Supporting Information

Random Boolean Network Models and the Yeast Transcriptional Network
S. Kauffman, C. Peterson, B. Samuelsson and C. Troein
Proceedings of the National Academy of Sciences USA **100**, 14796-14799 (2003)

Confronting Nested Canalyzing Functions with Compiled Data

In order to compare compiled and generated distributions of rules, we must ensure that every nested canalyzing function is always represented by the same set of parameters $I_1, \ldots, I_K$ and $O_1, \ldots, O_K$ (see *Appendix* in the printed article). All ambiguities in the choice of the representation can be derived from the following operations:

1. The transformation $I_K \rightarrow \text{NOT } I_K$ together with $O_K \rightarrow \text{NOT } O_K$ and $O_{\text{default}} \rightarrow \text{NOT } O_{\text{default}}$.

2. Permutations among a set of inputs $i_m, \ldots, i_{m+p}$ such that $O_m = \cdots = O_{m+p}$. The values of $I_m, \ldots, I_{m+p}$ are permutated in the same way as $i_m, \ldots, i_{m+p}$.

A unique representation is created from any choice of parameters in two steps. First, 1. is applied if $O_K \neq O_{K-1}$, which ensures that $O_K = O_{K-1}$. In order to handle the special case $K = 1$ in a convenient way we define $O_0 = \text{FALSE}$. Second, all intervals of inputs $i_m, \ldots, i_{m+p}$ such that 2. can be applied are identified and permutated so that $I_m = \cdots = I_{m+q} = O_m$ and $I_{m+q+1} = \cdots = I_{m+p} \neq O_m$ for some q, $0 \leq q \leq p$.

Using the above described procedure, we can compare a generated rule distribution with the compiled distribution. First, we take away all redundant inputs of each observed rule. An input is redundant if the output is never dependent on that input. That renders 2, 7, 70, 34 and 5 nested canalyzing rules with 1, 2, 3, 4 and 5 inputs respectively. Second, we let $\alpha = 7$ and generate rule distributions for each number of inputs. ($\alpha = 7$ is not based on a precise fit, it was picked by hand to fit the distribution of $I_1, \ldots, I_K$.) Table 2 shows the result for the most frequently observed rules, and Fig. 1 is a plot of the full rule distribution. The calculated distribution fits surprisingly well to the compiled one, considering that the model has only one free parameter, α.

n_{obs}	$(I_1 \to O_1)$, ..., $(I_K \to O_K)$	Boolean expression
29	$(0\to0), (0\to0), (0\to0)$	i_1 AND i_2 AND i_3
20	$(0\to0), (0\to0), (1\to0)$	i_1 AND i_2 AND NOT i_3
9	$(0\to0), (1\to1), (1\to1)$	i_1 AND (i_2 OR i_3)
9	$(0\to0), (0\to0), (0\to0), (0\to0)$	i_1 AND i_2 AND i_3 AND i_4
7	$(0\to0), (0\to0), (0\to0), (1\to0)$	i_1 AND i_2 AND i_3 AND NOT i_4
6	$(0\to0), (0\to0), (0\to1), (0\to1)$	i_1 AND i_2 AND NOT (i_3 AND i_4)
5	$(0\to0), (0\to1), (0\to1)$	i_1 AND NOT (i_2 AND i_3)
4	$(0\to0), (0\to0)$	i_1 AND i_2
3	$(0\to0), (1\to0)$	i_1 AND NOT i_2
3	$(0\to0), (1\to0), (1\to0)$	i_1 AND NOT (i_2 OR i_3)
3	$(0\to0), (1\to1), (0\to1)$	i_1 AND (i_2 OR NOT i_3)
3	$(0\to0), (0\to0), (1\to1), (0\to1)$	i_1 AND i_2 AND (i_3 OR NOT i_4)
3	$(0\to0), (1\to0), (1\to1), (1\to1)$	i_1 AND NOT i_2 AND (i_3 OR i_4)
2	$(0\to0)$	i_1
2	$(0\to0), (0\to0), (1\to0), (1\to0)$	i_1 AND i_2 AND NOT (i_3 OR i_4)
2	$(0\to0), (0\to0)$, (non-canalyzing)	i_1 AND i_2 AND (NOT i_3 AND i_4 OR NOT i_4 AND i_5)
1	$(0\to1), (0\to0), (0\to0)$	NOT i_1 OR i_2 AND i_3
1	$(0\to0), (1\to0), (0\to1), (0\to1)$	i_1 AND NOT (i_2 OR i_3 AND i_4)
1	$(0\to0), (1\to1), (0\to0), (0\to0)$	i_1 AND (i_2 OR i_3 AND i_4)
1	$(0\to0), (1\to1), (1\to1), (1\to1)$	i_1 AND (i_2 OR i_3 OR i_4)
1	$(1\to1), (0\to1), (0\to0), (1\to0)$	i_1 OR NOT i_2 OR i_3 AND NOT i_4
1	$(0\to0), (0\to0), (0\to0), (0\to0), (0\to0)$	i_1 AND i_2 AND i_3 AND i_4 AND i_5
1	$(0\to0), (0\to0), (0\to0), (0\to0), (1\to0)$	i_1 AND i_2 AND i_3 AND i_4 AND NOT i_5
1	$(0\to0), (0\to0), (1\to0), (0\to1), (0\to1)$	i_1 AND i_2 AND NOT (i_3 OR i_4 AND i_5)
1	$(0\to0), (0\to0), (1\to0), (1\to1), (0\to1)$	i_1 AND i_2 AND NOT i_3 AND (i_4 OR NOT i_5)
1	$(0\to0), (1\to0), (1\to1), (0\to0), (1\to0)$	i_1 AND NOT i_2 AND (i_3 OR i_4 AND NOT i_5)
1	$(0\to0), (0\to0)$, (non-canalyzing)	i_1 AND i_2 AND (i_3 XOR i_4)
1	$(0\to0)$, (non-canalyzing)	i_1 AND (i_2 XOR i_3 AND i_4)
1	$(0\to0)$, (non-canalyzing)	i_1 AND $(2 \leq)(i_2, i_3,$ NOT $i_4)$
1	$(1\to0)$, (non-canalyzing)	NOT i_1 AND (i_2 AND NOT i_3 OR i_3 AND NOT (i_4 OR i_5))
1	(non-canalyzing)	(i_1 AND i_2) OR NOT (i_1 OR i_2 OR i_3)

Table 1: The list of rules compiled in [?]. n_{obs} is the number of occurrences, and the rules are described both as an ordinary Boolean expression, and with the parameters $I_1, \ldots, I_K$ and $O_1, \ldots, O_K$, where $O_{\mathrm{default}} =$ NOT O_K. 0 and 1 correspond to FALSE and TRUE, respectively. The Boolean function $(2 \leq)$ is TRUE if at least two of its arguments are TRUE. (NOT has higher operator precedence than AND, whereas the precedences of OR and XOR are lower.)

Label	n_{obs}	n_{calc}	$(I_1 \to O_1), \ldots, (I_K \to O_K)$	Boolean expression
A	29	28	$(0\to0),(0\to0),(0\to0)$	i_1 AND i_2 AND i_3
B	20	26	$(0\to0),(0\to0),(1\to0)$	i_1 AND i_2 AND NOT i_3
c	9	1	$(0\to0),(1\to1),(1\to1)$	i_1 AND $(i_2$ OR $i_3)$
D	9	6	$(0\to0),(0\to0),(0\to0),(0\to0)$	i_1 AND i_2 AND i_3 AND i_4
E	7	9	$(0\to0),(0\to0),(0\to0),(1\to0)$	i_1 AND i_2 AND i_3 AND NOT i_4
F	6	2	$(0\to0),(0\to0),(0\to1),(0\to1)$	i_1 AND i_2 AND NOT $(i_3$ AND $i_4)$
G	5	3	$(0\to0),(0\to1),(0\to1)$	i_1 AND NOT $(i_2$ AND $i_3)$
H	4	5	$(0\to0),(0\to0)$	i_1 AND i_2
I	3	2	$(0\to0),(1\to0)$	i_1 AND NOT i_2
j	3	4	$(0\to0),(1\to0),(1\to0)$	i_1 AND NOT $(i_2$ OR $i_3)$
k	3	5	$(0\to0),(1\to1),(0\to1)$	i_1 AND $(i_2$ OR NOT $i_3)$
l	3	3	$(0\to0),(0\to0),(1\to1),(0\to1)$	i_1 AND i_2 AND $(i_3$ OR NOT $i_4)$
m	3	0	$(0\to0),(1\to0),(1\to1),(1\to1)$	i_1 AND NOT i_2 AND $(i_3$ OR $i_4)$
N	2	2	$(0\to0)$	i_1
o	2	4	$(0\to0),(0\to0),(1\to0),(1\to0)$	i_1 AND i_2 AND NOT $(i_3$ OR $i_4)$

Table 2: Compiled and generated rule distributions for all nested canalyzing rules observed more than once. n_{obs} is the number of observations in the compiled list of rules, whereas n_{calc} is the average number of rules in the generated distribution. Each rule is described both as an ordinary Boolean expression, and with the parameters $I_1, \ldots, I_K$ and $O_1, \ldots, O_K$, where $O_{\mathrm{default}} = \mathrm{NOT}\ O_K$. 0 and 1 correspond to FALSE and TRUE, respectively. The labels serve as references in Fig. 1, and capital labels mark rules that are chain functions.

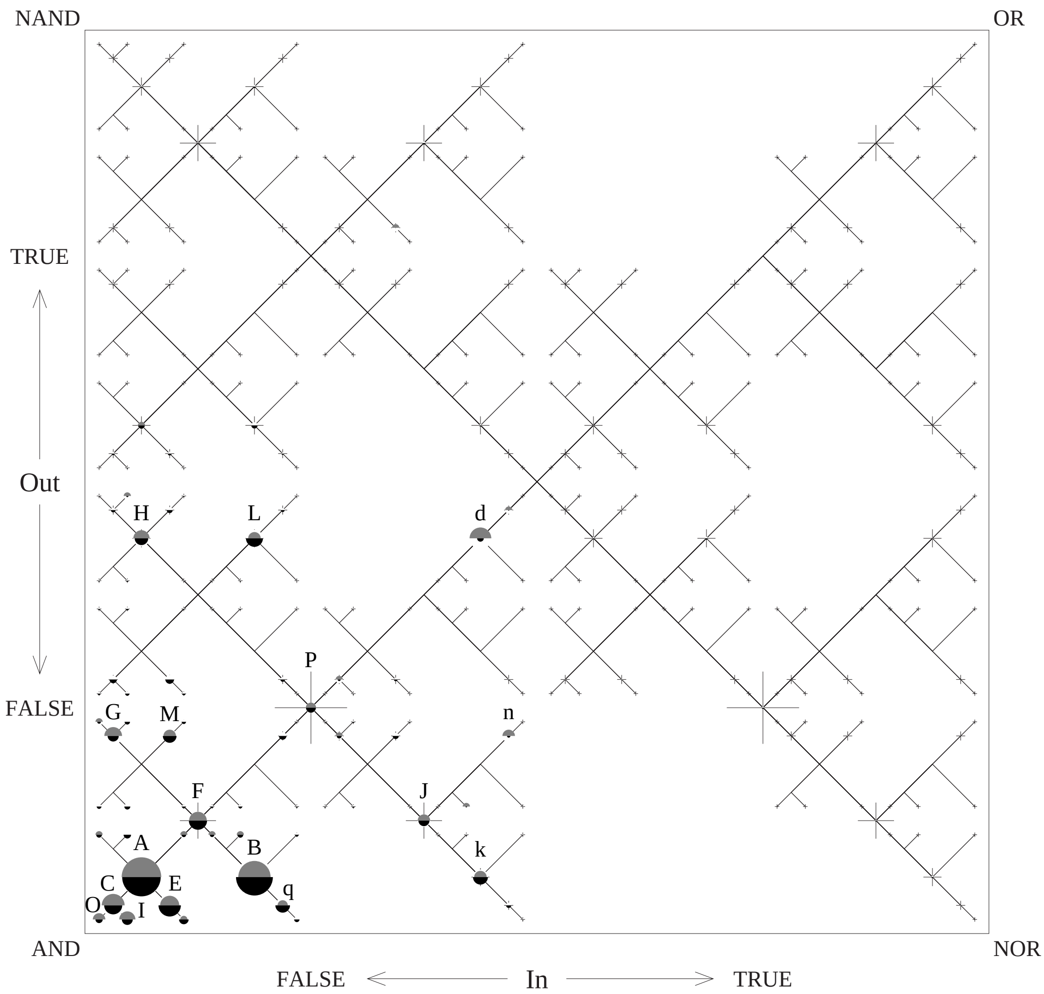

Figure 1: Compiled and generated rule distributions of nested canalyzing functions. The gray half-circles have an area proportional to the number of times each rule has been observed, while their black counterparts reflect the calculated distribution. The labeled rules are listed in Table 2. Capital labels mark rules that are chain functions. Each rule is assigned a coordinate in the unit square above (having $(0,0)$ as its lower left corner), according to $x = 1/2 + \sum_{m=1}^{K} 2^{-m}\phi(I_m)$, $y = 1/2 + \sum_{m=1}^{K} 2^{-m}\phi(O_m)$, where $\phi(\text{TRUE}) = 1/2$ and $\phi(\text{FALSE}) = -1/2$. The crosses mark the possible coordinates for a rule that is represented in its unique form. The lines indicate how the coordinates can change when new inputs are added to an existing rule.