Harmonic Analysis of Boolean Networks: Determinative Power and Perturbations

Reinhard Heckel,[*]   Steffen Schober[†] and Martin Bossert[†]

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 of the nodes 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 set of inputs over node i. We compare the determinative power of a set of inputs to the sensitivity to perturbations to these inputs, and find that, maybe surprisingly, an input that has 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 results, we analyze the large-scale regulatory network of *Escherichia coli*. We identify the most determinative nodes and show that a small subset of those reduces the overall uncertainty of the network state significantly. Furthermore, the network is found to be tolerant to perturbations of its inputs.

1   Introduction

A Boolean network (BN) is a discrete dynamical system, which is, for example, used to study and model a variety of biochemical networks such as genetic regulatory networks. BNs have been introduced in the late 1960s by Kauffman [1, 2] who proposed to study random BNs as models of gene regulatory networks. Kauffman investigated their dynamical behavior and a phenomena called self-organization. Aside from its original purpose, BNs were also used to model (small-scale) genetic regulatory networks; for example, in [3, 4, 5], it was demonstrated that BNs are capable of reproducing the underlying biological processes (i.e., the cell cycle) well. BNs are also used to model large-scale networks, such as the *Escherichia coli* regulatory network [6] which is analyzed in section 6. This network is, in contrast to Kauffman's automata and the regulatory networks considered in [3, 4, 5], not an autonomous system, since the gene's states are determined by external factors.

In the literature addressing the analysis of BNs, it is common to consider measures that quantify the effect of perturbations. Whether a random BN operates in the so called ordered or disordered regime is determined by whether a single perturbation, i.e., flipping the state of a node, is expected

---

[*]Department of Information Technology and Electrical Engineering, ETH Zurich
[†]Institute of Telecommunications and Applied Information Theory, University of Ulm

to spread or die out eventually. Kauffman [2] argues that biological networks must operate at the border of the ordered and disordered regime; hence, they must be tolerant to perturbations to some extent.

In contrast to measures of perturbations, determinative power in BNs has not received much attention, even though there are several settings where such a notion is of interest. For example, given a feed forward network where the states of the nodes are controlled by the states of nodes 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 section 6).

In this paper, we view the state of each node in the network as an independent random variable. This modeling assumption applies for networks with a tree-like topology, e.g., a feed forward network, and is often applied when studying the effect of perturbations. For this setting, determinative power of nodes and perturbation-related measures are properties of single functions; hence, the analysis of the BN reduces to the analysis of single functions. Our main tool for the theoretical results is Fourier analysis of Boolean functions. Fourier analytic techniques were first applied to BNs by Kesseli et al. [7, 8]. In [7, 8], results related to Derrida plots and convergence of trajectories in random BNs were derived. Ribeiro et al. [9] considered the pairwise mutual information in time series of random BNs, under a different setup that we use. Specifically, in [9], the functions are random; whereas here, the functions are deterministic, but the argument is random. Finally, note that part of this paper was presented at the 2012 International Workshop on Computational Systems Biology [10].

1.1 Contributions

Mutual information between a set of inputs to a node and the state of this node is a measure of the determinative power of this set of inputs, as mutual information quantifies mutual dependence of random variables. In order to understand the determinative power and mutual dependencies in Boolean networks, we systematically study the mutual information of sets of inputs and the state of a node. We relate mutual information to a measure of perturbations and prove that (maybe surprisingly) a set of inputs that is highly sensitive to perturbations might not necessarily have determinative power. Conversely, a set of inputs which has determinative power must be sensitive to perturbations. To prove those results, we show that the concentration of weight in the Fourier domain on certain sets of inputs characterizes a function in terms of tolerance to perturbations and determinative power of input nodes. Furthermore, we generalize a result by Xiao and Massey [11], which gives a necessary and sufficient condition of statistical independence of a set of inputs and a function's output in terms of the Fourier coefficients. This result can for instance be applied to decide for which classes of functions the algorithm presented in [12], which detects functional dependencies based on estimating mutual information, can succeed or fails. For *unate* functions, we show that any input and the function's output are statistically dependent and provide a direct relation between the mutual information and the influence of a variable. The class of unate functions is especially relevant for biological networks, as it includes all linear threshold functions and all nested canalizing functions, and describes functional dependencies in gene regulatory networks well [13]. 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 [6].

1.2 Outline

The paper is organized as follows. Boolean networks and Fourier analysis of Boolean functions are reviewed in section 2. In section 3, the influence and average sensitivity as measures of perturbations are reviewed, and their relation to the Fourier spectrum is discussed. In section 4, we study the mutual information of sets of inputs and the function's output. Section 5 is devoted to unate functions. Section 6 contains an analysis of the large-scale *E. coli* regulatory network, using the tools and ideas developed in previous sections.

2 Preliminaries

We start with a short introduction to Boolean networks and Fourier analysis of Boolean functions, and introduce notation.

2.1 Boolean networks

A (synchronous) 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 than $\{0, 1\}$ as more common in the literature on BNs will turn out to be advantageous later. However, both choices are equivalent. The state of the network at time t can be described by the vector $\mathbf{x}(t) = [x_1(t), ..., x_n(t)] \in \{-1, +1\}^n$. The network dynamic is defined by

$$x_i(t+1) = f_i(\mathbf{x}(t)), \tag{1}$$

where $f_i \colon \{-1, +1\}^n \to \{-1, +1\}$ is the Boolean function associated with node i. At time $t = 0$, an initial state $\mathbf{x}(0) = \mathbf{x}_0$ is chosen. In general, not all arguments $x_1, ..., x_n$ of a function $f_i(\mathbf{x})$ need to be *relevant*. The variable $x_j, j \in \{1, ..., n\}$ is said to be relevant for f_i if there exists at least one $\mathbf{x} \in \{-1, +1\}^n$, such that changing x_j to $-x_j$ changes the function's value. In most of the BN models in biology, the functions depend on a small subset of their arguments only. Furthermore, not every state must have a function associated with it; states can also be external inputs to the network.

To study the 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 with distribution $\mathrm{P}[X_i = x_i]$, $x_i \in \{-1, +1\}$. The assumption of independence holds for networks with tree-like topology, but is not feasible for networks with strong local dependencies and feedback loops. However, in many relevant settings, a BN has a tree-like topology, for instance the *E. coli* network analyzed in section 6. For a network with few local dependencies, assuming independence will lead to a small modeling error. Major results concerning the analysis of BNs have been obtained under the assumptions as stated above, e.g., the annealed approximation [14], an important result on the spread of perturbations in random BNs. Several important results on random BNs, e.g., [14], let the network size n tend to infinity; hence, there are no local dependencies.

2.2 Notation

We use $[n]$ for the set $\{1, 2, ..., n\}$, and all sets are subsets of $[n]$. With $\sum_{S \subseteq A} (\cdot)$, 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.

2.3 Fourier analysis of Boolean functions

In the following, we give a short introduction to 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 $P[X_i = x_i], x_i \in \{-1, +1\}$. Throughout this paper, probabilities $P[\cdot]$ and expectations $\mathbb{E}[\cdot]$ are with respect to the distribution of $\mathbf{X}$. We denote $p_i \triangleq P[X_i = 1]$, the variance of X_i by $\text{Var}(X_i)$, its standard deviation by $\sigma_i \triangleq \sqrt{\text{Var}(X_i)}$ and finally $\mu_i \triangleq \mathbb{E}[X_i]$. The inner product of $f, g : \{-1, +1\}^n \to \{-1, +1\}$ with respect to the distribution of $\mathbf{X}$ is defined as

$$\langle f, g \rangle \triangleq \mathbb{E}[f(\mathbf{X})g(\mathbf{X})] = \sum_{\mathbf{x} \in \{-1,1\}^n} P[\mathbf{X} = \mathbf{x}] f(\mathbf{x})g(\mathbf{x}) \tag{2}$$

which induces the norm $\|f\| = \sqrt{\langle f, f \rangle}$. An orthonormal basis with respect to the distribution of $\mathbf{X}$ is

$$\Phi_S(\mathbf{x}) = \prod_{i \in S} \frac{x_i - \mu_i}{\sigma_i}, \quad S \subseteq [n] \setminus \emptyset \tag{3}$$

and

$$\Phi_S(\mathbf{x}) = 1, \quad S = \emptyset.$$

This basis was first proposed by Bahadur [15]. Thus, each Boolean function $f \colon \{-1, +1\}^n \to \{-1, +1\}$ can be uniquely expressed as

$$f(\mathbf{x}) = \sum_{S \subseteq [n]} \hat{f}(S) \Phi_S(\mathbf{x}), \tag{4}$$

where $\hat{f}(S) \triangleq \langle f, \Phi_S \rangle$ are the Fourier coefficients of f. Note that (4) is a representation of f as a multilinear polynomial. As an example, consider the AND2 function defined as $f_{\text{AND2}}(\mathbf{x}) = 1$ if and only if $x_1 = x_2 = 1$, and let $p_1 = p_2 = 1/2$. According to (4)

$$f_{\text{AND}}(\mathbf{x}) = -\frac{1}{2} + \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_1 x_2.$$

As a second example consider PARITY2, i.e., the XOR function, defined as $f_{\text{PARITY2}}(\mathbf{x}) = 1$ if $x_1 = x_2 = 1$ or if $x_1 = x_2 = -1$, and $f_{\text{PARITY2}}(\mathbf{x}) = -1$ for all other choices of $\mathbf{x}$. Written as a polynomial, $f_{\text{PARITY2}}(\mathbf{x}) = x_1 x_2$. We conclude this section by listing properties of the basis functions which are used frequently throughout this paper.

Decomposition: Let $A \subseteq [n]$ and $S \subset A$, and denote $\bar{S} = A \setminus S$. Then,

$$\Phi_A(\mathbf{x}) = \Phi_S(\mathbf{x})\Phi_{\bar{S}}(\mathbf{x}).$$

Orthonormality: For $A, B \subseteq [n]$,

$$\mathbb{E}[\Phi_A(\mathbf{X})\Phi_B(\mathbf{X})] = \begin{cases} 1, \text{if } A = B \\ 0, \text{otherwise.} \end{cases}$$

Parseval's identity: For $f\colon \{-1, +1\}^n \to \{-1, +1\}$,

$$\mathbb{E}[f(\mathbf{X})^2] = \|f\|^2 = \sum_{S \subseteq [n]} \hat{f}(S)^2 = 1.$$

3   Influence and average sensitivity

Next, we discuss measures of perturbations and their relation to the Fourier spectrum. We start with a measure of the perturbation of a single input.

Definition 1 ([16]). *Define the influence of variable i on the function f as*

$$I_i(f) = \mathrm{P}[f(\mathbf{X}) \neq f(\mathbf{X} \oplus e_i)],$$

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 perturbing, i.e., flipping, input i changes the function's output. Influence can be viewed as the capability of input i to change the output of f. In BNs, usually, the sum of all influences, i.e., the *average sensitivity* is studied.

Definition 2. *The average sensitivity of f to the variables in the set A is defined as*

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

The average sensitivity of f is defined as $\mathrm{as}(f) \triangleq I_{\{1,\dots,n\}}(f)$.

$I_A(f)$ captures whether flipping an input chosen uniformly at random from A affects the function's output. Most commonly, all inputs are taken into account, i.e., the average sensitivity $\mathrm{as}(f)$ is studied. As an example, $\mathrm{as}(f_{\mathrm{PARITY2}}) = 2$ and $\mathrm{as}(f_{\mathrm{AND2}}) = 1$; hence, PARITY2 is more sensitive to single perturbations than AND2. Influence and average sensitivity have the following convenient expressions in terms of Fourier coefficients.

Proposition 1 (Lemma 4.1 of [17]). *For any Boolean function f,*

$$I_i(f) = \frac{1}{\sigma_i^2} \sum_{S \subseteq [n]\colon i \in S} \hat{f}(S)^2. \tag{5}$$

Proposition 2. *For any Boolean function f,*

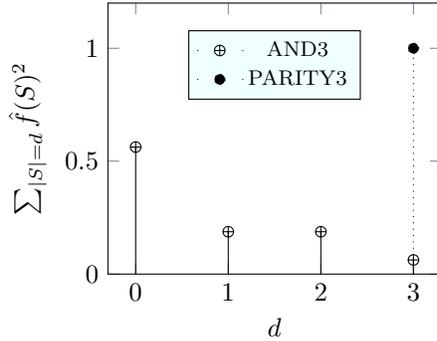$$I_A(f) = \sum_{S \subseteq [n]} \hat{f}(S)^2 \sum_{i \in S \cap A} \frac{1}{\sigma_i^2}. \tag{6}$$

Figure 1: The Fourier spectrum of AND3 and PARITY3.

Proposition 2 follows directly from Proposition 1 and the definition of $I_A(f)$. From (6), we see that as(f) is large if the Fourier weight is concentrated on the coefficients of high degree $d = |S|$, i.e., if $\sum_{S:\ |S| \geq d} \hat{f}(S)^2$ is large (i.e., close to one). For this case, Parseval's identity implies that the $\hat{f}(S)^2$ with $|S| < d$ must be small. Let's see an example: Suppose $p_1 = p_2 = p_3 = 1/2$ and consider the AND3 function, i.e., $f_{\mathrm{AND3}}(x_1, x_2, x_3) = 1$ if and only if $x_1 = x_2 = x_3 = 1$. f_{AND3} is tolerant to perturbations since as$(f_{\mathrm{AND3}}) = 0.75$, and as Figure 1 shows, its spectrum is concentrated on the coefficients of low degree. In contrast for $f_{\mathrm{PARITY3}}(x_1, x_2, x_3) \triangleq x_1 x_2 x_3$, as$(f_{\mathrm{PARITY}}) = 3$. Hence, PARITY3 is maximally sensitive to perturbations. Figure 1 shows that its spectrum is maximally concentrated on the coefficient of highest degree.

According to (6) as(f) is small only if the Fourier weight is concentrated on the coefficients of low degree. This is the case either if f is strongly biased (i.e., if $f(\mathbf{x}) = a$, for most inputs $\mathbf{x}$, where $a \in \{-1, 1\}$ is a constant) or if f depends on few variables only. This is in accordance with the results of Kauffman [1]; he found that a random BN operates in the ordered regime if the functions in the network depend on average on few variables.

We will state our result for measures of single perturbations. However, these results also apply to other noise models, specifically to the *noise sensitivity* of f. That is, because the noise sensitivity of f is small if f is tolerant to single perturbations. The noise sensitivity of a Boolean function is defined as the probability that the function's output changes if each input is flipped independently with probability ϵ. For uniformly distributed $\mathbf{X}$, ϵ as(f) is an upper bound for the noise sensitivity; for small values of ϵ, ϵ as(f) approximates the noise sensitivity well. For the X_i being equally but possibly nonuniformly distributed and a slightly different noise model, it was found in [18] that ϵ as(f) still upper bounds the noise sensitivity. This result was generalized to product distributed $\mathbf{X}$ in [19].

4   Mutual information and uncertainty

In this section, we study the determinative power of a subset of variables $\mathbf{X}_A$, where $\mathbf{X}_A$ consists of the entries of $\mathbf{X}$ corresponding to the indices in the set $A \subseteq [n]$, over the function's output $f(\mathbf{X})$. As a measure of determinative power, we take the mutual information $\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A)$ between $f(\mathbf{X})$ and $\mathbf{X}_A$, since $\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A)$ quantifies the statistical dependence between the random variable $\mathbf{X}_A$ and $f(\mathbf{X})$. Hence, this section is devoted to the study of $\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A)$.

Before giving a formal definition of mutual information, let us start with an example. Consider the PARITY2 function and let its inputs X_1, X_2 be uniformly distributed. Intuitively, if X_1 has determinative power, knowledge about X_1 should provide us with information about $f_{\text{PARITY2}}(\mathbf{X})$. Suppose we know the value of X_1, say $X_1 = 1$. Since $f_{\text{PARITY2}}(\mathbf{x}) = x_1 x_2$, we have with $\text{P}[X_2 = 1] = 1/2$ that $\text{P}[f_{\text{PARITY2}}(\mathbf{X}) = 1] = \text{P}[f_{\text{PARITY2}} = 1 | X_1 = 1]$. Hence, knowledge of X_1 does not help to predict the value of $f_{\text{PARITY2}}(\mathbf{X})$. Therefore, X_1 has no determinative power over $f_{\text{PARITY2}}(\mathbf{X})$. We indeed have $\text{MI}(f_{\text{PARITY2}}(\mathbf{X}); X_1) = 0$.

We next define mutual information. Mutual information is the reduction of uncertainty of a random variable Y due to the knowledge of X; therefore, we need to define a measure of uncertainty first, which is entropy. As a reference for the following definitions, see [20].

Definition 3. *The entropy $H(X)$ of a discrete random variable X with alphabet $\mathcal{X}$ is defined as*

$$H(X) \triangleq -\sum_{x \in \mathcal{X}} \text{P}[X = x] \log_2 \text{P}[X = x].$$

Definition 4. *The conditional entropy $H(Y|X)$ of a pair of discrete and jointly distributed random variables (Y, X) is defined as*

$$H(Y|X) \triangleq \sum_{x \in \mathcal{X}} \text{P}[X = x] \, H(Y|X = x).$$

Definition 5. *The mutual information $\text{MI}(Y; X)$ is the reduction of uncertainty of the random variable Y due to the knowledge of X*

$$\text{MI}(Y; X) \triangleq H(Y) - H(Y|X).$$

For a binary random variable X with alphabet $\mathcal{X} = \{x_1, x_2\}$ and $p \triangleq \text{P}[X = x_1]$, we have $H(X) = h(p)$, where $h(p)$ is the binary entropy function, defined as

$$h(p) \triangleq -p \log_2 p - (1 - p) \log_2 (1 - p). \tag{7}$$

The properties of mutual information are what we intuitively expect from a measure of determinative power: 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

$$\sum_{i=1}^{n} \text{MI}(f(\mathbf{X}); X_i) \leq \text{MI}(f(\mathbf{X}); \mathbf{X}) \leq 1, \tag{8}$$

which follows from the chain rule of mutual information (as a reference, see [20]) and independence of the $X_i, i \in [n]$. Hence, if $\text{MI}(f(\mathbf{X}); X_i)$ is large, i.e., close to 1, we can be sure that X_i has determinative power over $f(\mathbf{X})$ since (8) implies that $\text{MI}(f(\mathbf{X}); X_j)$ for $j \neq i$ must be small then.

4.1 Mutual information and the Fourier spectrum

In order to study determinative power, its relation to measures of perturbations, and statistical dependencies, we start by characterizing the mutual information in terms of Fourier coefficients. Our results are based on the following novel characterization of entropy in terms of Fourier coefficients.

Theorem 1. *Let f be a Boolean function, let $\mathbf{X}$ be product distributed, and let $\mathbf{X}_A = \{X_i \colon i \in A\}$ be a fixed set of arguments, where $A \subseteq [n]$. Then,*

$$H(f(\mathbf{X})|\mathbf{X}_A) = \mathbb{E}\left[h\left(\frac{1}{2}\left(1 + \sum_{S \subseteq A} \hat{f}(S)\Phi_S(\mathbf{X}_A)\right)\right)\right],$$

where $h(\cdot)$ is the binary entropy function as defined in (7).

Proof. See Appendix B. For the special case of uniformly distributed $\mathbf{X}$, a proof appears in [21], in the context of designing S-boxes. $\square$

Using the definition of mutual information, an immediate corollary of Theorem 1 is the following:

Corollary 1. *Let f be a Boolean function, $\mathbf{X}$ be product distributed, and $\mathbf{X}_A = \{X_i \colon i \in A\}$. Then,*

$$\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A) = h\left(1/2(1 + \hat{f}(\emptyset))\right) - \mathbb{E}\left[h\left(\frac{1}{2}\left(1 + \sum_{S \subseteq A} \hat{f}(S)\Phi_S(\mathbf{X}_A)\right)\right)\right]. \tag{9}$$

Theorem 1 (and Corollary 1) shows that the conditional entropy $H(f(\mathbf{X})|\mathbf{X}_A)$ and the mutual information $\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A)$ are functions of the coefficients $\{\hat{f}(S) \colon S \subseteq A\}$ only. This already hints at a fundamental difference to the average sensitivity, since the average sensitivity depends on the coefficients $\{\hat{f}(S) \colon |S \cap A| > 0\}$, according to (6).

We next discuss $\mathrm{MI}(f(\mathbf{X}); X_i)$ based on (9). First, note that $\mathrm{MI}(f(\mathbf{X}); X_i)$ has previously been studied under the notion *information gain* as a measure of 'goodness' for split variables in greedy tree learners [22] and also under the notion of *informativeness* to quantify voting power [23]. According to (9), the mutual information $\mathrm{MI}(f(\mathbf{X}); X_i)$ just depends on $\hat{f}(\{i\})$, $\hat{f}(\emptyset)$, and p_i. In contrast, the influence $I_i(f)$ is a function of the coefficients $\{\hat{f}(S) \colon S \in [n], i \in S\}$, according to (5). In Figure 2, 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 $\hat{f}(\{i\}) = 0$. That can be formalized as follows: $\mathrm{MI}(f(\mathbf{X}); X_i)$ is convex in $\hat{f}(\{i\})$. This can be proven by taking the second derivative of (9) and observing that it is larger than zero for all pairs of values $(\hat{f}(\emptyset), \hat{f}(\{i\}))$ for which $\mathrm{MI}(f(\mathbf{X}); X_i)$ is defined. Next, from (9), we see that $\mathrm{MI}(f(\mathbf{X}); X_i) = 0$ if $\hat{f}(\{i\}) = 0$; hence, it follows that $\mathrm{MI}(f(\mathbf{X}); X_i) = 0$ if and only if $\hat{f}(\{i\}) = 0$, which proves the following result:

Corollary 2. *Let f be a Boolean function, and $\mathbf{X}$ be product distributed. X_i and $f(\mathbf{X})$ are statistically independent if and only if $\hat{f}(\{i\}) = 0$.*
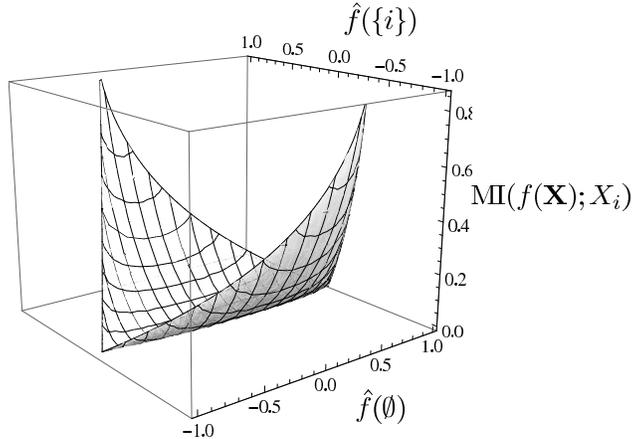
Figure 2: $\mathbf{MI}(\boldsymbol{f}(\mathbf{X}); \boldsymbol{X_i})$ as a function of $\hat{\boldsymbol{f}}(\{\boldsymbol{i}\})$ and $\hat{\boldsymbol{f}}(\emptyset)$ for $\boldsymbol{p_i} = \mathbf{0.3}$.

Corollary 2 also follows immediately from a more general result, namely Theorem 5, which is presented later. Recall that for PARITY2, $\mathrm{MI}(f_{\mathrm{PARITY2}}(\mathbf{X}); X_1) = 0$ and $\hat{f}(\{1\}) = 0$; hence, Corollary 2 comes at no surprise.

From Figure 2, it can be seen that the larger $|\hat{f}(\{i\})|$, the larger $\mathrm{MI}(f(\mathbf{X}); X_i)$ becomes. Formally, it follows from the convexity of $\mathrm{MI}(f(\mathbf{X}); X_i)$ and Corollary 2 that $\mathrm{MI}(f(\mathbf{X}); X_i)$ is increasing in $|\hat{f}(\{i\})|$. 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). $|\hat{f}(\{i\})|$ is trivially maximized for the *dictatorship function*, i.e., for $f(\mathbf{x}) = x_i$, or its negation, i.e., $f(\mathbf{x}) = -x_i$. The output $f(\mathbf{X})$ of the dictatorship function is fully determined by x_i.

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, $\mathrm{P}[f(\mathbf{X}) = 1] = 1/2$, 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 becomes. Let us finally relate the conditional entropy $H(f(\mathbf{X})|\mathbf{X}_A)$ to the concentration of the Fourier weight on the coefficients $\{S \colon S \subseteq A\}$, $A \subseteq [n]$.

Theorem 2. *Let f be a Boolean function, let $\mathbf{X}$ be product distributed, and let $\mathbf{X}_A = \{X_i \colon i \in A\}$ be a fixed set of arguments, where $A \subseteq [n]$. Then,*

$$\left(1 - \sum_{S \subseteq A} \hat{f}(S)^2\right)^{\frac{1}{\ln(4)}} \geq H(f(\mathbf{X})|\mathbf{X}_A) \geq 1 - \sum_{S \subseteq A} \hat{f}(S)^2.$$

Proof. See Appendix C. □

Theorem 2 shows that $H(f(\mathbf{X})|\mathbf{X}_A)$ can be approximated with $1 - \sum_{S \subseteq A} \hat{f}(S)^2$. It further shows that $H(f(\mathbf{X})|\mathbf{X}_A)$ is small if the Fourier weight is concentrated on the variables in the set A, i.e., if $\sum_{S \subseteq A} \hat{f}(S)^2$ is close to one. In contrast, as mentioned previously, for $I_A(f)$, it is relevant whether the Fourier weight is concentrated on the coefficients with high degree.

9

4.2 Relation to measures of perturbation

Mutual information and average sensitivity are related as follows.

Theorem 3. *For any Boolean function f, for any product distributed $\mathbf{X}$,*

$$I_A(f) \geq \min_{i \in A} \left(\frac{1}{\sigma_i^2} \right) (\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A) - \Psi(\mathrm{Var}\,(f(\mathbf{X})))) \tag{10}$$

with

$$\Psi(x) \triangleq (x)^{1/\ln(4)} - x. \tag{11}$$

Proof. See Appendix D. □

Note that the term $\Psi(\mathrm{Var}\,(f(\mathbf{X})))$ is close to zero. Specifically, for any $f(\mathbf{X})$ we have $0 \leq \Psi(\mathrm{Var}\,(f(\mathbf{X}))) < 0.12$, and for settings of interest, $\Psi(\mathrm{Var}\,(f(\mathbf{X})))$ is very close to zero, as explained in more detail in the following. Theorem 3 shows that if $\mathrm{MI}(f(\mathbf{X}); \mathbf{X}_A)$ if large (i.e., close to one), 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 determinative power). For the case that $A = [n]$, Theorem 3 states that the average sensitivity $\mathrm{as}(f)$ is lower-bounded by $\mathrm{MI}(f(\mathbf{X}); \mathbf{X})$ minus some small term.

We next discuss the special case that $A = \{i\}$. Theorem 3 evaluated for $A = \{i\}$ yields a lower bound on the influence of a variable in terms of the mutual information of that variable, namely

$$I_i(f) \geq \frac{1}{\sigma_i^2} (\mathrm{MI}(f(\mathbf{X}); X_i) - \Psi(\mathrm{Var}\,(f(\mathbf{X})))). \tag{12}$$

Again, $\Psi(\mathrm{Var}\,(f(\mathbf{X})))$ is close to zero for settings of interest, as the following argument explains. Equation (12) will not be evaluated for small $\mathrm{Var}\,(f(\mathbf{X}))$; since then, $f(\mathbf{X})$ is close to a constant function (i.e., close to $f(\mathbf{X}) = 1$ or $f(\mathbf{X}) = -1$), and $I_i(f)$ and $\mathrm{MI}(f(\mathbf{X}); X_i)$ must both be small (i.e., close to zero) anyway. Hence, (12) is of interest when $\mathrm{Var}\,(f(\mathbf{X}))$ is large, i.e., close to 1; for this case, the term $\Psi(\mathrm{Var}\,(f(\mathbf{X})))$ is small (e.g., for $\mathrm{Var}\,(f(\mathbf{X})) > 0.8$, $\Psi(\mathrm{Var}\,(f(\mathbf{X})) < 0.05$). Observe that, according to (12), 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 must be sensitive to perturbations. Conversely, as mentioned previously, an input i can have large influence and still $\mathrm{MI}(f(\mathbf{X}); X_i) = 0$. E.g., for the PARITY2 function, we have $I_i(f) = 1$ and $\mathrm{MI}(f(\mathbf{X}); X_i) = 0$.

Interestingly, the influence also has an information theoretic interpretation. The following theorem generalizes Theorem 1 in [23].

Theorem 4. *For any Boolean function f, for any product distributed $\mathbf{X}$,*

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

Proof. See Appendix E. For uniformly distributed $\mathbf{X}$, a proof appears in [23]. □

Theorem 4 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.

4.3 Statistical independence of inputs to a Boolean function

Next, 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 [11] from uniform to product distributed $\mathbf{X}$.

Theorem 5. *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*

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

Proof. See Appendix F. For uniformly distributed $\mathbf{X}$, i.e., $\mathrm{P}[X_i = 1] = 1/2$ for all $i \in [n]$, Theorem 5 has been derived by Xiao and Massey [11]. Note that the proof provided here is also conceptually different from the proof for the uniform case in [11], as it does not rely on the Xiao-Massey lemma. □

Theorem 5 shows that a function and small sets of its inputs are statistically independent if the spectrum is concentrated on the coefficients of high degree $d = |S|$. The most prominent example is the parity function of n variables, i.e., $f_{\mathrm{PARITYN}}(\mathbf{x}) = x_1 x_2 ... x_n$: For uniformly distributed $\mathbf{X}$, each subset of $n-1$ or fewer arguments and $f_{\mathrm{PARITYN}}(\mathbf{X})$ are statistically independent. Conversely, 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.

Theorem 5 also has an important implication for algorithms that detect functional dependencies in a BN based on estimating the mutual information from observations of the network's states, such as the algorithm presented in [12]. Theorem 5 characterizes the classes of functions for which such an algorithm may succeed and for which it will fail. Moreover, Theorem 5 shows that in a Boolean model of a genetic regulatory network, a functional dependency between a gene and a regulator cannot be detected based on statistical dependence of a regulator X_i and a gene's state $f_j(\mathbf{X})$, unless the regulatory functions are restricted to those for which $|\hat{f}(\{i\})| > 0$ holds for each relevant input i.

5 Unate functions

In this section, we discuss unate, i.e., locally monotone functions.

Definition 6. *A Boolean function f is said to be unate in x_i if for each $\mathbf{x} = (x_1, ..., x_n) \in \{-1, +1\}^n$ and for some fixed $a_i \in \{-1, +1\}$, $f(x_1, ..., x_i = -a_i, ..., x_n) \leq f(x_1, ..., x_i = a_i, ..., x_n)$ holds. f is said to be unate if f is unate in each variable x_i, $i \in [n]$.*

Each linear threshold function and nested canalizing function is unate. Moreover, most, if not all, regulatory interactions in a biological network are considered to be unate. That can be deduced from [13, 24], and the basic argument is the following: If an element acts either as a repressor or an activator for some gene, but never as both (which is a reasonable assumption for regulatory interactions[13, 24]), then the function determining the gene's state is unate by definition. For unate functions, the following property holds:

Proposition 3. *Let $f : \{-1, +1\}^n \to \{-1, +1\}$ be unate. Then,*

$$\hat{f}(\{i\}) = a_i \sigma_i I_i(f), \ \forall i \in [n], \tag{13}$$

where $a_i \in \{-1, +1\}$ is the parameter in Definition 6.

Proof. Goes along the same lines as the proof for monotone functions in Lemma 4.5 of [17]. $\square$

Note that conversely, if (13) holds for each $x_i, i \in [n]$, f is not necessarily unate. Inserting (13) into (9) yields

$$\mathrm{MI}(f(\mathbf{X}); X_i) = h\left(\frac{1}{2}(1 + \hat{f}(\emptyset))\right) - \mathbb{E}\left[h\left(\frac{1}{2}\left(1 + \hat{f}(\emptyset) + a_i \sigma_i I_i(f) \frac{X_i - \mu_i}{\sigma_i}\right)\right)\right], \tag{14}$$

where the expectation in (14) is over X_i. Based on (14), the discussion from section 4.1 on $\mathrm{MI}(f; X_i)$ applies by using $\hat{f}(\{i\})$ and $a_i \sigma_i I_i(f)$ synonymously. Hence, 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 functions actually depend on, then $|\hat{f}(\{i\})| > 0$. From this fact and the same arguments as given in section 4.1 follows:

Theorem 6. *Let $f \colon \{-1, +1\}^n \to \{-1, +1\}$ be unate. If and only if x_i is a relevant variable, then $\mathrm{MI}(f(\mathbf{X}); X_i) \neq 0$.*

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.

6   *E. coli* regulatory network

In [6], 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. Of the nodes, 636 represent genes and of the remaining 162 nodes, most (103) are external metabolites. The rest are stimuli, and others are state variables such as internal metabolites. The network has a layered feed-forward structure, i.e., no feedback loops exist. The elements in the first layer can be viewed as the inputs of the system, and the elements in the following seven layers are interacting genes that represent the internal state of the system. Our experiments revealed that all functions are unate; therefore, the properties derived in section 5 apply. Note that all functions being unate is a special property of the network, since if functions are chosen uniformly at random, it is unlikely to sample a unate function, in particular if the number of inputs n is large.

6.1   Determinative nodes in the *E. coli* network

We first identify the input nodes that have large determinative power (we will define what that means in a network setting shortly) and then show that a small number thereof reduces the uncertainty of the network's state significantly. Specifically, we show that on average, the entropy of the node's states conditioned on a small set of determinative input nodes, is small.

To put this result into perspective, we perform the same experiment for random networks with the same and different topology as the *E. coli* network. We denote by $\mathbf{X} = \{X_1, ..., X_n\}, n = 145$

the set of inputs of the feed forward network and assume that the X_i are independent and uniformly distributed. The remaining variables are denoted by $\mathbf{Y} = \{Y_1, ..., Y_m\}, m = 653$ and are a function of the inputs and the network's states, i.e., $Y_i = f_i'(\mathbf{X}, \mathbf{Y})$. For our analysis, the distributions of the random variables $Y_1, ..., Y_m$ need to be computed, since some of those variables are arguments to other functions. This can be circumvented by defining a collapsed network, i.e., a network where each state of a node is given as a function of the input nodes only, i.e., $Y_i = f_i(\mathbf{X})$. The collapsed network is obtained by consecutively inserting functions into each other, until each function only depends on states of nodes in the input layer, i.e., on $\mathbf{X}$. The collapsed network reveals the dependencies of each node on the input variables. Interestingly, in the collapsed network, it is seen that the variables chol_xt>0, salicylate, 2ddglcn_xt>0, mnnh>0, altrh>0, and his-l_xt>0 (here, and in the following, we adopt the names from the original dataset), which appear to be inputs when considering the original *E. coli* network, turn out to be not. Consider, for example, the node salicylate. The only node dependent on salicylate is mara = ((NOT arca OR NOT fnr) OR oxyr OR salicylate). However, arca = (fnr AND NOT oxyr), and it is easily seen that mara simplifies to mara = 1.

Next, we identify the determinative nodes. As argued in section 4, $\mathrm{MI}(f_i(\mathbf{X}); X_j)$ is a measure of the determinative power of X_j over $Y_i = f_i(\mathbf{X})$. This motivates the definition of the determinative power of input X_j over the states in the network as

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

Note that a small value of $D(j)$ implies that X_j alone does not have large determinative power over the network's states, but X_j may have large determinative power over the network states in conjunction with other variables. In principle $\sum_{i=1}^{m} \mathrm{MI}(f_i(\mathbf{X}); X_j, X_k)$ can be large for some $j, k \in [n]$, even though $D(j)$ and $D(k)$ are equal to zero. This is, however, not possible in the *E. coli* network since the functions are unate. Specifically, $\mathrm{MI}(f_i(\mathbf{X}); X_j, X_k) \neq 0$ implies that x_j or x_k are relevant variables, and according to Theorem 6, $\mathrm{MI}(f_i(\mathbf{X}); X_j) \neq 0$ or $\mathrm{MI}(f_i(\mathbf{X}); X_k) \neq 0$. We computed $D(j)$ for each input variable and found that $D(j)$ is large just for some inputs, such as o2_xt (37 bit), leu-l_xt (20.9 bit), glc-d_xt (19.3 bit), and glcn_xt>0 (17 bit), but is small for most other variables. Partly, this can be explained by the out-degree (i.e., the number of outgoing edges of a node) distribution of the input nodes. However, having a large out-degree does not necessarily result in large values of $D(j)$. In fact, in the *E. coli* network, glc-d_xt, glcn_xt>0, and o2_xt have 99, 93, and 73 outgoing edges, respectively. On the other hand, D(glc-d_xt) = 19.3 bit and D(glcn_xt>0) = 17 bit, whereas D(o2_xt) = 37 bit.

Denote τ as a permutation on $[n]$, such that $D(X_{\tau(1)}) \geq D(X_{\tau(2)}) \geq ... \geq D(X_{\tau(n)})$, i.e., τ orders the input nodes in descending order in their determinative power. We next consider $H(\mathbf{Y}|X_{\tau(1)}, ..., X_{\tau(l)})$ as a function of l to see whether knowledge of a small set of input nodes reduces the entropy of the overall network state significantly. $H(\mathbf{Y}|X_{\tau(1)}, ..., X_{\tau(l)})$ has an interesting interpretation which arises as a consequence of the so called asymptotic equipartition property [20] (as discussed in greater detail in [25]): Consider a sequence $\mathbf{y}_1, ..., \mathbf{y}_k$ of k samples of the random variable $\mathbf{Y}$. For $\epsilon > 0$ and k sufficiently large, there exists a set $A_\epsilon^{(k)}$ of typical sequences $\mathbf{y}_1, ..., \mathbf{y}_k$, such that

$$|A_\epsilon^{(k)}| \leq 2^{k(H(\mathbf{Y})+\epsilon)}$$

and
$$P\left[\mathbf{Y} \in A_\epsilon^{(k)}\right] > 1 - \epsilon,$$

where $|A_\epsilon^{(k)}|$ denotes the cardinality of the set $A_\epsilon^{(k)}$. This shows that the sequences obtained as samples of $\mathbf{Y}$ are likely to fall in a set of size determined by the uncertainty of $\mathbf{Y}$. Since the output layer consists of 653 nodes, the network's state space has maximal size 2^{653}. Since $\mathbf{Y}$ is a function of $\mathbf{X}$, $H(\mathbf{Y}) \leq H(\mathbf{X}) = 145$bit, where for the last equality, we assume uniformly distributed inputs. Thus, without knowing the state of any input variable, the network's state is likely to be in a set of size roughly 2^{145}. Given the knowledge about the states $X_{\tau(1)}, ..., X_{\tau(l)}$, the state of the network is likely to be in a set of size roughly $2^{H(\mathbf{Y}|X_{\tau(1)},...,X_{\tau(l)})}$. For a large network, however, $H(\mathbf{Y}|X_{\tau(1)}, ..., X_{\tau(l)})$ is expensive to compute as by definition:

$$H(\mathbf{Y}|\mathbf{X}_A) = \sum_{\mathbf{x}_A} P[\mathbf{X}_A = \mathbf{x}_A] \cdot \sum_{\mathbf{y}} P[\mathbf{Y} = \mathbf{y}|\mathbf{X}_A = \mathbf{x}_A] \log_2 P[\mathbf{Y} = \mathbf{y}|\mathbf{X}_A = \mathbf{x}_A]. \qquad (15)$$

Hence, the number of terms in the sum is exponential in n and $|A|$. An estimate of (15) can be obtained by sampling uniformly at random over $\mathbf{x}_A$ and $\mathbf{y}$. Instead, we will consider the following upper bound which is computationally inexpensive to compute:

$$H(\mathbf{Y}|X_{\tau(1)}, ..., X_{\tau(l)}) \leq A(l)$$

with

$$A(l) \triangleq \sum_{i=1}^{m} H(Y_i|X_{\tau(1)}, ..., X_{\tau(l)}).$$

The bound above follows from the chain rule for entropy [20]. $H(Y_i|X_{\tau(1)}, ..., X_{\tau(l)})$ is computationally inexpensive to compute, since Y_i depends on few variables only (in the *E. coli* network, on ≤ 8). For the *E. coli* network, $A(l)$ is depicted in Figure 3 as a function of l. Figure 3 shows that knowledge of the states of the most determinative nodes reduces the uncertainty about the network's states significantly. In fact, the upper bound $A(l)$ is loose; hence, we even expect $H(\mathbf{Y}|X_{\tau(1)}, ..., X_{\tau(l)})$ to lie significantly below $A(l)$. Also, note that when $A(l)$ is small, $H(Y_i|X_{\tau(1)}, ..., X_{\tau(l)})$ must be small on average; hence, $P\left[Y_i = 1|X_{\tau(1)}, ..., X_{\tau(l)}\right]$ is close to one or zero on average.

To put $A(l)$ for the *E. coli* network in Figure 3 into perspective, we compute $A(l)$ for random networks. First, we took the *E. coli* network and exchanged each function with one chosen uniformly at random from the set of all Boolean functions of corresponding degree. We also exchanged each function with one chosen uniformly at random from all unate functions. We performed the same experiment for the original *E. coli* network for 25 choices of random and random unate functions, respectively. The mean of $A(l)$, along with one standard deviation from the mean (dashed lines), is plotted in Figure 3 for random and random unate functions. It is seen that fewer inputs determine the output of the original *E. coli* network, compared to its random counterparts. For example, to obtain $A(l) = 50$, about twice as many inputs need to be known if the functions in the *E. coli* network are exchanged for functions chosen uniformly at random.

Next, we generated at random feed forward networks with $m = 653$ outputs and $n = 145$ inputs, each with out-degree 8, i.e., the average out-degree of the inputs in the collapsed *E. coli* network. Again, we computed $A(l)$ for 25 choices of random and random unate functions, respectively. The mean and one standard deviation from the mean are depicted in Figure 3. The results show that, as expected, for a random feed forward network, there seems to be no small set of inputs that determines the outputs.
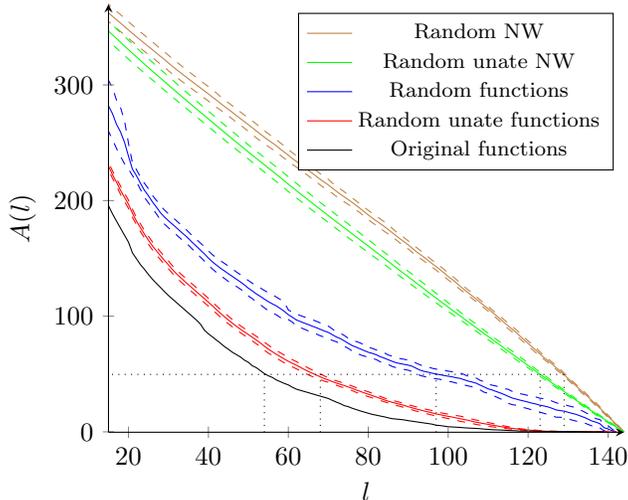
Figure 3: The upper bound $A(l)$ on $H(\mathbf{Y}|\mathbf{X}_{\tau(1)}, ..., \mathbf{X}_{\tau(l)})$ as a function of l for the *E. coli* network and random networks.

6.2 Tolerance to perturbations

Finally, we discuss the average sensitivity of individual functions in the *E. coli* network. In section 3, we found that the average sensitivity is small if the Fourier spectrum is concentrated on the coefficients of low degree. This appears to be the case for functions that are highly biased and for functions that depend on few variables only. Figure 4 shows pairs of values $(\mathrm{as}(f), \Pr[f(\mathbf{X}) = 1])$ for each function in the *E. coli* network, again assuming that the X_i are independent and uniformly distributed. We can see from Figure 4 that the average sensitivity of all functions is close to the lower bound on the average sensitivity. Note that the functions with high in-degree K (i.e., number of relevant input variables), which could have average sensitivity up to K, also have small average sensitivity, because those functions are highly biased. We, therefore, can conclude that the functions have small average sensitivity either because they depend on few variables only or because they are highly biased. For other input distributions, i.e., other values of $p = \mathrm{P}[X_i = 1], \forall i \in [n]$, we obtained the same results.

7 Conclusion

In a Boolean network, tolerance to perturbations, determinative power, and statistical dependencies between nodes are properties of single functions in a probabilistic setting. Hence, we analyzed single functions with product distributed argument. We used Fourier analysis of Boolean functions to study the mutual information between a function $f(\mathbf{X})$ and a set of its inputs $\mathbf{X}_A$, as a measure of determinative power of $\mathbf{X}_A$ over $f(\mathbf{X})$. We related the mutual information to the Fourier spectrum and proved that the mutual information lower bounds the influence, a measure of perturbation. We also gave necessary and sufficient conditions for statistical independence of $f(\mathbf{X})$ and $\mathbf{X}_A$. For the class of unate functions, which are particularly interesting for biological networks, we found that mutual information and influence are directly related (not just via an inequality). We also found
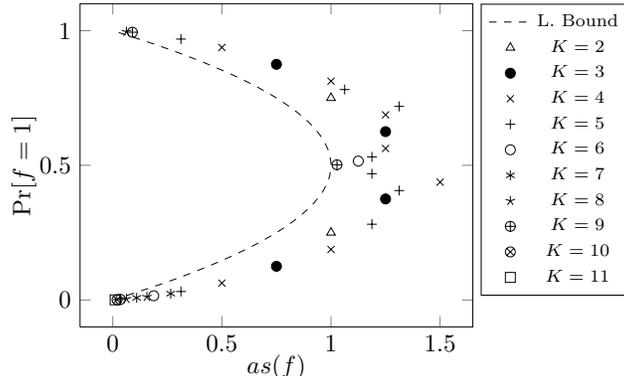
Figure 4: Average sensitivity in the *E. coli* network. Pairs of values $(\text{as}(f), \Pr[f(\mathbf{X}) = 1])$ of each function in the *E. coli* network for different in-degrees K and uniformly distributed $\mathbf{X}$. Moreover, a lower bound on the average sensitivity $\text{as}(f)$, i.e., Poincare's inequality, is plotted.

that $\text{MI}(f(\mathbf{X}); X_i) > 0$ for each relevant input i, which, as an application, implies that in a unate regulatory network, a gene and its regulator must be statistically dependent. As an application of our results, we analyzed the large-scale regulatory network of *E. coli*. We identified the most determinative input nodes in the network and found that it is sufficient to know only a small subset of those in order to reduce the uncertainty of the overall network state significantly. This, in turn, reduces the size of the state space in which the network is likely to be found significantly.

A possible direction for future work is to provide an analysis similar to that of the *E. coli* regulatory network for other Boolean models of biological networks, and see if similar conclusions as in section 6 can be reached. One of the main assumptions in our work is the independence among the input variables of the network. It would be interesting to provide methods that can be used beyond this setup. However, deriving such results is challenging because for dependent inputs, the basis functions $\Phi_S(\mathbf{x})$ do not factorize as in (3), and many results cited and derived in this paper make use of this particular form of the basis functions. In this paper, we focused on generic properties of information-processing networks that may help identify possible principles that underlie biological networks. Assessing our findings from a biological perspective would be an interesting next step.

Appendices

A   Lemma 1

For the proof of Theorems 1 and 5, we will need the following lemma:

Lemma 1. *Let f be a Boolean function, let $\mathbf{X}$ be product distributed, and let $A \subseteq [n]$ and some fixed $\mathbf{x_A} \in \{-1, +1\}^{|A|}$ be given. Then,*

$$\mathbb{E}[f(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A] = \sum_{S \subseteq A} \hat{f}(S)\Phi_S(\mathbf{x}_A). \tag{16}$$

Proof. Inserting the Fourier expansion of $f(\mathbf{X})$ given by (4) in the left-hand side of (16) and utilizing the linearity of conditional expectation yields

$$\mathbb{E}[f(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A] = \sum_{S \subseteq [n]} \hat{f}(S) \, \mathbb{E}[\Phi_S(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A].$$

For $S \subseteq A$,

$$\mathbb{E}[\Phi_S(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A] = \Phi_S(\mathbf{x}_A).$$

Conversely, for $S \not\subseteq A$,

$$\mathbb{E}[\Phi_S(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A] = 0.$$

To see this, assume without loss of generality that $S = A \cup \{j\}$ and $j \notin A$. Using the decomposition property of the basis function as given in section 2.3,

$$\mathbb{E}[\Phi_S(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A] = \mathbb{E}\left[\prod_{i \in S} \Phi_{\{i\}}(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A\right]$$

$$= \prod_{i \in S} \mathbb{E}[\Phi_{\{i\}}(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A]$$

which is equal to zero as

$$\mathbb{E}[\Phi_{\{j\}}(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A] = \mathbb{E}[\Phi_{\{j\}}(\mathbf{X})] = 0.$$

$\square$

B Proof of Theorem 1

First,

$$\mathrm{P}[f(\mathbf{X}) = 1|\mathbf{X}_A = \mathbf{x}_A] = \frac{1}{2}\left(1 + \mathbb{E}[f(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A]\right)$$

$$= \underbrace{\frac{1}{2}\left(1 + \sum_{S \subseteq A} \hat{f}(S)\Phi_S(\mathbf{x}_A)\right)}_{q(\mathbf{x}_A)}, \tag{17}$$

where (17) follows from an application of Lemma 1. By definition of the conditional entropy,

$$H(f(\mathbf{X})|\mathbf{X}_A) = \sum_{\mathbf{x}_A \in \{-1,1\}^{|A|}} \mathrm{P}[\mathbf{X}_A = \mathbf{x}_A] \, H(f(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A)$$

$$= \sum_{\mathbf{x}_A \in \{-1,1\}^{|A|}} \mathrm{P}[\mathbf{X}_A = \mathbf{x}_A] \, h(\mathrm{P}[f(\mathbf{X}) = 1|\mathbf{X}_A = \mathbf{x}_A])$$

$$= \sum_{\mathbf{x}_A \in \{-1,1\}^{|A|}} \mathrm{P}[\mathbf{X}_A = \mathbf{x}_A] \, h(q(\mathbf{x}_A)) \tag{18}$$

$$= \mathbb{E}[h(q(\mathbf{X}_A))], \tag{19}$$

where $h(\cdot)$ is the binary entropy function as defined in (7). To obtain (18), we used (17). The expectation in (18) is with respect to the distribution of $\mathbf{X}_A$. Inserting $q(\mathbf{X}_A)$ as given by (17) in (19) concludes the proof.

C Proof of Theorem 2

First, note that with $q(\cdot)$ as defined in (17), we have

$$
\begin{aligned}
\mathbb{E}[4q(\mathbf{X}_A)(1 - q(\mathbf{X}_A))] &= \mathbb{E}\left[1 - \left(\sum_{S \subseteq A} \hat{f}(S)\Phi_S(\mathbf{X_A})\right)^2\right] \\
&= \sum_{S \subseteq A}\sum_{U \subseteq A} \hat{f}(S)\hat{f}(U)\mathbb{E}[\Phi_S(\mathbf{X}_A)\Phi_U(\mathbf{X}_A)] \\
&= 1 - \sum_{S \subseteq A} \hat{f}(S)^2,
\end{aligned}
\tag{20}
$$

where (20) follows from the orthogonality of the basis functions.

We start with proving the lower bound in Theorem 2. Applying the lower bound on the binary entropy function $h(p) \geq 4p(1 - p)$, given in Theorem 1.2 of [26], on (19) yields

$$
H(f(\mathbf{X})|\mathbf{X}_A) = \mathbb{E}[h(q(\mathbf{X}_A))] \geq \mathbb{E}[4q(\mathbf{X}_A)(1 - q(\mathbf{X}_A))],
$$

and the lower bound in Theorem 2 follows using (20).

Next, we prove the upper bound in Theorem 2. Applying the upper bound on the binary entropy function $h(p) \leq (p(1 - p))^{1/\ln(4)}$, given in Theorem 1.2 of [26], on (19) yields

$$
\begin{aligned}
H(f(\mathbf{X})|\mathbf{X}_A) &= \mathbb{E}[h(q(\mathbf{X}_A)] \\
&\leq \mathbb{E}[(\underbrace{4q(\mathbf{X}_A)(1 - q(\mathbf{X}_A))}_{Y})^{1/\ln(4)}].
\end{aligned}
\tag{21}
$$

The term Y in (21) is a random variable, and the function $(Y)^{1/\ln(4)}$ is concave in Y. An application of Jensen's inequality (see e.g. [20]) yields $\mathbb{E}[(Y)^{1/\ln(4)}] \leq (\mathbb{E}[Y])^{1/\ln(4)}$; hence, the right-hand side of (21) can be lower as

$$
H(f(\mathbf{X})|\mathbf{X}_A) \leq (\mathbb{E}[4q(\mathbf{X}_A)(1 - q(\mathbf{X}_A))])^{1/\ln(4)}.
\tag{22}
$$

Finally, the upper bound in Theorem 2 follows from combining (22) and (20).

D Proof of Theorem 3

According to Proposition 2,

$$
\begin{aligned}
I_A(f) &= \sum_{S \subseteq [n]} \hat{f}(S)^2 \sum_{i \in S \cap A} \frac{1}{\sigma_i^2} \\
&\geq \sum_{S \subseteq [n] \setminus \emptyset} \hat{f}(S)^2 |S \cap A| \min_{i \in A}\left(\frac{1}{\sigma_i^2}\right) \\
&\geq \min_{i \in A}\left(\frac{1}{\sigma_i^2}\right) \sum_{S \subseteq A \setminus \emptyset} \hat{f}(S)^2.
\end{aligned}
\tag{23}
$$

18

Next, we rewrite the lower bound on $H(f(\mathbf{X})|\mathbf{X}_A)$ given by Theorem 2 as

$$\sum_{S\subseteq A\backslash\emptyset}\hat{f}(S)^2 \geq 1-\hat{f}(\emptyset)^2-H(f(\mathbf{X})|\mathbf{X}_A). \tag{24}$$

By adding $H(f(\mathbf{X}))-H(f(\mathbf{X}))$ on the right-hand side of (24) and using the definition of mutual information, (24) becomes

$$\sum_{S\subseteq A\backslash\emptyset}\hat{f}(S)^2 \geq \mathrm{MI}(f(\mathbf{X});\mathbf{X}_A)-H(f(\mathbf{X}))+1-\hat{f}(\emptyset)^2. \tag{25}$$

With $\mathrm{Var}\,(f(\mathbf{X}))=1-\hat{f}(\emptyset)^2$ and by using the inequality $H(f(\mathbf{X}))\leq (\mathrm{Var}\,(f(\mathbf{X})))^{1/\ln(4)}$, given in Theorem 1.2 of [26], (25) becomes

$$\sum_{S\subseteq A\backslash\emptyset}\hat{f}(S)^2 \geq \mathrm{MI}(f(\mathbf{X});\mathbf{X}_A)-\Psi(\mathrm{Var}\,(f(\mathbf{X}))), \tag{26}$$

with $\Psi(\cdot)$ as defined in (11). Finally, Theorem 3 follows by combining (23) and (26).

E   Proof of Theorem 4

For notational convenience, let $A=[n]\setminus\{i\}$. By definition of the conditional entropy,

$$\begin{aligned} H(f(\mathbf{X})|\mathbf{X}_A) &= \sum_{\mathbf{x}_A\in\{-1,1\}^{|A|}} \mathrm{P}[\mathbf{X}_A=\mathbf{x}_A]\,H(f(\mathbf{X})|\mathbf{X}_A=\mathbf{x}_A) \\ &= \sum_{\mathbf{x}_A\in\{-1,1\}^{|A|}} \mathrm{P}[\mathbf{X}_A=\mathbf{x}_A]\,h(\mathrm{P}[f(\mathbf{X})=1|\mathbf{X}_A=\mathbf{x}_A]), \end{aligned} \tag{27}$$

where $h(\cdot)$ is the binary entropy function as defined in (7). Observe that

$$h(\mathrm{P}[f(\mathbf{X})=1|\mathbf{X}_A=\mathbf{x}_A])=h(\mathrm{P}[X_i=1])$$

if

$$\begin{aligned} f(X_1=x_1,...,X_i=1,...,X_n=x_n) \\ \neq f(X_1=x_1,...,X_i=-1,...,X_n=x_n) \end{aligned}$$

and

$$h(\mathrm{P}[f(\mathbf{X})=1|\mathbf{X}_A=\mathbf{x}_A])=0$$

otherwise. Hence, 27 becomes

$$H(f(\mathbf{X})|\mathbf{X}_A) = \sum_{\mathbf{x}_A\in\{-1,1\}^{|A|}} \mathrm{P}[\mathbf{X}_A=\mathbf{x}_A]\,h(p_i)\mathbf{1}_{\{f(\mathbf{X})\neq f(\mathbf{X}\oplus e_i)\}},$$

where $\mathbf{x}\oplus e_i$ is the vector obtained from $\mathbf{x}$ by flipping its ith entry, and Theorem 4 follows by using the definition of the influence.

F   Proof of Theorem 5

By definition, $f(\mathbf{X})$ and $\mathbf{X}_A$ are statistically independent if and only if for all $\mathbf{x}_A \in \{-1, +1\}^{|A|}$

$$P[f(\mathbf{X}) = 1 | \mathbf{X}_A = \mathbf{x}_A] = P[f(\mathbf{X}) = 1]. \tag{28}$$

With

$$P[f(\mathbf{X}) = 1 | \mathbf{X}_A = \mathbf{x}_A] = \frac{1}{2} + \frac{1}{2}\mathbb{E}[f(\mathbf{X})|\mathbf{X}_A = \mathbf{x}_A]$$

and application of Lemma 1 given in Appendix 1, (28) becomes

$$\sum_{S \subseteq A} \hat{f}(S)\Phi_S(\mathbf{x_A}) = \hat{f}(\emptyset)$$

$$\Leftrightarrow \sum_{S \subseteq A \setminus \emptyset} \hat{f}(S)\Phi_S(\mathbf{x_A}) = 0. \tag{29}$$

It follows from the Fourier expansion (4) that (29) holds for all $\mathbf{x}_A \in \{-1, +1\}^{|A|}$ if and only if $\hat{f}(S) = 0$ for all $S \subseteq A \setminus \emptyset$, which proves the theorem.

Acknowledgments

References

[1] S. Kauffman, Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. **22**(3), 437–467 (1969)

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

[3] S. Davidich, M.I. Bornholdt, Boolean network model predicts cell cycle sequence of fission yeast. PLoS ONE. **3**(2), e1672 (2008)

[4] J. Saez-Rodriguez, L. Simeoni, J.A. Lindquist, R. Hemenway, U. Bommhardt, B. Arndt, U Haus, R Weismantel, E.D. Gilles, S. Klamt, B. Schraven, A logical model provides insights into T cell receptor signaling. PLoS Comput. Biol. **3**(8), e163 (2007)

[5] F. Li, T. Long, Y. Lu, Q. Ouyang, C. Tang, The yeast cell-cycle network is robustly designed. Proc. Natl. Acad. Sci. USA **101**(14), 4781–4786 (2004)

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

[7] J. Kesseli, P. Rämö, O. Yli-Harja, On spectral techniques in analysis of Boolean networks. Physica D: Nonlinear Phenomena. **206**(1–2), 49–61 (2005)

[8] J. Kesseli, P. Rämö, O. Yli-Harja, Tracking perturbations in Boolean networks with spectral methods. Phys Rev E **72**(2), 026137 (2005)

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

[10] R. Heckel, S. Schober, M. Bossert, Determinative power and tolerance to perturbations in Boolean networks. Paper presented at the 9th international workshop on computational systems biology, Ulm, Germany, 4-6 June 2012

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

[12] S. Liang, S. Fuhrman, R. Somogyi, Reveal, a general reverse engineering algorithm for inference of genetic network architectures. Pacific Symposium on Biocomputing **3** 18–29 (1998)

[13] J. Grefenstette, S. Kim, S. Kauffman, An analysis of the class of gene regulatory functions implied by a biochemical model. Bio. Syst. **84**(2), 81–90 (2006)

[14] B. Derrida, Y. Pomeau, Random networks of automata: a simple annealed approximation. Europhysics. Lett. **1**(2), 45–49 (1986)

[15] R.R. Bahadur, A representation of the joint distribution of responses to n dichotomous items. in *Studies in Item Analysis and Prediction*, ed. by H. Solomon (Stanford University Press, Stanford, 1961), pp.158–168

[16] M. Ben-Or, N. Linial, Collective coin flipping, robust voting schemes and minima of Banzhaf values. Paper presented at the 26th annual symposium on foundations of computer science, Portland, Oregon, USA, 21–23 October 1985

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

[18] S. Schober, About Boolean networks with noisy inputs. Paper presented at the fifth international workshop on computational systems biology, Leipzig, Germany, 11–13 June 2008

[19] V. Matache, M.T. Matache, On the sensitivity to noise of a Boolean function. J. Math. Phys. **50**(10), 103512 (2009)

[20] J.A. Cover, T.M. Thomas, *Elements of Information Theory*, 2nd edn. (Wiley-Interscience, New York, 2006)

[21] R. Forre, Methods and instruments for designing S-boxes. J. Cryptology **2**(3), 115-130 (1990)

[22] B. Rosell, L. Hellerstein, S. Ray, Why skewing works: learning difficult Boolean functions with greedy tree learners. Paper presented at the 22nd international conference on machine learning, Bonn, Germany, 7–11 August 2005

[23] A. Diskin, M. Koppel, Voting power: an information theory approach. Soc. Choice Welfare **34**, 105–119 (2010)

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

[25] S. Schober, *Analysis and Identification of Boolean Networks using Harmonic Analysis*, (Der Andere Verlag, Germany, 2011)

[26] F. Topsoe, Bounds for entropy and divergence for distributions over a two-element set. Inequalities Appl. Pure Math. **2**, (2001)