henry, JL Reed, M Krummenacker, AR Joyce, PD Karp, LJ Broadbelt, VBO Hatzimanikatis, V Palsson, A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG 1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **3**, 121 (2007). 10.1038/msb4100155
13. A Bernasconi, *Mathematical techniques for the analysis of Boolean functions. PhD thesis.* (University of Pisa, Italy, 1998)
14. RR Bahadur, in *Studies on Item Analysis and Prediction, no. 6 in Stanford Mathematical Studies in the Social Sciences*, ed. by H Solomon. A representation of the joint distribution of responses to n dichotomous items (Stanford University Press, Stanford, CA, 1961), pp. 158–176
15. ML Furst, JC Jackson, SW Smith, in *Proceedings of the Fourth Annual Workshop on Computational Learning Theory*. Improved learning of AC0 functions (Morgan Kaufmann Publishers Inc., Santa Cruz, 1991), pp. 317–325
16. NH Bshouty, C Tamon, On the Fourier spectrum of monotone functions. *J. ACM.* **43**(4), 747–770 (1996)
17. J Shanks, Computation of the fast Walsh-Fourier transform. *IEEE Trans. Comput.* **C-18**(5), 457–459 (1969)
18. I Benjamini, G Kalai, O Schramm, Noise sensitivity of Boolean functions and applications to percolation. *Publications mathématiques de l’IHES.* **90**, 5–43 (1999)
19. J Kahn, G Kalai, N Linial, in *Proceedings of the 29th Annual Symposium on Foundations of Computer Science*. The influence of variables on Boolean functions (White Plains, (New York, USA, 24–26 Oct 1988), pp. 68–80
20. E Friedgut, Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica.* **18**, 27–35 (1998). 10.1007/PL00009809
21. R O’Donnell, in *Proceedings of the 40th annual ACM symposium on Theory of computing*. Some topics in analysis of boolean functions (ACM, Victoria, 2008), pp. 569–578. <http://portal.acm.org/citation.cfm?id=1374458>
22. R Heckel, S Schober, M Bossert, Harmonic analysis of Boolean networks: determinative power and perturbations (2011). arXiv:1109.0807
23. A Samal, S Jain, The regulatory network of *E. coli* metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response. *BMC Syst. Biol.* **2**, 21 (2008). doi:10.1186/1752-0509-2-21
24. B Derrida, Y Pomeau, Random networks of automata—a simple annealed approximation. *Europhys. Lett.* **2**, 45–49 (1986)
25. S Schober, *Analysis and identification of Boolean networks using harmonic analysis.* (Dissertation, Ulm University, Ulm, Germany, 2011)
26. S Schober, M Bossert, Analysis of random Boolean networks using the average sensitivity (2007). arXiv:nl.cg/0704.0197
27. SA Kauffman, Metabolic stability and epigenesis in randomly constructed nets. *J. Theor. Biol.* **22**, 437–467 (1969)
28. S Schober, D Kracht, M Heckel, R Bossert, Detecting controlling nodes of Boolean regulatory networks. *EURASIP J. Bioinf. Syst. Biol.* **27**(11), 1529–1536 (2011). <http://www.ncbi.nlm.nih.gov/pubmed/21989141>
29. M Maucher, B Kracher, M Kühl, HA Kestler, Inferring Boolean network structure via correlation. *Bioinformatics* (2011). <http://bioinformatics.oxfordjournals.org/content/early/2011/04/05/bioinformatics.btr166.abstract>
30. R Feuer, K Gottlieb, JG Klotz, S Schober, M Bossert, O Sawodny, G Sprenger, M Ederer, in *Proceedings of the 8th International Workshop on Computational Systems Biology (WCSB)*. Model-based analysis of adaptive evolution, (Zuerich, Switzerland, 2011), pp. 108–111

doi:10.1186/1687-4153-2013-1

Cite this article as: Klotz et al.: Properties of Boolean networks and methods for their tests. *EURASIP Journal on Bioinformatics and Systems Biology* 2013 **2013**:1.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
