Compensatory interactions to stabilize multiple steady states or mitigate the effects of multiple deregulations in biological networks

Gang Yang, Colin Campbell, and Réka Albert

Compensatory interactions to stabilize multiple steady states or mitigate the effects of multiple deregulations in biological networks

Gang Yang,[1, *] Colin Campbell,[1, 2] and Réka Albert[1, †]

[1]*Department of Physics, Pennsylvania State University, University Park, Pennsylvania 16802, USA*
[2]*Department of Physics, Washington College, Chestertown, Maryland 21620, USA*

Complex diseases can be modeled as damage to intra-cellular networks that results in abnormal cell behaviors. Network-based dynamic models such as Boolean models have been employed to model a variety of biological systems including those corresponding to disease. Previous work designed compensatory interactions to stabilize an attractor of a Boolean network after single node damage. We generalize this method to a multi-node damage scenario and to the simultaneous stabilization of multiple steady state attractors. We classify the emergent situations, with a special focus on combinatorial effects, and characterize each class through simulation. We explore how the structural and functional properties of the network affect its resilience and its possible repair scenarios. We demonstrate the method′s applicability to two intra-cellular network models relevant to cancer. This work has implications in designing prevention strategies for complex disease.

I. INTRODUCTION

Complex networks are increasingly used to understand and simulate the behavior of biological systems such as cellular signaling networks [1–6]. The network-based dynamic modelling approach aims to capture the biological function and behavior of these systems as an emergent property that arises from the totality of interactions among the components [7]. Several researchers have successfully used network-based approaches such as Boolean and logical models to study specific biological processes [2, 3, 8]. Complex diseases including diabetes and cancers can be modeled as network damage due to temporary or permanent node perturbation (e.g. constitutive activation of a protein arising from a genetic mutation) [2, 9]. Thus the topics of network repair and network control have drawn significant attention in the scientific community [10–13]. Most approaches aim to influence network dynamics by controlling the states of certain nodes of the network [11, 12]. Recently, another approach to the network control problem, namely modifying the interactions in the network, was proposed [10]. Using this approach, compensatory interventions can be found to stabilize an attractor (e.g. steady state) of the network after damage to a single node [10]. These interventions can be implemented as preventive measures or applied immediately after the onset of damage. The effect of the intervention is that the perturbation does not propagate to the rest of the network, and a close-to-normal behavior is restored [10]. Ultimately, a combination of node-based and edge-based approaches will provide researchers more potential therapeutic strategies.

Recent research suggests that complex diseases such as cancer often involve multiple gene mutations and the "one disease, one target, one drug" approach may not be effective to battle these diseases [9, 14, 15]. Thus it is worthwhile to use the network paradigm to explore the combinatorial effect of multiple gene mutations, and to

* gzy105@psu.edu
† rza1@psu.edu

design control measures to prevent these effects. Moreover, many biological systems were shown to have several possible steady states (e.g. several possible cell types), each reachable for alternative histories (time courses) [2, 3, 7]. Repair interventions should be cognizant of these alternative states and maintain (or eliminate) them as necessary or desired in the specific context. Here we generalize the method of Campbell *et al.* [10] to a multiple node damage setting, and to systems that have multiple steady states, aiming to provide a theoretical platform to mitigate damage more realistically.

This paper presents three key results. First, we use analytical and computational methods to study how network structure and regulatory logic affect the resilience of the network′s steady states to single node perturbation. Second, we present an algorithm to design compensatory interventions to stabilize a steady state of the network after double node damage and evaluate it on random Boolean networks. Third, we apply the algorithm on stabilizing two steady states simultaneously after a single node damage and discuss the emerging situations and their corresponding frequencies. We apply the above algorithms to two biological examples and also adapt the latter algorithm to the alternative goal of stabilizing a steady state and destabilizing another.

II. BACKGROUND: BOOLEAN MODELING

A network is a mathematical abstraction of a set of relationships between various elements. The network consists of nodes that represent the different elements and edges that specify the pairwise relationships between them [16, 17]. In biological networks at the molecular level, nodes are molecular species such as small molecules, RNA, protein, and edges indicate interactions and regulatory relationships [7, 18]. A substantial amount of studies have characterized the topological properties of networks, such as degree distribution, heterogeneity and community structure [16, 17]. Biological networks were found to exhibit interesting topological properties such

as a heterogeneous degree distribution [19, 20]. However, in order to understand the biological function of a system, the network's topological information alone is not enough and dynamical information should be incorporated. More specifically, in a dynamical model, each node i is characterized by a state variable σ_i, which can be continuous or discrete, and the vector $(\sigma_1, \cdots, \sigma_n)$ represents the state of the system [7]. The state of the system can be followed in continuous time or at discrete time intervals. In discrete time models, the activity of each node σ_i is described by a regulatory rule $\sigma_i(t + \tau_i) = f(\sigma_{i_1}(t), \cdots, \sigma_{i_k}(t))$, where $i_1, \cdots, i_k$ are the regulating nodes of i and τ_i is a discrete time delay.

Here we focus on discrete time Boolean network models, where node states are binary, 1(ON) or 0(OFF), and the regulatory rule is specified by a truth table. This is motivated by the fact that biological species are frequently observed to have highly nonlinear regulation and switch-like behavior; thus the node state 1 means the molecular species is above a threshold concentration or activity and the node state 0 means it is below a threshold concentration or activity [7, 18]. The time trajectory of the system is simulated deterministically or stochastically depending on the updating scheme. A simple deterministic updating scheme is synchronous updating, where $\tau_i = 1$ for every node. For this scheme, given a specific initial state, the system will deterministically evolve into an attractor, which can be a steady state (fixed point) or several states that repeat regularly, called a limit cycle. Steady states can be interpreted as cell types and limit cycles correspond to a cell cycle or circadian rhythms [7]. A commonly used stochastic updating scheme is general asynchronous updating, where a random node is selected to be updated at each time step [21]. This type of update is motivated by the fact that different biological processes have various timescales, and often the timescales of specific processes are not known [22]. Fixed points (steady states) do not depend on the updating scheme [23]. However, limit cycles are generally unstable to infinitesimal deviation from synchronous updating [23], and the known variety of time scales in biological processes makes limit cycles observed in synchronous updating schemes inherently suspect. While stochastic update may lead to attractors that involve irregular repetitions of a set of states (so-called complex attractors) [7], we here focus our attention on steady state attractors. Abnormal behavior of a certain element can be modelled as a change in the node state, either a temporary perturbation or permanent damage [2, 9]. For example, a loss-of-function mutation or the knockout of a gene can be represented as a permanent OFF state of the corresponding node in the network.

III. RESULTS

A. The influence of single node damage on a steady state of a system

We consider a Boolean model of a biological system; this model will have one or several attractors. We start from a steady state s. Then we consider damage to a node i by permanent knockout (sustained OFF state) or constitutive expression or activity (sustained ON state). If the damaged state s^* is a new steady state (i.e. other nodes are not affected by the perturbation), we say that steady state s is stable against the damage. In the converse case, the state of one or more nodes will change, which then has a cascading effect in the biological system. We say the steady state s needs repair in order to prevent damage propagation. We define the sensitive node set S_i as the set of nodes that would change their state as a direct consequence of the damage to node i.

Previous research has studied the relationship between a network's structure and its topological resilience to incremental node loss [24] and the relationship between average degree and the effect of single node damage [10]. It was shown that the larger the average node degree, the less stable a steady state is against single node damage. Another related result is that random Boolean network ensembles will go through a phase transition from a frozen phase to a chaotic phase as the average node degree increases. Two states that initially differ in a single node's state will diverge on average in the chaotic phase. The critical boundary is average degree $<K> = 2$ when considering unbiased Boolean logic (all Boolean functions) and using an annealed approximation (at every time step the input nodes and Boolean functions are randomized for each node)[25–30]. We note that our setting of a steady state damaged by a single node knockout is different from what was considered in previous work on random Boolean network ensembles.

As biological networks have been observed to exhibit degree heterogeneity and long-tailed decreasing degree distributions [1, 19, 31, 32], we explore the effect of degree heterogeneity on the resilience of a steady state following single node knockout. To probe a variety of regulatory rules consistent with a given number of regulators, we first consider random Boolean rules, and then focus on more realistic nested canalizing Boolean functions.

1. Theoretical estimation of resilience probability in case of single node damage

We define the resilience probability (RP) of a steady state as the probability that the steady state of the network is stable against single node damage. It follows that the damage probability $DP = 1 - RP$. We define $\alpha(I_j)$ to be the probability that a node j with in-degree I_j is stable (does not change state) if one of its randomly chosen inputs, i, is knocked out. Knocking out a node i will directly affect the state of at most O_i nodes, where O_i is the out-degree of node i. If we denote the nodes regulated by i as $n_1, n_2, \cdots, n_{O_i}$, then the probability that the state of the system is a steady state after we knock out node i alone is $p(i) = \prod_{i=1}^{O_i} \alpha(I_{n_i})$ since every regulated node must be stable for the overall network to be stable. The average RP is given by $RP = \frac{1}{N} \sum_{i=1}^{N} p(i)$. Under the mean-field assumption that every node follows the same node in-degree distribution $f(I_i)$ and out-degree

distribution $g(O_i)$, the average RP can be estimated as[1]

$$RP = \sum_{O_i} g(O_i)(\sum_{I_j} f(I_j)\alpha(I_j))^{O_i} \qquad (1)$$

In all cases $\alpha(I_j = 0) = 1$ as a source node cannot, by definition, be disrupted by any other node. If each possible Boolean function occurs with equal chance, $\alpha(I_j) = \frac{1}{2}$ for $I_j > 0$; this is due to the equal probability of having 0 or 1 values in each position of the function, which leads to a chance of one half that a change in value of an input variable does not lead to a change in value of the output[2]. However, to make sure that the regulatory logic correctly reflects the desired topology, we use effective Boolean rules wherein no input is redundant or spurious [34]. That is, for any input node i, $f(\cdots, \sigma_i = 1, \cdots) \neq f(\cdots, \sigma_i = 0, \cdots)$ for at least one pair of input configurations. We find by exhaustive enumeration that for effective rules, the probability $\alpha(I_j)$ changes with the in-degree of the affected node j: $\alpha(I_j = 1) = 0, \alpha(I_j = 2) = 0.4, \alpha(I_j = 3) \approx 0.477, \alpha(I_j = 4) \approx 0.498$. A Monte Carlo calculation shows that α approaches 0.5 as node in-degree increases. Thus one can readily see from the estimated average RP (formula 1) that $(\sum_{I_j} f(I_j)\alpha(I_j))^{O_i}$ decreases from 1 exponentially as O_i increases from 0. Thus sink nodes and nodes with smaller out-degree have a greater contribution to the resilience probability, as they affect no or few other nodes. Given an average node out-degree, heterogeneity in the out-degree distribution tends to make the steady state of the network more stable against single node damage because it leads to more low-degree nodes. However, since $\alpha(I_j)$ increases relatively slowly and saturates at 0.5 as I_j increases, it is less straightforward to see the dependence between in-degree heterogeneity and the resilience probability of a steady state. We note that our mean-field approximation takes out-degree distribution and effective Boolean rules into consideration compared with annealed approximation.

We also analyze the effect of restricting the Boolean rules to nested canalizing rules, as research shows that the regulation in biological networks is frequently described in this way [35]. A nested canalizing Boolean function with k inputs can be generated by determining two sequences, the input sequence $(I_1, I_2, \cdots, I_k)$ and the output sequence $(O_1, O_2, \cdots, O_k)$, where I_i or O_i is either 0 or 1. The output o as a function of input configuration $(i_1, \cdots, i_k)$ is thus determined through the hierarchy $o = O_1$ if $i_1 = I_1$; $o = O_2$ if $i_1 \neq I_1$ and $i_2 = I_2$; $\cdots$; $o = O_k$ if $i_1 \neq I_1, \cdots, i_{k-1} \neq I_{k-1}, i_k = I_k$; $o = NOT \ O_k$ if $i_1 \neq I_1, \cdots, i_{k-1} \neq I_{k-1}, i_k \neq I_k$. The last condition is used to guarantee that the rule is an effective rule [35]. All nested canalizing functions can be written in the above form up to a permutation of node order. We determine analytically, and verify by numeri-

cal simulations, that the probability that a node's state will not change after knockout of one of its x regulators is $\alpha(x) = \frac{x-1}{x}$ for nested Boolean functions generated by the method above with no bias in I_i or O_i. This is because knocking out the first dominant canalizing variable i_1 (the probability of this is $1/x$), will change the input configuration; the output will be changed with probability $1/2$, which is the probability that two outputs O_1 and $O_l(l \neq 1)$ of the nested Boolean function hierarchy are different. Knocking out the second dominant canalizing variable changes the output only if $i_1 \neq I_1$ (the probability of this is $1/2$), the probability that output is changed is $1/2$ as before under the condition $i_1 \neq I_1$, and so on. Also notice that the order of the last two inputs in the hierarchy of the canalizing function does not affect the resilience probability, thus the probability of needing repair is $\frac{1}{x}(\frac{1}{2} + (\frac{1}{2})^2 + \cdots + (\frac{1}{2})^{x-1} + (\frac{1}{2})^{x-1}) = \frac{1}{x}$, and therefore $\alpha(x)$ is $\frac{x-1}{x}$. Notice that two different sequences may give the same rule, for example, for a one-input rule, $(I_1 = 1, O_1 = 1)$ is actually the same as $(I_1 = 0, O_1 = 0)$. Also, nested canalizing function ensembles generated by the input and output sequence with no bias lead to a different degeneracy of the Boolean functions in a Boolean table representation, in which the output of the Boolean function is specified for each possible input configuration. Simulations show that nested canalizing Boolean functions randomly picked from the Boolean table representation with equal probability have a different $\alpha(x)$ function: $\alpha(x = 2) = 0.5, \alpha(x = 3) = 0.625, \alpha(x = 4) = 0.712, \alpha(x = 5) = 0.766$. Regardless of the representation, α is larger for nested canalizing functions compared with random Boolean functions or effective random Boolean functions. This indicates that steady states of networks with nested Boolean functions will have an increased resilience against single node damage [35–37]. Since α is smaller than 1, the conclusion that heterogeneity in the out-degree distribution tends to make the steady state of the network more resilient against single node damage holds for nested canalizing functions.

2. Damage probability in simulations of random network ensembles

To estimate the resilience probability, we consider five random Boolean network ensembles with different in-degree/out-degree distributions, namely (a) constant in-degree and scale-free out-degree distribution (SF_out), (b) constant in-degree and Poissonian out-degree (NK_out), (c) constant in-degree and constant out-degree (NKK), (d) Poissonian in-degree distribution and constant out-degree (NK_in), and (e) scale-free in-degree distribution and constant out-degree (SF_in). The algorithm we used in generating these networks will give scale-free (power-law) degree distribution or Poisson degree distribution in the limit of very large network size. Even for small network sizes, the heterogeneity of these two types of networks is significantly different, e.g. the standard deviation of the first's is approximately twice

---

[1] To be exact, the in-degree distribution $f(I_i)$ in formula 1 should be the conditional in-degree distribution conditioned on a node being knocked out. The conditional in-degree distribution can be obtained through the in-degree distribution reweighted by in-degree.
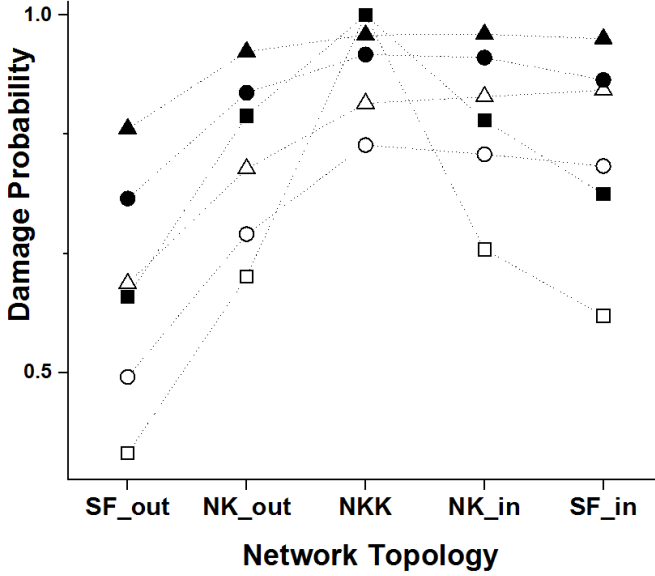
[2] A similar result was obtained in [33].

FIG. 1: The estimated damage probability across five ensembles with different degree heterogeneity. Different symbol shapes of the series represent different average degree, $<K> = 1$ (squares), $<K> = 2$ (circles), and $<K> = 3$ (triangles). Single node knockout results are shown with empty symbols and double node knockout results are shown with solid symbols. The standard error of the average damage probability is estimated to be in the order of 0.001, which is negligible compared with the size of the symbol.

the second's [3]. For each ensemble, we generate 1000 networks with 20 nodes. To make sure that the generated ensemble has the desired topology and degree distribution, we only accept at least weakly connected networks and use effective rules when assigning a Boolean function to each node. We study ensembles with average degree $<K> = 1$, 2 and 3, which would be in frozen phase for $<K> = 1$ and chaotic phase for $<K> = 2$ or 3 when considering the annealed approximation, the infinite network size limit and unbiased *effective* Boolean rules [26–30]. Note that knowing the phase is not enough to predict the damage probability. For each network, we find all the steady states. For each steady state, we individually knock out (keep in the OFF state) every node that has the ON state in the steady state. A similar procedure can be followed to consider the constitutive expression (sustained ON state) of nodes that are currently OFF in the attractor; we do not explicitly consider this latter case.

We estimate the resilience probability (RP) and damage probability ($DP = 1 - RP$) for networks with given topological characteristics by considering all steady states and all possible node knockouts with equal proba-

---

[3] When the average degree equals 2, the standard deviation of the node out-degree of one sample ensemble is 1.336 for Poisson distribution, 2.675 for scale-free distribution. When the average degree equals 3, the standard deviation is 1.588 for Poisson distribution, 3.188 for scale-free distribution.

bility in the corresponding network ensemble. We do not weight our analysis based on the size of the attraction basin of steady states, as this property is not necessarily relevant in biological systems [38]. Fig. 1 summarizes the simulation results for the estimated damage probability. In agreement with the theoretical result, for single node damage, given a fixed node in-degree, heterogeneity in out-degree leads to a smaller damage probability for the steady state (compare SF_out, NK_out and NKK results). In contrast, with node out-degree fixed, heterogeneity in in-degree distribution does not show a general trend and is connectedness dependent: the damage probabilities of the NKK, NK_in and SF_in ensembles are close for $<K> = 2$ or $<K> = 3$. Thus the theoretical analysis (see III A 1 2nd paragraph) is consistent with the computational result. A quantitative comparison of damage probability estimation by simulations and mean-field theory is shown in Table I for selected ensembles.

TABLE I: A quantitative comparison of damage probability estimation

Average degree/Method		SF_out	NK_out	NKK	NK_in	SF_in
$<K> = 2$	Simulation	0.494	0.694	0.818	0.805	0.789
$<K> = 2$	Mean-field	0.496	0.711	0.84	0.842	0.824
$<K> = 3$	Simulation	0.624	0.785	0.877	0.886	0.895
$<K> = 3$	Mean-field	0.609	0.805	0.891	0.905	0.906

Single node damage probability estimation by simulations (1st and 3rd row) and mean-field theory (2nd and 4th row). Different columns correspond to different ensembles. Mean-field calculations employ the degree distributions of the generated ensemble.

B. Double node damage

1. Classification of the resilience scenarios of double node damage

In this section, we generalize the single node damage repair algorithm proposed in Ref. [10] and investigate the properties of interventions that prevent the cascading effect of knocking out two nodes in a network. The potential combinatorial effects of simultaneous damage to two nodes have been named genetic interactions in biological systems [9]. A specific example of cases where combined knockout of two genes has a stronger effect than the sum of the effects of the individual knockouts is synthetic lethality [39–41]. The converse case was termed synthetic viability [41, 42]. Both of these genetic interactions have been studied experimentally [39, 40] and theoretically [41, 42]. Here we go beyond the identification of genetic interactions by determining the specific edge additions through which the cascading effects of cumulative damage can be prevented.

When repair is necessary, for each sensitive node, we define candidate nodes as nodes that are neither its pre-existing regulators nor the sensitive node itself, and we add a suitable interaction starting from a candidate node to prevent the state change. (We avoid using pre-existing regulators since it is less biologically feasible

TABLE II: Classifications of different situations comparing single node damage and double node damage.

Class/ subclass	Status of SS after single node damage	Status of SS after double node damage
1 (b)	stable for both	stable
2 (c)	stable for both	Needs repair
3 (a)	stable in one case, needs repair in the other case	stable
4 (a,b,c,d)	stable in one case, needs repair in the other case	Needs repair
5 (a)	Both need repair	stable
6 (a,b,c,d)	Both need repair	Needs repair
a	$S_A \cup S_B \supsetneq S_{AB}$	
b	$S_A \cup S_B = S_{AB}$	
c	$S_A \cup S_B \subsetneq S_{AB}$	
d	$(S_A \cup S_B)\backslash S_{AB} \neq \emptyset \quad and \quad S_{AB}\backslash(S_A \cup S_B) \neq \emptyset$	

The first column is the class and its possible subclasses, whose definitions are given in the last four rows. The subclasses are defined based on the relationships between S_{AB} and $S_A \cup S_B$. $S_A \cup S_B \supsetneq S_{AB}$ means S_{AB} is a true subset of $S_A \cup S_B$. $(S_A \cup S_B)\backslash S_{AB} \neq \emptyset$ and $S_{AB}\backslash(S_A \cup S_B) \neq \emptyset$ means that S_{AB} is not a subset of $S_A \cup S_B$ and $S_A \cup S_B$ is also not a subset of S_{AB}, where $\backslash$ means relative complement. SS means steady state.

[10].) This way, we preserve the steady state aside from the immediate impact on the damaged node and block the cascading effect as soon as possible. Specifically, say node i is regulated by nodes that belong to set A, $x_i = f(x_{j_1}, \cdots, x_{j_k})$, where $j_1, \cdots, j_k \in A$. If one wants to repair node i so that it remains ON ($x_i = 1$), one needs to find a candidate node l (i.e. $l \notin A$ and $l \neq i$) and modify the rule such that $x_i = f(x_{j_1}, \cdots, x_{j_k})$ OR x_l if $x_l = 1$ or such that $x_i = f(x_{j_1}, \cdots, x_{j_k})$ OR (NOT x_l) if $x_l = 0$. Here AND, OR and NOT are Boolean functions. Similarly, if one wants to repair node i so that it remains OFF ($x_i = 0$), one can modify the rule to be $x_i = f(x_{j_1}, \cdots, x_{j_k})$ AND (NOT x_l) if $x_l = 1$; or $x_i = f(x_{j_1}, \cdots, x_{j_k})$ AND x_l if $x_l = 0$ [10]. Assuming that a candidate node with the appropriate x_l value exists, which is generally the case in realistic networks, regulation of this sort is always possible in principle [10]. We say that a repair solution exists if each sensitive node can be repaired. In the algorithm for double node damage, the sensitive node set is determined after knockout of both nodes; then for each sensitive node, a candidate node set is identified with the additional restriction of excluding both damaged nodes. Then, similarly to single node knockout, an interaction is added from an appropriate candidate node to each sensitive node.

When considering the damage of node A, damage of a different node B, and damage of both nodes, six outcomes are possible, which are summarized in the first six rows of Table II. These six outcomes have been discussed in the context of synthetic lethality in random threshold networks [41]. In order to compare the repair solutions,

TABLE III: The ten two-input effective Boolean functions and their classification in double node damage

Function Name	(1,1)	(1,0)	(0,1)	(0,0)	Class
(NOT A) OR (NOT B)	0	1	1	1	6b
(NOT A) OR B	1	0	1	1	3a
A OR (NOT B)	1	1	0	1	3a
A OR B	1	1	1	0	2c
XNOR(A,B)	1	0	0	1	5a
XOR(A,B)	0	1	1	0	5a
A AND B	1	0	0	0	6b
A AND (NOT B)	0	1	0	0	3a
(NOT A) AND B	0	0	1	0	3a
(NOT A) AND (NOT B)	0	0	0	1	2c

The first column indicates the name of the function of A and B, where XOR(A,B)= (A AND (NOT B)) OR ((NOT A) AND B), XNOR(A,B)=NOT XOR(A,B). The second to fifth columns list the value of the respective function for input combinations (A=1, B=1), (A=1, B=0), (A=0, B=1), (A=0, B=0). The sixth column gives the classification (as in Table II) of a network motif composed of three nodes, wherein A and B are source nodes and the regulation of C is described by the respective Boolean function of A and B. This classification assumes that the input nodes are $A = 1, B = 1$ initially. The letter in the last column gives the subclass based on the relationship between $S_A \cup S_B$ and S_{AB} as described in Table II.

we denote the sensitive node sets after damage to node A, B, and both A and B as S_A, S_B, and S_{AB}, respectively. If no node is a child node of both node A and node B, $S_{AB} = S_A \cup S_B$. However, if a node is a child node of both nodes A and B, different situations can emerge as indicated in the last four rows of Table II. For completeness, for each class (i.e., situation) we list the possible subclasses in the first column of the table. Classes 1, 2, 3, and 5 admit a single subclass only, while classes 4 and 6 can have any of the four subclasses.

In order to gain insight into how different networks lead to the different outcomes of Table II, we consider a simple network motif, in which two nodes (A, B) regulate a third node C. We determine which class and subclass each two-variable Boolean function belongs (Table III). We start from state (1, 1) for the two nodes (in the order A, B). If the output of state (0,1) is different from that of state (1,1), the state needs repair after damage to node A; similar conclusions apply to all the cases. The symmetrical AND rule and its negation belong to class 6, the symmetrical OR rule and its negation belong to class 2, four cases of unsymmetrical two-variable regulation belong to class 3, and the XOR/XNOR functions belong to class 5. The three node motif holds the same properties when embedded within a larger network. However, we emphasize that a network containing a three-node motif and additional nodes does not necessarily fall into the same category as the three-node motif alone, since different parts of the network may be different for different damage situations.
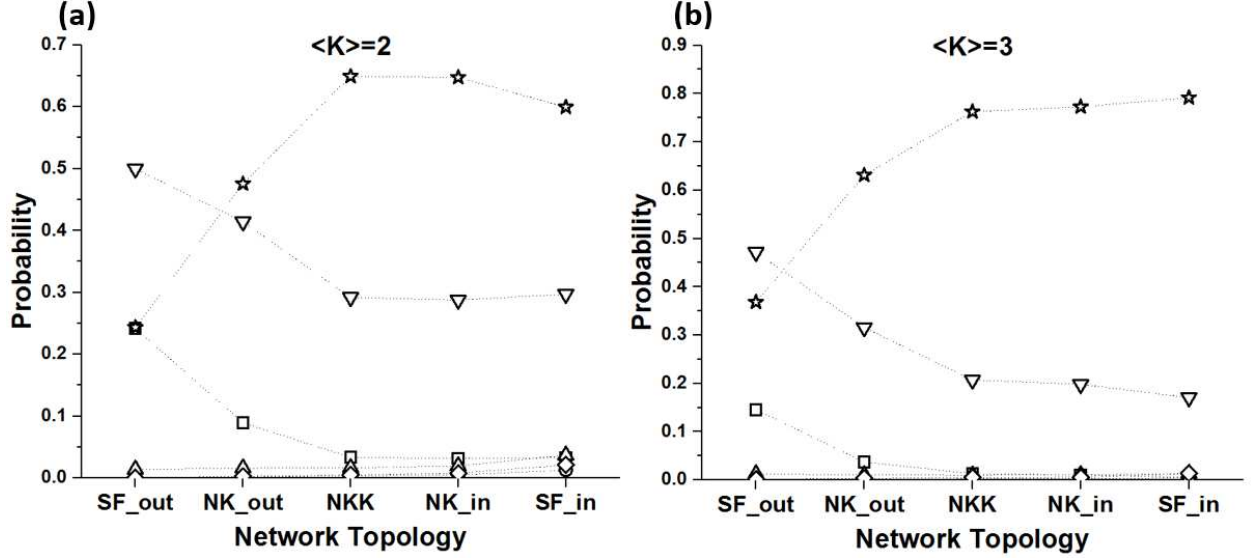
FIG. 2: Probability of each class of double node knockout across the five ensembles with different degree distributions. The left and right graph shows the result for networks with average degree $<K> = 2$ and $<K> = 3$ respectively. Classes 1 to 6 are drawn in square, circle, up triangle, down triangle, diamond, and star symbols respectively.

We find that in subclass b, that is, $S_A \cup S_B = S_{AB}$, the repair solution for the double node damage will always be a subset of the "direct product" of the single node damage repair solution. More rigorously, let $S_A = \{A_1, \cdots, A_m\}$ and $S_B = \{B_1, \cdots, B_n\}$. A repair solution after knocking out node A has the form $(r_{A_1}, \cdots, r_{A_m})$, where r_{A_1} represents a way to stabilize node A_1. Let R_{A_1} be the set containing all r_{A_1} that appears in all possible repair solutions. The set of all possible solutions after knocking out node A will be denoted as $G_A = \{(r_{A_1}, \cdots, r_{A_m}) : r_{A_i} \in R_{A_i}\}$. For the direct product of single node damage repair solution, $S_D = \{D_1, \cdots, D_p\}$, $p = |S_A \cup S_B|$; $R_{D_i} = R_{A_i} \cup R_{B_i}$ if $D_i \in S_A \cap S_B$ and $D_i = A_i = B_i$; $R_{D_i} = R_{A_i}$ if $D_i \in S_A \cap S_B^c$ and $D_i = A_i$; $R_{D_i} = R_{B_i}$ if $D_i \in S_A^c \cup S_B$ and $D_i = B_i$. Then the direct product of single node damage repair solution is given by $G_D = \{(r_{D_1}, \cdots, r_{D_m}) : r_{D_i} \in R_{D_i}\}$. This can be explained in the following way: since a node only has two states in a Boolean network, it will either be stable or will need repair. When the node needs repairing, damage to an additional node reduces the number of candidate nodes that can be used as starting points of the repair edges; nothing else should happen. Thus, the individual single node repair solutions are compatible with each other. Another observation is that $S_{AB} = S_A = S_B$ can only happen if A and B are regulating the same node(s). Otherwise, part of the damage, and thus also of the repair solutions, would be independent of each other.

2. Damage probability and class distribution in simulations of random network ensembles

Similarly to Sec. III A 2, we study the effect of degree heterogeneity on the resilience probability in a double

knockout setting. We also explore the distribution of the repair categories introduced in Sec. III B 1 using simulations of random Boolean networks. The computational details are similar as in Sec. III A 2 except we consider all possible pairs of knockouts to obtain the estimation for the damage probability and the classification of each category.

As shown in Fig. 1, the damage probability after double knockout is rather high regardless of the degree distribution and is higher than the damage probability after single knockout. As one can see, the double-knockout damage probability is higher in a network with higher average degree, which is consistent with the established conclusion that the complexity of the dynamics increases with larger average node in-degree [26–28]. The NKK model with K=1 is an exception; here the damage probability is 1 whether one or two nodes are damaged. This is because this network forms a single cycle. The only possible effective Boolean rules for K=1 are the identity (the output equals the input) and the negation. Thus knocking out any currently-ON node in the network will induce a change in its child node, which means the network will need to be repaired.

Based on the simulations, the damage probabilities of ensembles with fixed out-degree (K=2 and K=3) for double node knockout are rather close to each other; in-degree heterogeneity does not significantly change the damage probability. However, when we compare the three ensembles with fixed in-degree, out-degree heterogeneity leads to a decrease in the damage probability; this is because of the abundance of sink nodes. These results are similar to the results of the single node knockout.

To illustrate the distribution of the double damage classes introduced in Table II, in Fig. 2 we plot the prob-
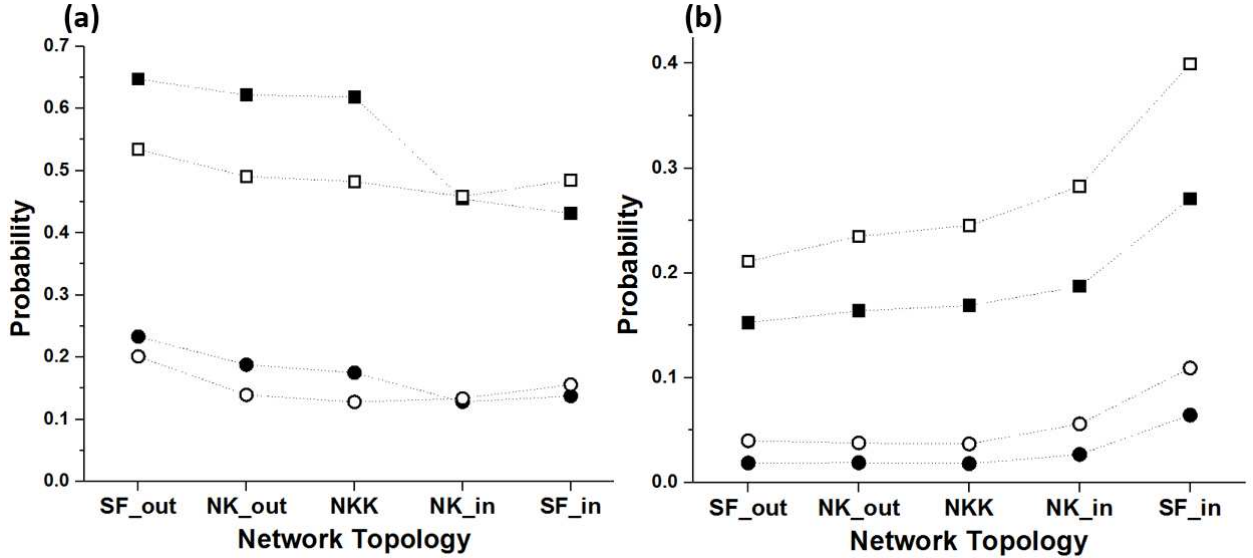
FIG. 3: The probability that one needs to repair more or different (circles) or fewer (squares) nodes for simultaneous damage of two nodes compared to the union of the repairs needed for individual damage to each of the nodes. The probability that one needs to repair the same nodes is 1 minus the sum of the two shown probabilities. The same network ensembles as in Fig. 1 and 2 are used. Solid symbols represent $<K> = 2$ results and empty symbols represent $<K> = 3$. (a) Simultaneous damage of two nodes that share a downstream target. (b) The general case of simultaneous damage of two nodes.

ability of each class in the five ensembles. Based on the simulations, class 2 (both single damage cases are stable, repair is needed for double damage), class 3 (repair is needed for one case of single damage, stable after double damage) and class 5 (repair is needed for each single damage, stable after double damage) have very low probability of occurrence. The reason is that the occurrence of these situations requires that the nodes being knocked out are regulating a common target. In contrast, most randomly chosen node pairs are independent. Class 6 (see Table II) tends to have the highest probability, followed by class 4 and class 1 ; the probability of these cases also varies more in the different ensembles. Comparing the three ensembles with a fixed out-degree (K=2 or K=3), the probability of each class is fairly close according to the simulation. Comparing the three ensembles with a fixed node in-degree, we can readily see that heterogeneity in node out-degree leads to a smaller probability for class 6 (stars) and larger probability for classes 4 (down triangles) and 1 (squares). This is related to the fact that heterogeneity in node out-degree leads to more sink nodes in the network.

As we are interested in combinatorial effects of double node knockout, we marginalize all the (sub)classes into three categories based on whether we need to repair more or different nodes (class 2, 4c, 4d, 6c, and 6d), the exact same set of nodes (class 1, 4b and 6b), or less nodes (classes 3, 5, 4a, and 6a) in case of double damage compared to the union of the two single damage cases. We estimate the probability of each category by simulation using the five ensembles. If the two nodes being knocked out do not share a target, the two dam-

age processes are independent and there will not be any combinatorial effect. It is therefore of particular interest to calculate the probability of each category in just the cases wherein the two nodes share a target (Fig. 3(a)), and compare with the general case (Fig. 3(b)). Since prior research shows that the average degree of biological networks is around 2 [33, 43], we focus on $<K> = 2$ and $<K> = 3$. According to both Fig. 3(a) and 3(b), the probability that we need to repair fewer nodes (subclass a) in double knockouts is larger than the probability that we need to repair more or different nodes (subclasses c and d). This is consistent with the fact that in Table III, there are more motifs corresponding to subclass a than to subclass c. For $<K> = 2$, compared with networks with constant in-degree (the left three ensembles in Fig. 3(a)), networks with constant out-degree and heterogeneous in-degree distribution (the right two ensembles in Fig. 3(a)) demonstrate a lower probability for cases wherein one needs to repair more or fewer nodes; for $<K> = 3$, the probabilities are close to each other. As the network topology changes from constant in-degree and scale-free out-degree distribution to constant in-degree and out-degree to constant out-degree and scale-free in-degree distribution (from left to right in Fig. 3(b)), the percentage of node pairs sharing a target node among all possible pairs increases. This change is more dramatic than the change in the probability of the three categories across different ensembles in Fig. 3(a). Thus as shown in Fig. 3(b), the probability of cases wherein one needs to repair more nodes (circles) or less nodes (squares) among all node pairs increases across the five ensembles.

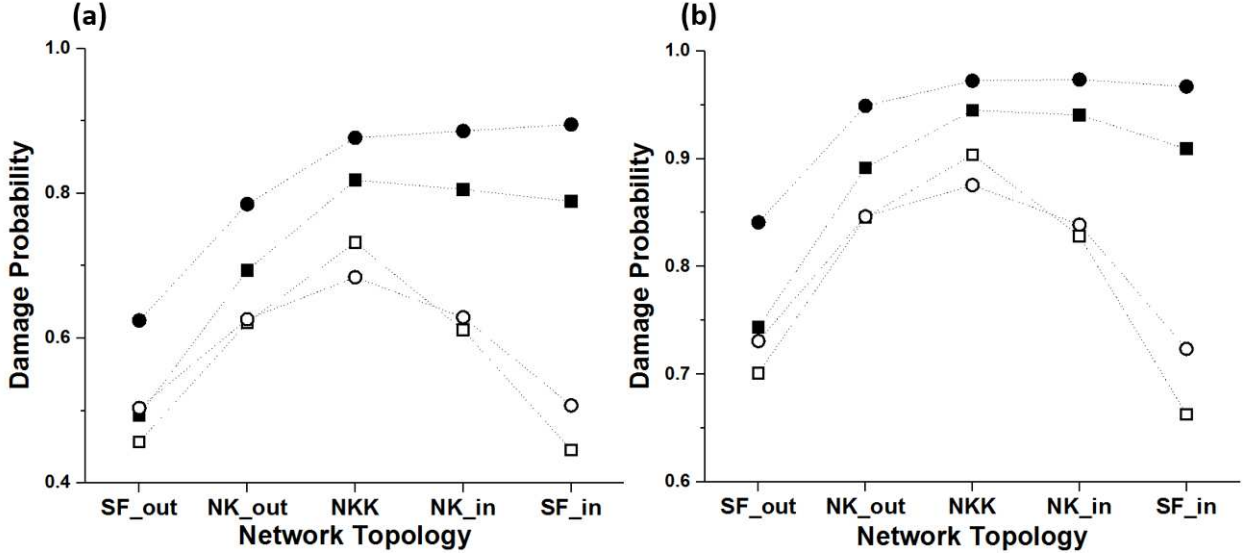As discussed in Sec. III A 1, using nested canalizing

FIG. 4: Damage probability in network ensembles using effective Boolean rules (solid symbols) or nested canalizing rules (empty symbols). Square symbols represent $<K> = 2$ and circular symbols represent $<K> = 3$. (a) Single node knockout; (b) double node knockout.

functions helps make a steady state more resilient to single node damage under the same network topology. This is confirmed by simulation results summarized in Fig. 4(a). The damage probability is smaller for networks with nested canalizing functions (empty symbols) for all five ensembles with $<K> = 2$ or $<K> = 3$. This conclusion holds for double node damage, as shown in Figure 4(b). As discussed in Sec. III A 1, out-degree heterogeneity leads to lower damage probability for both effective and canalizing functions (compare the first three ensembles). In the case of nested canalizing Boolean functions, in-degree heterogeneity also leads to a lower damage probability, in contrast with its minor effect in case of effective Boolean functions. This is because higher in-degree leads to more stability for nested canalizing functions, reflected in the fact that the stability probability $\alpha(I_j)$ of nested canalizing functions keeps increasing steadily and is much larger than that of effective Boolean functions for higher I_j (see Sec. III A 1).

C. Single node damage in networks with two steady states

1. General Discussion

Another follow-up direction is to explore the effect of single node damage on two different steady states of a network. The goal is to see whether a single solution can remedy the damage in multiple attractors (steady states here) at the same time. To classify all the situations of knockout damage to a single node, we observe that the damaged node may be normally (when undamaged) ON in both steady states, or ON in one steady state and OFF in the other. (We do not consider the situation that the node is OFF in both steady states, as the knockout damage will not change anything to either steady state). The

categorization of the constitutive expression type damage will be analogous.

If the node is ON in both steady states, the steady states can be both stable, both in need of repair or one is stable and the other needs repair. If the node has different states in the two steady states, only one of them needs repair, as summarized in Table IV.

We explore the probability distribution of the classification shown in Table IV in random Boolean networks. The computational details are similar to Sec. III A 2 except we consider all possible single node knockouts for every pair of steady states for a specific network. As there are 9 classes and each class may have a small probability, we marginalize class 2 and 3 (where the node is ON in both steady states before damage and one of the steady states needs repair after damage), class 4, 5 and 6 (where both need repair), class 8 and 9 (where node has different states before damage and one of the steady states needs repair after damage). As shown in Fig. 5, we found that the class in which both steady states need repair (up triangle) is less probable in heterogeneous networks. The class in which both steady states are stable (squares) is more probable in out-degree heterogeneous networks as sink nodes contribute to the resilience probability of steady state as in Sec. III A.

We are particularly interested in determining whether or not there are common repair solutions in the cases where both steady states need repair (classes 4, 5 and 6). We find from our simulations on network ensembles that not having common solutions is less probable (the fraction of classes 5 and 6 is between 5% and 7% for different ensembles with $<K> = 2$), thus we enumerate these situations. Similar to Sec. III B 1, we start by looking for three-node motifs that will lead to no common solutions. Reexamining the 10 motifs in Table III, and consider-
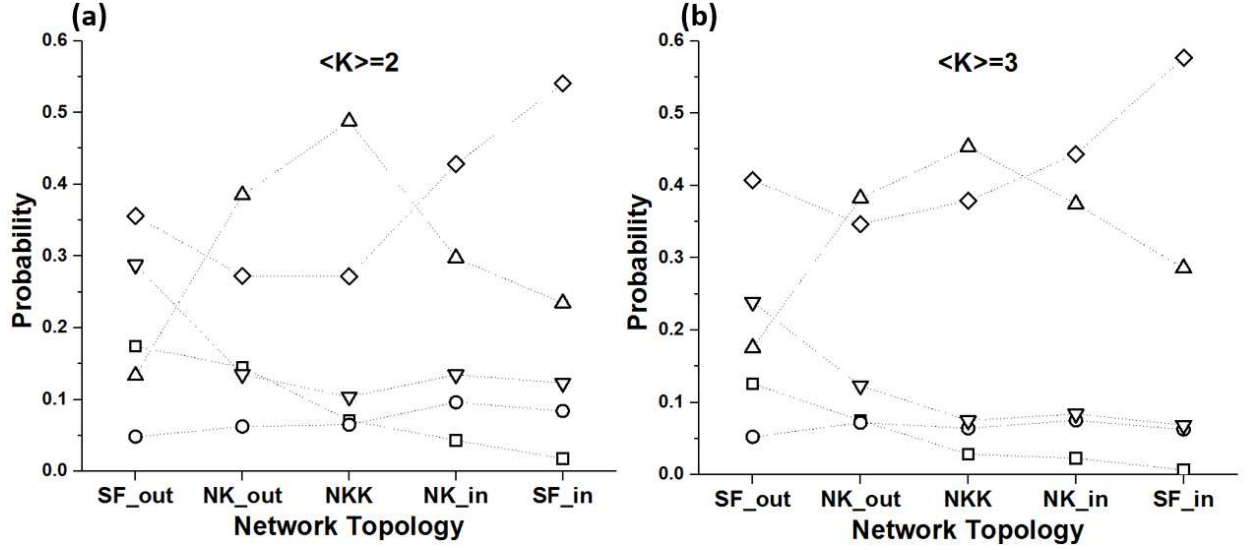
FIG. 5: Probability of outcomes of node knockout on pairs of steady states across five network ensembles with different degree distributions. The left and right graph shows the result for networks with average degree $<K> = 2$ and $<K> = 3$ respectively. The classes are grouped as 1 (squares), the union of 2 and 3 (circles), the union of 4, 5, 6 (up triangles), 7 (down triangle), and union of 8 and 9 (diamonds).

TABLE IV: Classification of the outcomes of single node knockout for two steady states of a network

Situation Index	Node state before damage in the two SSs	Status of the two SSs after node damage
1	ON in both	Both stable
2	ON in both	One SS is stable, the other SS needs repair, compatible solutions exist
3	ON in both	One SS is stable, the other SS needs repair, no compatible solutions
4	ON in both	Both SSs need repair, common solution(s) exist
5	ON in both	Both need repair, compatible (but not common) solution(s) exist
6	ON in both	Both need repair, no compatible solutions
7	ON, OFF	Both stable
8	ON, OFF	SS with node ON needs repair, compatible solutions exist
9	ON, OFF	SS with node ON needs repair, no compatible solutions

The table lists nine situations that will happen after knocking out a single node in the network that has two steady states (SSs). The second column indicates the state of the knocked-out node before damage. The third column specifies whether the damaged steady state is stable against the damage or needs repair. We use the term "common solution" for when we need to repair exactly the same set of nodes for the two steady states after single node knockout, we say "compatible solution" for when there exists a solution that can be used to stabilize both steady states. Thus all common solutions are compatible solutions.

ing pairs of possible steady states for these motifs, we find that only the XOR/XNOR motif will forbid a common solution for repairing the two steady states. The XOR/XNOR motif is rarely observed in biological networks, as they represent cases where each regulator can switch between being an activator or inhibitor depending on the state of the other regulator.

Another mechanism that will lead to no common solutions for repairing two steady states is that there is no valid candidate to use as a starting point of an additional edge. One such situation is that all the candidate nodes have different states in the two steady states, thus none of them can be used to realize the same function in the two steady states. This is exemplified in Fig. 6(a). Another situation is that the sensitive node is regulated by almost every node (other than the node itself) and there are no nodes left to be repair candidates since current regulators cannot be used. A combination of the two situations can also lead to no valid candidate for repair.

When the node is ON in one steady state and OFF in the other steady state, the damage will do nothing to the second steady state. However, the repair solution for the first steady state may or may not be compatible with the second steady state. Our simulations using random networks suggest that the incompatible situation is rarer. Incompatibility can arise in a lot of simple motifs of two or three nodes, including a single regulating edge (positive or negative), OR gate, AND gate, XOR gate, XNOR gate. The reason why this situation is rare in a real network is that if the network has nodes that have different states in the two steady states, any of these nodes can be used as starting points to an additional edge to node B. This additional edge will have an opposite effect in the two steady states and thus it can solve the incompatibility problem. It is rare, but still possible, that two
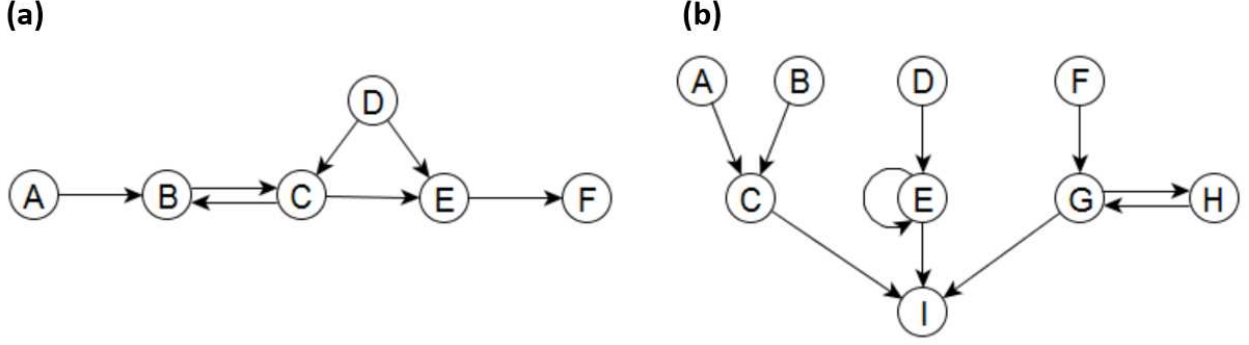
(a)

(b)

FIG. 6: (a) A simple network to illustrate the situation when there are no valid repair candidates. All the edges are positive, the updating rules are B = A AND C, C = B OR D, E = C AND D. There are two steady states, (1, 1, 1, 1, 1, 1) and (1, 1, 1, 0, 0, 0). If node A is knocked out, we need to repair node B. All candidate nodes (namely D, E and F) are in different states in the two steady states, thus no common solutions exist. (b) An example network to illustrate incompatibility in stabilizing two steady states at the same time because the two steady states only differ in the state of the knocked-out node and of the sensitive node. C = A OR B, E = D OR E, G = F OR H, I = A OR E OR G. First, we consider the effect of knockout of source node A on the steady state pair (1, 0, 1, 1, 1, 1, 1, 1, 1) and (0, 0, 0, 1, 1, 1, 1, 1, 1). When knocking out node A, we need to repair node C to be ON for the first steady state. However, fixing C to be ON will eliminate the other steady state (where C is OFF) as all the candidate nodes (B, D, E, F, G, H or I) have the same state in the two steady states; thus no compatible repair solutions exist. The incompatibility mechanism is the same for knockout of node E in case of the steady state pair (0, 0, 0, 0, 1, 0, 0, 0, 1) and (0, 0, 0, 0, 0, 0, 0, 0, 0), and the knockout of node G in case of the state pair (1, 1, 1, 1, 1, 0, 1, 1, 1) and (1, 1, 1, 1, 1, 0, 0, 0, 1).

steady states of a network have the same state for most of nodes and only differ in the state of the knocked-out node, the sensitive node, and possibly its current regulators. This can happen if the knocked-out node is part of a bistable motif connected with the rest of the network with a canalizing function such as an OR gate. Thus the bistable motif neither affects nor is affected by the rest of the network in a steady state. Examples of bistable motifs are a source node, a node with a self-loop and a two-node feedback loop (see Fig. 6(b)).

D. Biological Examples

In this section, we apply our algorithms to study the resilience of the T cell large granular lymphocyte (T-LGL) leukemia network [44] and the epithelial-to-mesenchymal transition (EMT) network [45].

T-LGL leukemia is a rare blood cancer. While normal T cells undergo activation induced cell death (apoptosis) after successfully fighting a virus, leukemic T-LGL cells survive. The network model constructed by Zhang *et al.* [44] includes the proteins involved in the activation of T cells, in activation induced cell death, as well as a number of proteins that were observed to be abnormally highly expressed or active in T-LGL cells. The model describes the regulation of each of these proteins with Boolean rules, and captures the normal (apoptosis) and leukemic (survival) states of the system [44]. The original network has 60 nodes, including three source nodes, and 142 regulatory edges. By fixing all the states of source (unregulated) nodes in the biologically relevant condition and iteratively replacing fixed node states in the Boolean rules, one can reduce the network to a smaller network,

whose nodes' states are not determined by the source nodes alone but rather by the specific dynamic trajectory of the system [46]. We perform additional network simplification as specified in Appendix A. The reduced network model (Fig. 7) has two steady states, namely a disease (T-LGL) state (0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1) and a normal T cell state committed to the path to apoptosis (1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0), where the nodes are in the alphabetic order, BID, CREB, Caspase, Ceramide, DISC, FLIP, Fas, GPCR, IAP, IFNG, MCL1, S1P, SMAD, sFas.

EMT is a cell fate change involved in embryonic development which can be reactivated during cancer metastasis [45]. During EMT, epithelial cells lose their original adhesive property, leave their primary site, invade neighboring tissue, and migrate to distant sites as mesenchymal cells. A Boolean network model of EMT in the context of hepatocellular carcinoma invasion has been established by Steinway *et al.* [47]. The EMT network has 70 nodes and 135 edges. Steinway *et al.* performed a network reduction to obtain a network with 19 nodes and 70 edges (Fig. 8). This type of network reduction has been shown to have no effect on the permitted dynamics and enables us to fully explore the state space [47]. In the reduced network, the adhesion factor E-cadherin is the sink node and its OFF state will indicate the transition to a mesenchymal state. The reduced network has two steady states, the epithelial state (0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1) and the mesenchymal state (1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0), written in the order of AKT, AXIN2, β-catenin_memb, β-catenin_nuc, Dest_compl, E-cadherin, GLI, GSK3β, MEK, NOTCH,
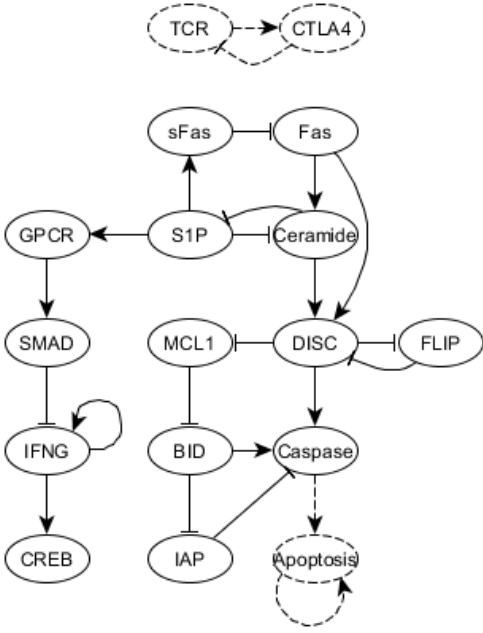
FIG. 7: Reduced T-LGL leukemia signaling network. An arrowhead or flat bar end indicates a positive or negative regulation edge respectively. The nodes and edges drawn in dashed lines are ignored in the analysis (see Appendix A).
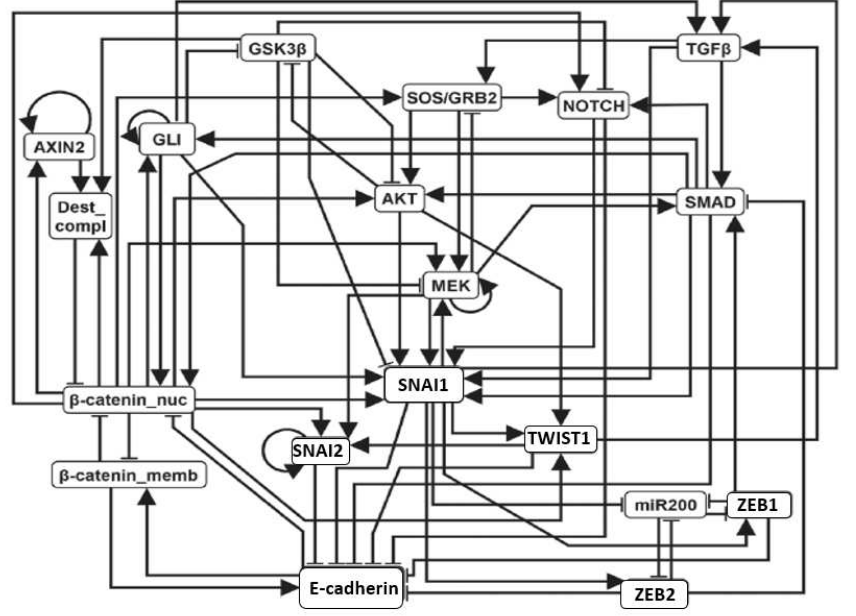

FIG. 8: Reduced EMT Network. Nodes represent proteins and miRNAs involved. An arrowhead or flat bar end indicates positive or negative regulation, respectively.

SMAD, SNAI1, SNAI2, SOS/GRB2, TGFβR, TWIST1, ZEB1, ZEB2, miR200.

The average degree of the T-LGL leukemia network is 1.43, the standard deviation is 0.65 for the in-degree and 0.94 for the out-degree. The average degree of the EMT network is 3.68, the standard deviation is 1.80 for the in-degree and 2.38 for the out-degree. To explore the relationship between the biological case studies and random ensembles, we randomize the biological networks to form an ensemble of 1000 networks for each. We consider two types of randomization, one that preserves the degree of each node (DPR) and one that additionally preserves the regulatory function of each node (DFPR). To preserve node degree, for each randomization, we exchange the child nodes of two randomly selected edges for $50 \times M$ times, where M is the total number of edges. Then we generate effective Boolean functions with the same bias as the original network (in DRP) or we just keep the original function (in DFPR).

As shown in Table V, the average damage probability of the two steady states of the T-LGL network is 0.667 for single node damage and 0.871 for double node damage. The average damage probability of the two steady states of the EMT network is lower, 0.316 for single node damage and 0.449 for double node damage. Table V also indicates that after either single node damage or double node damage, the damage probability of the T-LGL leukemia network does not statistically deviate from the randomized ensemble average. The damage probability of the EMT network deviates from the damage probability of the degree preserving randomized ensemble. The

main contributing factor is that the Boolean functions of the EMT network are all canalizing functions and a majority of them are also nested canalizing, which tends to decrease the damage probability, as discussed in the last paragraph of Sec. III A 1. This nested canalizing nature is destroyed when random functions are used, even if they have the same bias. Indeed, the damage probability of the DFPR ensemble is much closer to the result of the EMT network. Though the Boolean functions used in T-LGL leukemia networks are also mostly nested canalizing, the T-LGL network has an average node degree smaller than 2 and most two-input effective functions are also nested canalizing functions, so the randomized functions would be similarly canalizing.

As shown in Table VI, the probability distribution of each class of double node knockout for the T-LGL network is close to the ensemble average in that class 6 and class 4 are the most well represented. In contrast, the probability distribution of the double node knockout classes for the EMT network is different from the ensemble average in that class 1 is more represented and class 6 is less represented, which is consistent with the deviation in damage probability between the EMT network and its randomized ensemble. In both cases, the result of the biological network is closer to the result of the DFPR ensemble than the result of the DPR ensemble, as expected. We include the detailed classification of double knockout pairs in the two biological networks in Appendix B .

We also apply our algorithm of stabilizing two steady states simultaneously after a single node damage to both networks. While one generally wishes to eliminate rather

TABLE V: Comparison between the damage probability of the two biological networks and their randomizations

network\ensemble	DP	DPFR DP, std	DPR DP, std
T-LGL SD	0.667	0.674, 0.115	0.725, 0.135
T-LGL DD	0.871	0.875, 0.078	0.905, 0.092
EMT SD	0.316	0.449, 0.122	0.792, 0.096
EMT DD	0.473	0.689, 0.122	0.943, 0.041

The 2nd column is the damage probability (DP) of the T-LGL leukemia network (first two rows) and the EMT network (last two rows) after single node damage (SD) or double node damage (DD). The 3rd and 4th columns report the average damage probability and its standard deviation (std) among networks for the corresponding randomized network ensembles.

TABLE VI: Class probability distribution after double node knockout in the two biological networks

Ensemble \ Class	1	4	6
T-LGL	0.097	0.484	0.387
T-LGL DFPR	0.095	0.444	0.420
T-LGL DPR	0.076	0.397	0.495
EMT	0.505	0.398	0.065
EMT DFPR	0.274	0.479	0.202
EMT DPR	0.041	0.326	0.609

The first and fourth row show the results of the real biological networks, while the rest are the results of their DFPR and DPR ensembles. We only list the three classes (columns) from Table II that have the highest probability. The probability of the rest of the classes is smaller than 0.04 and does not show much difference between the real network and its randomization.

than repair a disease state in a biological network, these networks nonetheless provide a useful framework for applying our methodology; after considering joint repair, we will turn our attention to removing the disease state. Most of the nodes have opposite states in the two steady states of both networks. This is not surprising since the two steady states correspond to two opposite biological outcomes in each case (apoptosis versus survival in the T-LGL network; epithelial versus mesenchymal state in the EMT network). Furthermore, the network reduction used in both cases eliminates nodes that are fixed by source nodes and have the same state in both steady states [45, 46].

The only two nodes having the same state in the T-LGL leukemia network are CREB and IFNG, which exist in a sink branch of the network and do not directly determine the cell state. Thus when we consider the simultaneous repair of the two steady states, there will be 12 cases wherein the damaged node is ON in one state and OFF in the other. Among them, 4 cases (Caspase, FLIP, IAP, or SMAD knockout) fall into class 7 (see Table IV). Directly damaging the node Caspase may be not biologically interesting as we treat this node to be the sink node of the signaling network here. All the other 8 situations (BID, Ceramide, DISC, Fas, GPCR, MCL1, S1P, sFas) fall into class 8. The class distribution of the classifications of the DFPR ensemble concentrates on class 8, class 7 and class 4, and the result of the DPR ensemble concentrates on class 8, class 4 and class 7, both in decreasing probability. The class distribution of the randomized ensembles is consistent with, but with higher spread than the class distribution of T-LGL leukemia network, which is restricted to class 8 and class 7. The latter is mainly due to the fact that the two steady states are almost exactly opposite.

When stabilizing the two steady states of the EMT network simultaneously after a single node damage, there are 17 nodes whose knockout can be considered (Dest_compl and SOS/GRB2 are OFF in both steady states). Among them, two cases (AXIN2, SNAI2 knockout) belong to class 1 (see Table IV); nine cases

(AKT, β-catenin_nuc, GLI, NOTCH, SMAD, TGFβR, TWIST1, ZEB2, miR2000) belong to class 7 ; six cases (β-catenin_memb, E-cadherin, GSK3β, MEK, SNAI1, ZEB1) belong to class 8. As an example, let us consider permanently knocking out node GSK3β, which is ON in the epithelial steady state and OFF in the mesenchymal steady state. One needs to repair node AKT, MEK, SNAI1, NOTCH and there are 14, 14, 10, 14 simple repair choices for each corresponding node (see Appendix C). As most of the nodes have opposite states in the two steady states, the majority of the repair solutions will be compatible with the other steady state. The algorithm then calculates that there will be 11, 11, 6, 11 repair choices for each corresponding node. The specific choices are listed in Appendix C.

The class distribution of the DFPR ensemble concentrates on class 1, class 8 and class 4 and the result of the DPR ensemble concentrates on class 8, class 5 and class 4, both in decreasing probability. The distributions of the randomized ensembles are more uniformly spread compared with the distribution of the classifications of the EMT network, which is restricted to class 8, class 7 and class 1. This restriction is largely due to the fact that the two steady states are almost exactly opposite.

In summary, we compared the biological networks with their randomized ensembles, focusing on their damage probability, class distribution after double node knockout and class distribution of repairing two steady states after single node knockout. The T-LGL leukemia network agrees with its randomized ensembles and the EMT network deviates from its degree-preserving randomized ensemble in all tested aspects.

In both networks, a more biologically meaningful intervention than preserving both steady states is to keep the normal steady state as intact as possible and destabilize the disease steady state. If the node to be repaired has opposite states in the two steady states, adding a new edge starting from a node that has the same state in the two steady states will destabilize the disease state.

For example, if Fas is knocked out in the T-LGL network, we need to repair Ceramide to be ON to avert

cascading damage to the healthy steady state. The algorithm will give 9 repair solutions involving a new independent edge, shown in Appendix D. Two edge repair solutions ("Ceramide= $\cdots$ OR NOT IFNG" and "Ceramide= $\cdots$ OR NOT CREB", where $\cdots$ stands for the original rule for Ceramide) are not compatible with the second steady state since these two nodes have the same node state in the two steady states. Thus either of these two repair strategies will make the disease state a transient state and the system will keep evolving. Whether the system evolves toward the healthy steady state depends on the node knocked out and the repair solution. In the example above, if we knock out Fas and fix Ceramide to be ON by adding an edge from CREB, this will make the system evolve towards another steady state wherein Caspase is ON, a state biologically similar to the healthy steady state.

Similarly to the T-LGL leukemia network, a repair solution using nodes with the same node state in the two steady states can preserve the epithelial steady state and perturb the mesenchymal steady state of the EMT network. However, in some cases the new attractor is not an epithelial one (E-cadherin is not guaranteed to be ON).

IV. DISCUSSION AND CONCLUSION

One promising approach to mitigating the effects of diseases is to proactively manipulate the interactions in the relevant biological network. For example, cancerous cells fail to undergo natural cell death; compensatory interactions in the cancer signaling network may in principle drive cancerous cells to undergo cell death. While a theoretical basis for such manipulation has been established in the case of deregulation of a single node (e.g. a single genetic mutation) [10], complex diseases are triggered by several co-existing gene mutations [9, 14, 15]. The algorithm presented here can be used to design preventive interventions for combinations of multiple dysfunctions of the network. Our identified repair strategy classes provide a framework to explore the short-term combinatorial effects of double knockouts and can be straightforwardly adapted to other types of multiple perturbations.

The network ensembles most considered here exist in the chaotic phase for very large networks according to the well-studied annealed approximation (due to the average in-degree of 2 or 3), where the topology and update functions are randomized after each time step [25–28]. Thus, we expect the effects of network perturbations to propagate throughout the network. However, to gain detailed insight into the dynamic behavior of the network and to determine specific repair strategies, it is necessary to consider a fixed network topology and interaction rules. We therefore consider two specific biological case studies in this report.

As patients are often diagnosed with complex diseases after symptoms already developed, the cascading effect of the initial gene mutation or protein dysfunction is already in progress. Thus it is interesting to consider the long-term effects of damage when aiming to repair the effects of single or multiple dysregulations. One can define a node's region of influence as the nodes whose states will be changed due to the cascading effect of its perturbation. Similar to what we have done in the short-term setting, if the regions of influence of two nodes do not intersect and are not co-regulating another target, then the two damage processes are independent of each other, and one would expect to be able to mitigate their effects independently. If the regions of influence of two initially damaged nodes intersect or co-regulate a third node, combinatorial effects will appear and can be analyzed in a similar way as we did here.

In some cases two or more steady states with distinct biological meanings, such as natural cell death and cancerous persistence, may exist [9, 48]. As demonstrated in two biological case studies, our algorithm provides strategies to find compatible ways to stabilize two steady states or stabilize one and destabilize the other. The approach we take here is most useful in designing preventive interventions for disease, as the repair is assumed to be effective on a faster timescale than the propagation of damage. Model-based design of therapeutic methods for complex diseases entails an understanding of the disease state and the identification of manipulations that drive the system from the disease state back to a normal state [49]. As a first step, our method provides choices to destabilize the disease state and a framework to test the feasibility of simple edge modifications. A systemic study of the trajectories from a destabilized disease state into a normal state would be another interesting area for future work.

Appendix A: Additional simplification of T-LGL leukemia network

When the sink node Apoptosis is activated, the cell is going to die. Zhang *et al.* chose to represent cell death by a state in which Apoptosis is ON and all the other nodes are OFF and implemented it by adding to every node's Boolean function the clause "AND (NOT Apoptosis)" [44]. Here for simplicity we do not use this additional clause; this is equivalent with considering any steady state that includes Apoptosis=ON as a normal steady state. In the reduced network (Fig. 7) a small motif consisting of TCR and CTLA4 is isolated from the main part of the network. Since the small motif does not influence the apoptotic decision, we ignore it in the analysis. An auxiliary node P2 in [44] is removed and we incorporate the effect in the Boolean rule of IFNG. Also, if the cell is already dead, node knockout and constitutive expression have no biological meaning. However, the activation of Apoptosis requires the node Caspase to be ON first. Thus, we delete the node Apoptosis and consider that Caspase is determining the state of the cell.

Appendix B: Classification of double knockout pairs in the T-LGL leukemia network and EMT network

We apply the algorithm to stabilize a steady state after double node damage to the T-LGL leukemia network.

TABLE VII: Classification of double knockout pairs in the T-LGL leukemia network

Node pair	S_A	S_B	S_{AB}	class
BID,Caspase	IAP	∅	IAP	4b
BID,Ceramide	IAP	S1P	IAP,S1P	6b
BID,DISC	IAP	MCL1,FLIP	IAP,Caspase, MCL1,FLIP	6c
BID,Fas	IAP	Ceramide	IAP,Ceramide	6b
Caspase,Ceramide	∅	S1P	S1P	4b
Caspase,DISC	∅	MCL1,FLIP	MCL1,FLIP	4b
Caspase,Fas	∅	Ceramide	Ceramide	4b
Ceramide,DISC	S1P	MCL1,FLIP	MCL1,S1P,FLIP	6b
Ceramide,Fas	S1P	Ceramide	S1P,DISC,	6d
DISC,Fas	MCL1, FLIP	Ceramide	MCL1,FLIP, Ceramide,	6b

All ten node pairs (A, B) are list in the first column. The node to be repaired after knocking out A, B, both A and B are listed in the second to fourth column respectively, where ∅ means that no node need to be repaired. The class and subclass index as in Table 1 is listed in the fifth column.

There are five nodes ON in the healthy steady state, and thus there are 10 double knockout cases (see Table VII). Among them, four cases belong to class 4b (see Table II). Six cases belong to class 6: four in 6b and one each in 6c and 6d . Thus in this example, there are no cases where less repair is needed, and there are two cases where combinatorial effect occurs. For the disease steady state, there are 7 nodes in the ON state and thus 21 double knockout cases. Among them one case belong to class 1, ten cases belong to class 4b and ten cases belong to class 6: three in 6a and seven in 6b.

We also apply the algorithm to stabilize a steady state after double node damage to the EMT network. For the epithelial steady state, there are 6 nodes in the ON state and thus there are 15 double knockout cases. Among them, three cases belong to class 1, nine cases belong to class 4b, two cases belong to class 6b and one case to class 6d. For the mesenchymal steady state there are 13 nodes with ON states and thus 78 node pairs. Among them, 44 cases belong to class 1, 1 case belongs to class 2, 2 cases belong to class 3, 28 cases belong to class 4 (one in 4a, twenty-six in 4b, one in 4c), 3 cases belong to class 6 (one in 6a and two in 6b). Thus there are four cases where less repair is needed and two cases where more repair is needed for double knockout compared to the union of two individual single knockouts.

Appendix C: Simple solutions after knocking out GSK3β in the healthy steady state of EMT network

All the solutions have similar format as above, where ⋯ stands for the original rule for that node. Solutions in normal text format are simple solutions compatible with disease steady state after knockout GSK3β in the healthy steady state, i.e., solutions in *italics* are simple solutions incompatible with disease steady state after knockout GSK3β in the healthy steady state.

Modifications for node AKT:
AKT=⋯ AND GLI
AKT=⋯ AND MEK
AKT=⋯ AND NOTCH
AKT=⋯ AND SNAI1
AKT=⋯ AND TGFβR
AKT=⋯ AND TWIST1
AKT=⋯ AND ZEB1
AKT=⋯ AND ZEB2
AKT=⋯ AND NOT β-catenin_memb
AKT=⋯ AND NOT E-cadherin
AKT=⋯ AND NOT miR200
AKT = ⋯ AND Dest_compl
AKT = ⋯ AND NOT AXIN2
AKT = ⋯ AND NOT SNAI2

Modifications for node MEK:
MEK= ⋯ AND AKT
MEK= ⋯ AND GLI
MEK= ⋯ AND NOTCH
MEK= ⋯ AND SMAD
MEK= ⋯ AND TGFβR
MEK= ⋯ AND TWIST1
MEK= ⋯ AND ZEB1
MEK= ⋯ AND ZEB2
MEK= ⋯ AND NOT β-catenin_memb
MEK= ⋯ AND NOT E-cadherin
MEK= ⋯ AND NOT miR200
MEK = ⋯ AND Dest_compl
MEK = ⋯ AND NOT AXIN2
MEK = ⋯ AND NOT SNAI2

Modifications for node SNAI1:
SNAI1= ⋯ AND TWIST1
SNAI1= ⋯ AND ZEB1
SNAI1= ⋯ AND ZEB2
SNAI1= ⋯ AND NOT β-catenin_memb
SNAI1= ⋯ AND NOT E-cadherin
SNAI1= ⋯ AND NOT miR200
SNAI1 = ⋯ AND Dest_compl
SNAI1 = ⋯ AND SOS/GRB2
SNAI1 = ⋯ AND NOT AXIN2
SNAI1 = ⋯ AND NOT SNAI2

Modifications for node NOTCH:
NOTCH= ⋯ AND AKT
NOTCH= ⋯ AND GLI
NOTCH= ⋯ AND MEK
NOTCH= ⋯ AND SNAI1
NOTCH= ⋯ AND TGFβR
NOTCH= ⋯ AND TWIST1
NOTCH= ⋯ AND ZEB1
NOTCH= ⋯ AND ZEB2
NOTCH= ⋯ AND NOT β-catenin_memb
NOTCH= ⋯ AND NOT E-cadherin
NOTCH= ⋯ AND NOT miR200
NOTCH = ⋯ AND Dest_compl
NOTCH = ⋯ AND NOT AXIN2
NOTCH = ⋯ AND NOT SNAI2

Appendix D: Modifications for node Ceramide after knockout of Fas in the T-LGL leukemia network:

All the solution have the same format: "Ceramide = $\cdots$ OR New Rule", where $\cdots$ stands for the original rule.

Ceramide=$\cdots$ OR BID
Ceramide=$\cdots$ OR Caspase
Ceramide=$\cdots$ OR DISC
Ceramide=$\cdots$ OR NOT FLIP
Ceramide=$\cdots$ OR NOT GPCR
Ceramide=$\cdots$ OR NOT IAP
Ceramide=$\cdots$ OR NOT MCL1
Ceramide=$\cdots$ OR NOT SMAD
Ceramide=$\cdots$ OR NOT sFas

[1] B. Palsson, *Systems Biology: Properties of Reconstructed Networks* (2006).

[2] S. N. Steinway, J. G. Zañudo, W. Ding, C. B. Rountree, D. J. Feith, T. P. Loughran, and R. Albert, Cancer Research **74**, 5963 (2014).

[3] Z. Sun, X. Jin, R. Albert, and S. M. Assmann, PLoS Computational Biology **10**, e1003930 (2014).

[4] G. Michal and A. H. Brian, Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1804**, 476 (2010).

[5] H. Ikushima and K. Miyazono, Nature Reviews Cancer **10**, 415 (2010).

[6] G. Karlebach and R. Shamir, Nature Reviews Molecular Cell Biology **9**, 770 (2008).

[7] R.-S. Wang, A. Saadatpour, and R. Albert, Physical biology **9**, 055001 (2012).

[8] N. Miskov-Zivanov, M. S. Turner, L. P. Kane, P. A. Morel, and J. R. Faeder, Science signaling **6**, ra97 (2013).

[9] P. Creixell, E. M. Schoof, J. T. Erler, and R. Linding, Nature biotechnology **30**, 842 (2012).

[10] C. Campbell and R. Albert, BMC systems biology **8**, 53 (2014).

[11] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, Nature **473**, 167 (2011).

[12] A. Mochizuki, B. Fiedler, G. Kurosawa, and D. Saito, Journal of Theoretical Biology **335**, 130 (2013).

[13] D. K. Wells, W. L. Kath, and A. E. Motter, Phys. Rev. X **5**, 031036 (2015).

[14] D. Hanahan and R. A. Weinberg, Cell **144**, 646 (2011).

[15] T. Geppert and H. Koeppen, Drug Development Research **75**, 271 (2014).

[16] M. E. J. Newman, *Networks: an introduction* (2010).

[17] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[18] S. Bornholdt, Journal of The Royal Society Interface **5**, S85 (2008).

[19] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, Nature **407**, 651 (2000).

[20] J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, *et al.*, Nature **437**, 1173 (2005).

[21] M. Chaves, R. Albert, and E. D. Sontag, Journal of Theoretical Biology **235**, 431 (2005).

[22] J. A. Papin, T. Hunter, B. O. Palsson, and S. Subramaniam, Nature Reviews Molecular Cell Biology **6**, 99 (2005).

[23] K. Klemm and S. Bornholdt, Phys. Rev. E **72**, 055101 (2005).

[24] R. Kinney, P. Crucitti, R. Albert, and V. Latora, The European Physical Journal B **46**, 101 (2005).

[25] M. Aldana, S. Coppersmith, and L. P. Kadanoff (2003) Chap. Boolean Dynamics with Random Couplings, pp. 23–89.

[26] B. Derrida and Y. Pomeau, Europhysics Letters **1**, 45 (1986).

[27] M. Aldana and P. Cluzel, Proc. Natl. Acad. Sci. USA **100**, 8710 (2003).

[28] M. Aldana, Physica D **185**, 45 (2003).

[29] A. Pomerance, E. Ott, M. Girvan, and W. Losert, Proc. Natl. Acad. Sci. USA (2009).

[30] B. Luque and R. V. Solé, Phys. Rev. E **55**, 257 (1997).

[31] S. Li, C. M. Armstrong, N. Bertin, H. Ge, S. Milstein, M. Boxem, P.-O. Vidalain, J.-D. J. Han, A. Chesneau, T. Hao, *et al.*, Science **303**, 540 (2004).

[32] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, Nature Reviews Genetics **12**, 56 (2011).

[33] I. Shmulevich and S. A. Kauffman, Phys. Rev. Lett. **93**, 048701 (2004).

[34] H. Ebadi and K. Klemm, Phys. Rev. E **90**, 022815 (2014).

[35] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, Proc. Natl. Acad. Sci. USA **100**, 14796 (2003).

[36] Y. Li, J. O. Adeyeye, D. Murrugarra, B. Aguilar, and R. Laubenbacher, Theoretical Computer Science **481**, 24 (2013).

[37] P. T. Peixoto, The European Physical Journal B **78**, 187 (2010).

[38] R. Albert and H. G. Othmer, Journal of Theoretical Biology **223**, 1 (2003).

[39] I. Hartman, John L., B. Garvik, and L. Hartwell, Science **291**, 1001 (2001).

[40] D. A. Chan and A. J. Giaccia, Nat Rev Drug Discov **10**, 351 (2011).

[41] G. Boldhaus, F. Greil, and K. Klemm, Theory in Biosciences **132**, 17 (2013).

[42] A. E. Motter, N. Gulbahce, E. Almaas, and A.-L. Barabási, Molecular Systems Biology **4** (2008).

[43] J. E. S. Socolar and S. A. Kauffman, Phys. Rev. Lett. **90**, 068702 (2003).

[44] R. Zhang, M. V. Shah, J. Yang, S. B. Nyland, X. Liu, J. K. Yun, R. Albert, and T. P. Loughran, Proc. Natl. Acad. Sci. USA **105**, 16308 (2008).

[45] Y. Nakaya and G. Sheng, Development, Growth and Differentiation **50**, 755 (2008).

[46] A. Saadatpour, R.-S. Wang, A. Liao, X. Liu, T. P. Loughran, I. Albert, and R. Albert, PLoS Computational Biology **7**, e1002267 (2011).

[47] S. N. Steinway, J. G. Zañudo, W. Ding, C. B. Rountree, D. J. Feith, T. P. Loughran, and R. Albert, Cancer research **74**, 5963 (2014).

[48] A. Naldi, J. Carneiro, C. Chaouiya, and D. Thieffry, PLoS computational biology **6**, e1000912 (2010).

[49] J. G. T. Zañudo and R. Albert, PLoS Computational Biology **11** (2015).