Seeing the net for the nodes: Coarse-graining modular Boolean networks

Harold Fellermann*
Institute for Physics and Chemistry, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark and ICREA-Complex Systems Lab, Universitat Pompeu Fabra (GRIB), Dr Aiguader 80, 08003 Barcelona, Spain

Maria Davidich†
Institute for Theoretical Physics, University of Bremen, Otto-Hahn-Allee, NW1, D-28359 Bremen Germany

We propose a novel method to reduce the dimensionality of modular Boolean networks, in which an original network is represented by a coarse-grained representation. We analyzed the performance of the method by simulations of randomly wired modular Kauffman and activator-inhibitor networks with respect to reliability, determinism, and efficiency. It is shown that the reliability of the method depends on the separation of time scales between intermodular and intermodular dynamics, and that this requirement is more likely fulfilled for activator-inhibitor networks than for Kauffman networks. If the requirement is fulfilled, the proposed coarse-graining method produces correct and mainly deterministic projections with high dimensional reduction.

Keywords: modular Boolean networks, activator-inhibitor networks, coarse-graining

I. INTRODUCTION

When building predictive models in systems biology one has to deal with complex networks consisting of numerous components such as proteins, amino acids, and polysaccharides whose interplay determines the functioning of living organisms. Recent development in automated high throughput techniques has significantly increased the knowledge of biomolecular interactions which is now available through vast databases for gene regulation, protein interactions, etc. In order to be not lost in the complexity of the acquired details, it is desirable to find a way to reduce the amount of information without loosing characteristic properties of the system as a whole. For many questions, the identification of such system-level features is even preferred to the knowledge of its precise molecular details.

Biological networks are known to be modular [1–3]: each module is an entity of several components that performs an identifiable task, different from the functions of other modules. Most molecules that perform one function mainly interact with molecules that are involved in the same task, whereas some components participate in several tasks or link one task to another, as protein p53, for example, which is involved in the cell cycle control process and triggers apoptosis in case when cell damage is unrepairable. Recent research has focused on methods for detecting and analyzing community structure and modularity in networks (for a recent review of existing approaches see [4] and references therein). The dynamical properties of modular networks, however, have received relatively little attention – despite the potential importance of the subject. More recent work has addressed the role of the modularity on dynamics of cascading failures in scale-free networks [5], and synchronization patterns of networked oscillators [6]. In particular, Ref. [6] demonstrated that the modularity of networks is strongly reflected in their synchronization dynamics, so that communities emerge as connected groups of synchronized oscillators.

It is clearly easier to determine the properties of one specific module then the complex properties of an organism as a whole. Therefore, the question arises whether one can use the property of modularity to reduce the complexity in the description of biological processes and, in particular, whether it is possible to reduce the global dynamics of the system to dynamics of interacting modules.

To answer this question, we use a Boolean networks approach which we consider appropriate because of its simplicity. Boolean networks use a minimum of information about biochemical reactions: all components are represented by nodes, each of them being in one of two states – ON (1) or OFF (0). Interactions between components are encoded in Boolean functions that map the presence or absence of a set of components onto their future presence or absence – knowledge of the exact concentrations of components is not needed. Boolean networks have been originally proposed by Kauffman more than 40 years ago [7]. While Kauffman's original approach used arbitrary Boolean functions, later models restricted the set of possible interactions to a biological plausible subset, namely interactions by activation and inhibition. In spite of its minimalist character, the method was successfully used in modeling different biological processes in a variety of organisms [8–14].

The notion of modules in the context of Boolean networks can be reformulated in the following way: a module is an entity within which the internal nodes are connected with higher density than nodes among different modules. The connections between one particular module

*Electronic address: harold.fellermann@upf.edu
†Electronic address: davidich@itp.uni-bremen.de

and other modules can be interpreted as external perturbations. Owing to this definition, one has the following reformulated problem: To reduce the global dynamics of Boolean network to dynamics of interacting modules.

In the current paper, we present and test a new method to reduce the dimensionality of modular Boolean networks. Starting from a network with modular topology, the procedure constructs a representative network, where each original module is represented by a single node with links reflecting the connectivity of the original modules. In the next step, we define states and transition rules for this topology such that the global attractors of the original networks are preserved.

The remainder of this work is organized as follows: section II describes the coarse-graining method, which is analyzed with respect to its ability to generate reliable, useful, and efficient results (sections III A through III C) by coarse-graining randomly generated modular Boolean networks with activator-inhibitor or Kauffman dynamics. Finally, we discuss the obtained results and limits of the proposed method.

II. COARSE-GRAINING METHOD

We denote a Boolean network with the tuple $(\mathcal{N}, A, S, f)$, where $\mathcal{N} = \{n_1, \ldots, n_N\}$ is the set of nodes, and $A \in \mathcal{L}^{N\times N}$ the adjacency matrix. $\mathcal{L}$ is the set of link types which is either $\{0, 1\}$ for Kauffman networks or $\{-1, 0, 1\}$ for activator-inhibitor networks. At each time step, a node n_i is in state $s_i(t) \in S = \{0, 1\}$ and its dynamics is defined by a transition function f_i which maps the input configuration of node n_i onto the future node state $s_i(t+1)$. In this work, we consider only synchronous networks where all nodes are updated at the same time. For Kauffman networks, f_i is a random Boolean function. For activator-inhibitor networks the transition function is $s_i(t+1) = \Theta\left(\sum_j A_{ij} s_j(t)\right)$ where Θ is the Heaviside function. Furthermore, let k_i denote the indegree of node n_i, and J_i an index set, such that $j \in J_i \Leftrightarrow A_{ij} \neq 0$. Finally, we employ bars to denote the cardinality (number of elements) of a set, e.g $|\mathcal{N}| = N$.

The attractor landscape of a Boolean network is a graph with 2^N nodes, each one representing a distinct network state $s = (s_1, \ldots, s_N)$. Nodes $s^{(1)}$ and $s^{(2)}$ are connected by a directed link $s^{(1)} \to s^{(2)}$ if $f(s^{(1)}) = s^{(2)}$, i.e. if the Boolean network moves from state $s^{(1)}$ to state $s^{(2)}$. Paths in the attractor landscape correspond trajectories of the network. Since the dynamics are deterministic, each node in the attractor landscape has outdegree 1. Recursively removing all nodes from the attractor landscape that have indegree 0 yields the set of non-transient states.

A coarse-graining procedure can be defined as a projection

$$\mathcal{P} : (\mathcal{N}, A, S, f) \longrightarrow (\mathcal{N}^*, A^*, S^*, f^*) \tag{1}$$

with $N^* < N$, i.e. the dimensionality of the network is reduced. We are interested in projections that preserve the original global dynamics (attractors) of $(\mathcal{N}, A, S, f)$.

In general, dimensional reduction procedures draw from a specific structure in the dynamics of the systems under consideration (see [15] and references therein): when the state space of a system is coarse-grained, the system will reside longer in each coarse-grained state, such that spatial and temporal coarse-graining necessarily come as a pair. Reduction techniques start from the assumption that the dynamics of the system can be decomposed into slow and fast components in a way that the system exhibits fast motion below the spatial resolution of the coarse-graining, while moving comparably slow on the length scale preserved by the reduction process. The performance of the procedures is usually better, the more apparent the time scales are separated [16].

In the context of modular Boolean networks, time scale separation means that information exchange between modules is slower than information processing within each module. More technical, this implies that a module can relax to an attractor after each change in its inputs from other modules. This approach is similar in spirit to perturbation analysis of continuous dynamical systems, where the motion of the system along the fast manifold is assumed to happen instantaneously.

Depending on the coupling between the fast and slow dynamics, reduction techniques allow to either a) completely neglect fast degrees of freedom if they have no influence on the slow dynamics, b) subsume their influence on slow dynamics by its time average, or c) replace the actual temporal progression of fast degrees of freedom by their statistical distribution, thereby approximating an originally deterministic system by a stochastic process. As it will turn out, the coarse-graining method proposed in this work will belong to this class of non-deterministic or stochastic projections.

A schematic of the proposed procedure is shown in Fig. 1.

A. Coarse-grained topology

In this work, we assume that the partition of nodes into modules is given as a partition $\mathcal{M} = \{m_0, \ldots, m_M\}$ of $\mathcal{N}$, where the m_i's are non-overlapping subsets of $\mathcal{N}$. The question on how modules can be detected in a given network has been extensively studied [4, 17–19] and is not the subject of this article.

Each module of the original network is represented by a single node in the coarse-grained network. Therefore we can use $\mathcal{M}$ as the node set of the coarse grained network:

$$\mathcal{N}^* = \mathcal{M} \tag{2}$$

Nodes $m_i, m_j \in \mathcal{M}$ in the coarse-grained network are connected ($A^*_{ij} = 1$) if there exists at least one link between any two $n_k \in m_i$ and $n_l \in m_j$ (see Fig. 1.A). This includes the case $i = j$: nodes in the coarse-grained

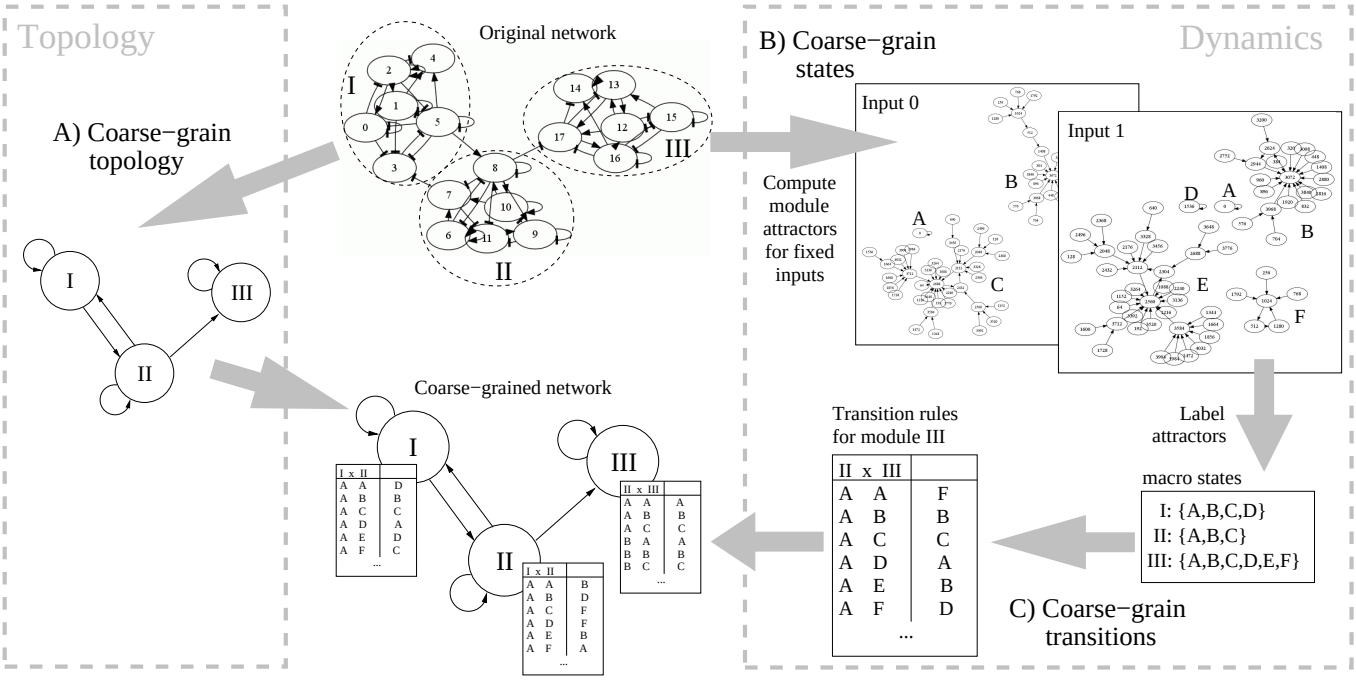

FIG. 1: Schematics of the coarse-graining procedure: **A)** Starting from a network with given modular structure (top center), the coarse-grained topology lumps together each module into a single node (left). These nodes a linked to reflect the wiring of the original modules. **B)** For each module, the set of attractors is computed for all possible input configurations (shown in the upper right for module III). These attractors are labeled and the labels serve as macro states of the coarse-grained notes (letters A to F). Note that an attractor can appear unaltered for different input configurations (attractors A and B in the example). **C)** Transition rules for each module are generated from the micro dynamics (see text for details) the resulting transition rules can be non-deterministic if a macro state can lead to several following macro states. Steps B and C are performed for every module in the network.

network are self-linked, which will be important when defining transition rules.

B. Coarse-grained states

As nodes in the coarse-grained network correspond to modules in the original network, states of the coarse-grained network correspond to attractors of the respective module.

The relaxation requirement ensures that no module input changes before the module has reached one of its attractors, i.e. each input link is either set to 0 or 1 as long as the dynamics is transient. For each module m_i exist 2^{k_i} possible input configurations and each input configuration will deterministically approach its attractor.

Algorithmically, the procedure iterates over all modules m_i and all input configurations of this module and computes the resulting attractor landscape. Let $\omega_i = \left\{\omega_i^j\right\}$ denote the set of possible attractors (where j runs over all attractors). By labeling this set, we can define the set of states for each macro node (see Fig. 1.B):

$$S_i^* = \left\{\omega_i^j\right\} \tag{3}$$

and the total set of states for the coarse-grained system:

$$S^* = \left\{s_i^{*j} \in S_i^* \mid i = 1, \ldots, M\right\}. \tag{4}$$

We emphasize that the resulting coarse-grained networks are not Boolean networks, since macro nodes can have more than two distinct states. They are, however, still discreet systems with a finite set (S^*) of possible states.

C. Coarse-grained transition rules

Assuming macro node m_i has input modules m_j for $j \in J_i$. The procedure needs to define transition rules for every combination of input states of the form

$$\underbrace{s_{j_1}^* \times s_{j_2}^* \times \ldots}_{\text{input configuration}} \longrightarrow \underbrace{s_i^*}_{\text{result}}. \tag{5}$$

Since macro nodes will be self-linked ($i \in J_i$), the future state will also depend on the current state of the respective module, itself. For any given condition (l.h.s.), the result can be found as follows (see Fig. 1.C.):

1. For each module input link, find its micro state (0 or 1) for the current macro state. Since the

state sequence of every attractor is known, the micro state can be extracted from the knowledge of the macro state just by considering the state of the input node. If the separation of time scale requirement is fulfilled, the micro state does not fluctuate but is either always 0 or always 1.

2. For each input configuration, find the attractors in the landscape for this input configuration that can be reached from any state in the current attractor. The resulting subset of S_i^* is the r.h.s. of the transition rule. Note that this can result in non-deterministic dynamics: if the attractor is not a fixed point but a limit cycle, different micro states of this attractor can lie in different basins of attraction in the attractor landscape for the switched input configuration, and can thus evolve to one of several macro states.

Since the resulting projection is a non-deterministic system, anyway, we can disregard the restriction of time scale separation in step 1 and consider both alternatives in the set of input configurations if the input module is in an attractor where the connecting node fluctuates between 0 and 1.

Knowing the transition rules for all super-nodes, one can generate the transition function

$$f^* : s^*(t) \longrightarrow s^*(t+1) \tag{6}$$

and the attractor landscape of the coarse-grained network.

The procedure may easily be extended to include probabilities for the macro transitions: the probability to approach a certain macro state is proportional to the number of states in the current attractor of the module which lead to the attractor that is associated with the respective macro state. Second, the probability for a given micro input configuration is given by the relative occurrences of 0's and 1's in the attractors realized by the input modules. The overall probability for a transition rule is simply the probability that the associated input configuration occurs times the probability that the systems runs in the attractor associated with the new macro state. Though relatively simple to implement, the forthcoming analysis focuses on topological features of the projection. The stochastic process defined by the above mentioned probabilities is not part of the current study.

Analytical considerations

To appreciate the efficiency of the coarse-graining procedure, let us consider the computational effort to construct the global attractor landscape for a given network. The original network requires 2^N update steps. Coarse-graining of the network requires $\sum_{i=1}^{M} 2^{|m_i|+k_i}$ steps, since the landscapes of all modules under all input configurations need to be computed. It takes $\prod_{i=1}^{M} |S_i^*|$ additional steps to compute the global attractor landscape of the coarse-grained network. Assuming that $|m_i| = N/M$, $k_i = k$ (all modules have the same size and indegree), and $|S_i^*| = \bar{S}$ (the worst case, in which the product is maximized), the overall computational effort is

$$M 2^{N/M+k} + \bar{S}^M. \tag{7}$$

We are interested in the scaling of the algorithm for $N \to \infty$, $M \to \infty$ with constant N/M and constant k, meaning that we increase the overall network size by adding ever more modules, but keep the size of the modules constant. The efficiency of the procedure (ratio of necessary steps in the original network divided by necessary steps in the coarse-grained network) is then

$$\frac{2^N}{Mc + \bar{S}^M} \in \mathcal{O}\left(\frac{2^N}{\bar{S}^M}\right) \leq \mathcal{O}\left(2^{N-M^2}\right), \tag{8}$$

where $c = 2^{N/M+k}$ is constant, and the equality is fulfilled only if $\bar{S} = 2^M$, i.e. if all states of a module are fixed states. In practice, the set of attractors is much smaller, and known to grow superpolynomial in M for synchronous Kauffman networks [20]. There is no closed analytical estimate for $\bar{S}$ as a function of M and we will estimate the efficiency of the method by simulation (see section III C).

III. RESULTS

We start this section with some general observations: when the original network has attractors that involve rapidly changing intermodular links (i.e. when the requirement of time scale separation is not fulfilled), the nodes of a module can reside in states that are transient under the static input assumed in the coarse-graining process. If this happens, the module's state is not captured by any macro state. The procedure still generates a projection, but the result is invalid since it does not preserve the original dynamics. Therefore, a check of the reliability of the method is needed.

Even if all attractors are preserved and the macro dynamics are deterministic, the coarse-grained network can yield less attractors than the original one. Consider a network with (non-fixed-point) attractors in several modules that can be aligned to each other in different phases. The coarse-graining procedure discards the information which exact state of the attractor the system is in. As a consequence, the phase shift of these attractors is ignored and the procedure will map all possible combinations on a single attractor in the landscape recovered from the coarse-grained network. This loss of attractors can only happen if the phase shift does not influence the global dynamics, and the reduction of attractors is therefore desirable.

Non-deterministic attractors of the coarse-grained network are subgraphs where some nodes have an outdegree higher than 1. As said, there are two causes for non-deterministic attractors: first, micro states correspond-

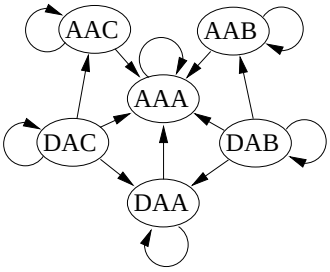

FIG. 2: Non-deterministic attractor obtained from simulating a Kauffman network with 3 modules and one input from other modules per module. Node labels denote the global macro states as a concatenation of the macro states of each module (Module I has 4 possible macro states, module II has one possible macro state, and module III has 3 possible macro states). Links show possible transitions from one macro state to the next.

ing to the current macro state of a module can lie in different basins of attraction after a changing input configuration; second, the input configuration itself can fluctuate. Due to the way the procedure deals with input fluctuations, nodes in the non-deterministic attractor tend to be self-linked. This occurs when one of the combinations of all possible micro-inputs from other modules is equal to the input configuration that lead to the current macro state (an example of this is shown in Fig. 2). In principle, the original network might take any path through this graph (assuming that the coarse-grained representation is a correct projection). However, the actually realized paths are restricted by synchronization of the micro dynamics: in the graph of Fig. 2, it might for example happen that the micro input configuration which is responsible for moving the system from macro state DAB to AAB occurs only at times where the micro dynamics of module 1 are insusceptible to any input from other modules, such that the transition from DAB to AAB never happens. This synchronization effect is lost in the coarse-graining process.

In the context of biological modularity, however, we should reexamine Fig. 2: given the absence of a global clock, molecular interactions occur asynchronous, so that the micro synchronization of module states and input configurations is likely to be an artifact of the method [21, 22]. In that case, there will eventually come a time where the module is susceptible to its input and respond with the transition to state AAA. Therefore, state DAB is an unstable attractor [23] and can be removed from the attractor graph. Recursively removing all states that have just themselves as their only input, we arrive at a further reduced state space that reflects the attractor of the system with respect to the mentioned asynchronicity (leaving state AAA as a deterministic fixed-point in the example).

A useful coarse-graining method ought to be reliable, efficient, and yield a mainly deterministic result. We will analyze each of these aspects in the following.

A. Reliability

In order to check the reliability of the coarse-graining results, we run simulations of randomly generated networks and compare the original attractor landscape with the attractor landscape recovered from the coarse-grained transition rules. An original attractor is preserved if and only if there exists a coarse-grained attractor such that all the original states occur as macro state in the coarse-grained attractor. We emphasize that the length of the attractor does not need to be preserved, since several micro configurations can be described by the same macro configuration.

We coarse-grain randomly generated modular networks with both Kauffman and activator-inhibitor transition rules and varying number of modules M and nodes within modules N/M. The networks are generated as follows: the N nodes are grouped into M modules of size N/M. Within each module, nodes are randomly wired such that each node has k input links. Additionally, nodes among different modules are connected such that each module receives l inputs from nodes in other modules. Link types (for activator-inhibitor networks) and transition rules (for Kauffman networks) are assigned randomly. We performed simulations for a total of 1040 networks with $k = 3$, $l = 1$, M ranging from 2 to 6, and N/M ranging from 3 to 7 such that $N < 20$. For each parameter set, we compute the projections of 40 differently wired networks, and for each network, we count the fraction of recovered attractors. The relatively small network sizes allow us to systematically sample the whole state space which grows exponentially with the number of nodes.

Fig. 3 shows the fraction of successfully coarse-grained networks, i.e. networks where all global attractors are preserved under coarse-graining. Each data point corresponds to networks with equal number of modules and nodes per module. Data points for networks with equal number of nodes per module are connected by lines labeled with the number of nodes. The results show that activator-inhibitor networks are generally more likely to be projected correctly than Kauffman networks of the same size. For both classes, the success rate decreases with the number of modules as well as with the number of nodes per module, but the trends are significantly stronger for Kauffman networks than for activator-inhibitor networks. Furthermore, the data appears to be more noisy for the former class than for the latter. The procedure performs comparable for networks with the same number of nodes but different partitioning (e.g. $N = 12 = 6 \times 2 = 4 \times 3 = 3 \times 4$).

In the activator-inhibitor networks generated before, 97% of all attractors can be projected correctly on the average – practically independent of the number of modules and number of nodes per module (see Fig. 4). Therefore, the slight trend in Fig. 3 toward less reliable projections for bigger networks is most likely due to the increased number of attractors and the fact that the networks are

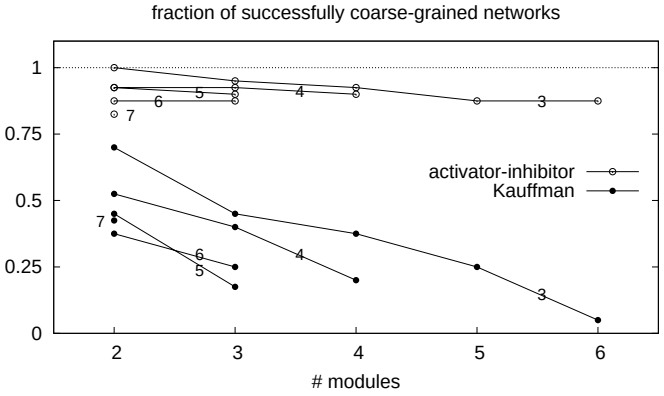

FIG. 3: Data points show the fraction of successfully coarse-grained networks, i.e. networks where all global attractors are preserved under coarse-graining. Each data point is computed from a sample of 40 randomly generated Kauffman and activator-inhibitor networks with equal number of modules and nodes per module. Data points for networks with equal number of nodes per module are connected by lines labeled with the number of nodes.

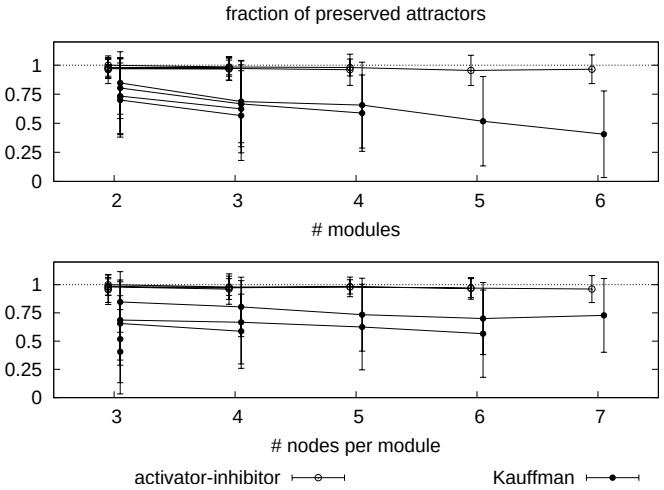

FIG. 4: Fraction of attractors that are preserved by the coarse-graining method for randomly generated modular networks (see text) as a function of the number of modules and nodes per module. Both Kauffman and activator-inhibitor dynamics are analyzed. Each error bar shows the average fraction and standard deviation of 40 networks with identical topology.

considered only to be coarse-grained successfully if all their attractors can be preserved. Fig. 4, however, shows that even if a network is not coarse-grained absolutely correctly, the majority of its attractors are still represented accurately. These observations do not hold for Kauffman networks: despite the high standard deviation, the results show that the successful projection of an attractor is less likely in networks with many modules (but does not depend on the number of nodes per modules). This suggests that attractors extend over several modules in Kauffman networks but are constrained to intramodular attractors in activator inhibitor-networks.

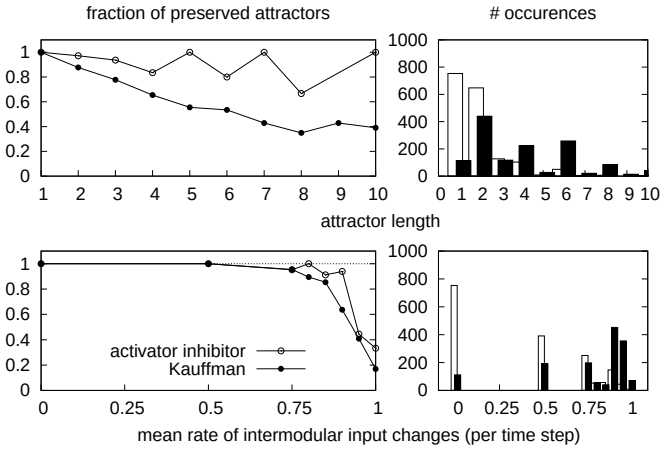

FIG. 5: Reliability of the method as a function of attractor length and mean rate of intermodular input change for activator-inhibitor and Kauffman networks with 3 modules of 5 nodes each. Each node has 3 inputs from nodes within the module and each module has 1 link from another module. For each network class, results have been binned and averaged over 110 randomly wired networks. Upper left) Fraction of preserved attractors as a function of attractor length. Upper right) Frequency histograms of attractors over attractor length. Lower left) Fraction of preserved attractors as a function of the mean rate of intermodular input change (see text). Lower right) Frequency histograms of attractors over mean intermodular input change.

So far, the reliability of the coarse-graining procedure has been analyzed with respect to topological features, and non-topological properties have only been discussed with respect of the different classes of updating rules. It has become apparent in the previous analysis that the latter choice has major impact on the reliability of the procedure. Consequently, we turn our attention to other dynamic properties of the networks.

In particular, the randomly generated networks have not been tested to exhibit separation of time scales. Therefore, we need to analyze the performance of the procedure with respect to this requirement. We define the following measure: for each node in the original network, we identify those nodes that provide input for other modules. For a given attractor, we measure the number of time steps between two subsequent input changes for each input link and take the inverses of these time spans to get the respective change rates (for fixed points, we define a change rate of 0). We then take the average of all these rates as an estimate for the global intermodular input change rate for a network in a given attractor.

Fig. 5 shows the fraction of preserved attractors as a function of their length and average input change rate together with frequency histograms of encountered attractor lengths and rates. Results have been obtained from coarse-graining 110 randomly wired networks with 3 modules of 5 nodes each for both Kauffman and activator-inhibitor networks. The results show that the reliability of coarse-graining decreases monotonically

with the attractor length in case of Kauffman networks (upper panels of Fig. 5). This tendency is not clearly visible in case of activator-inhibitor networks, where the fraction of preserved attractors fluctuates with increased attractor length. This might be due to statistical fluctuations given the small number of long attractors in this class of networks. Nevertheless, the figure shows that the method performs more reliable for activator-inhibitor networks no matter the length of the attractors. Consulting the frequency histogram, one can conclude that the high number of short attractors in activator-inhibitor networks paired with the general tendency toward better projections explains why the reliability of the method is more reliable in case of activator-inhibitor networks than in case of Kauffman networks.

The results show further that the method fails for networks that do not exhibit separation of time scales (mean input change rate close to one), while performing accurately when modular input changes slow (lower panels of Fig. 5). For both classes of networks, more than 80% of the attractors are preserved if the input change rate is lower than 0.9. Whether one class of networks can be projected more reliable with respect to time-scale separation cannot be decided from the results due to statistical fluctuations. However, the frequency histograms reveal that activator-inhibitor networks yield dynamics with much stronger separation of time scales than Kauffman networks, even when one disregards fixed-point solutions (which are always projectable accurately). Summarizing, the results confirm the initial hypothesis that motivated our method.

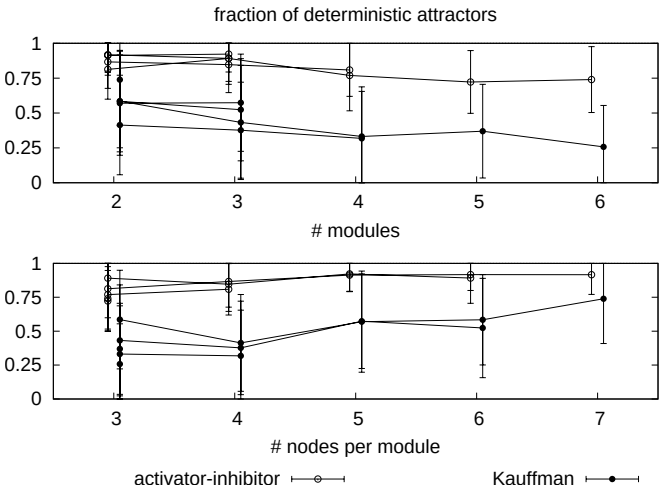

FIG. 6: Fraction of deterministic projected attractors for randomly generated modular Kauffman and activator-inhibitor networks. Only networks that can be projected reliably are taken into account. Each error bar shows the average fraction and standard deviation of 40 networks with identical topology.

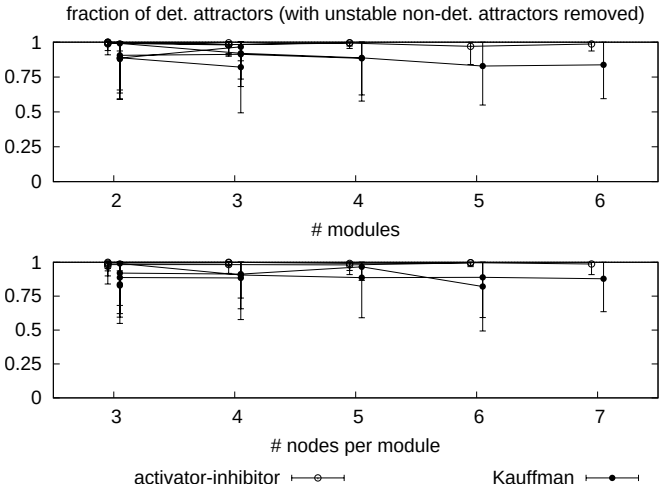

FIG. 7: Fraction of deterministic projected attractors for the networks of Fig. 6 after unstable non-deterministic attractors have been removed from the attractor landscape.

B. Determinism

Beside being reliable, a coarse-graining procedure ought to yield useful results. Given the possibility that the method produces non-deterministic projections, one might be afraid, that the procedure – though able to compute a projection – does so by merely lumping together all attractors into a single non-deterministic component, basically stating that "everything is possible after anything". In order to address this concern, we are analyzing the determinism of randomly generated projected networks.

We measure the determinism of a coarse-graining result as the fraction of deterministic attractors in the attractor landscape of the coarse-grained network. When doing so, one has to bear in mind the comment on asynchronicity that was given in the introduction of this section. In the forthcoming analysis, we both give the fraction of originally deterministic attractors as well as the fraction of deterministic attractors that remains after recursively removing unstable attractors, i.e. nodes of the state space that have just themselves as the only input link (but leaving those with no outlink other than to themselves).

Randomly wired networks are generated with the procedure used in the last section, but the analysis is confined to only those networks, that can be projected with 100% reliability – i.e. all original attractors are preserved by the coarse-graining. For each parameter combination, 40 networks are analyzed.

Fig. 6 shows the fraction of deterministic attractors in the unmodified attractor landscape of the coarse-grained system. Fig 7 shows the fraction of deterministic attractors after unstable non-deterministic attractors have been removed. As has been observed for the reliability of the procedure, its tendency to produce deterministic results is higher and shows less deviation for activator-inhibitor networks than for Kauffman networks. For the first network class, no clear trend is visible with respect

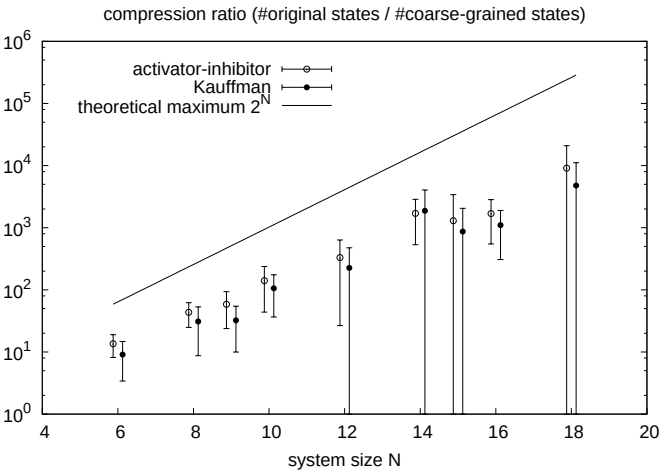

FIG. 8: Compression ratio (number of coarse-grained states over number of original states) as a function of the network size N for randomly wired modular Kauffman and activator-inhibitor networks. The solid line marks the theoretical maximum 2^N which corresponds to a compression of the whole original state space onto a single state.

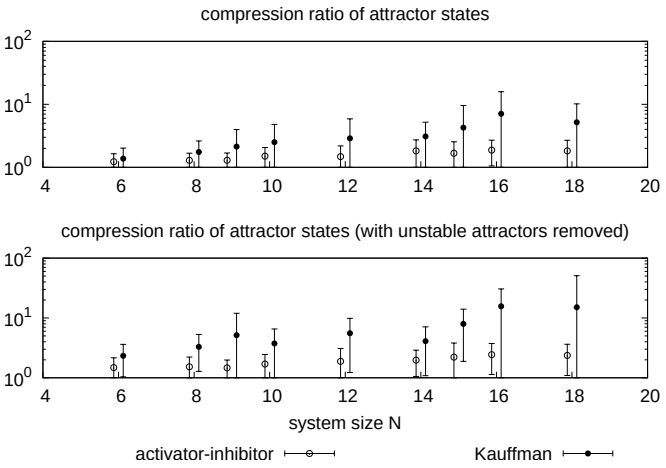

FIG. 9: Compression ratio of the method with respect to the number of attractor states of the original versus coarse-grained networks. The upper panel refers to all coarse-grained attractors, the lower panel to those attractor states that remain when removing unstable non-deterministic attractors (see text).

to network and module sizes, whereas projected attractors of Kauffman networks tend to be more deterministic for bigger modules.

Removing unstable non-deterministic nodes from the attractor-landscape of the coarse-grained network significantly increases deterministic results: in Fig. 7, 99% of all coarse-grained attractors turn out to be deterministic for activator-inhibitor networks, while for Kauffman networks, the fraction of deterministic attractors increases to 90%, with no visible trend, implying that in both networks, most non-deterministic components are of the type of Fig. 2 and approach a deterministic attractor when asynchronicity is assumed.

C. Efficiency

We define the efficiency of the method as the compression ratio of the state space (number of number of original states divided by coarse-grained states):

$$E = \frac{|S|}{|S^*|} = 2^N \Pi_{i=1}^{M} |S_i^*|^{-1} \tag{9}$$

Fig. 8 shows the efficiency of the method for the networks analyzed in section III B as a function of the network size N. The solid line marks the theoretical maximum 2^N which corresponds to a compression of the whole original state space onto a single state (unique fixed-point). We find that the efficiency of the method roughly follows an exponential advance – though with a smaller exponent than its theoretical maximum. Activator-inhibitor networks yield a slightly higher compression, but no significant trend can be spotted in their difference over the examined range of network sizes.

In general, the compression is remarkably high: for the tested networks, one finds a compression ratio about 10 for networks of only 6 nodes (i.e. 6 coarse-grained instead of 64 original network states) up to a ratio about 1000 for networks of 18 nodes (corresponding to about 260 coarse-grained rather than more than 260.000 original network states).

Observing alternatively the number of original attractor states versus the number of coarse-grained attractor states (see Fig. 9) yields significantly smaller compression ratios in the range of 1 to 50. This, in turn, implies that most of the compression results from the removal of transient nodes in the attractor landscape of the original network. Fig. 9 further shows a trend toward higher compression of large networks in case of Kauffman networks which is absent in case of activator-inhibitor networks. This is likely due to the compression of long attractors which are more numerous in Kauffman networks.

IV. DISCUSSION

In this work, we have developed a novel method to coarse-grain modular Boolean networks. We have analyzed the performance of the method on randomly wired modular Kauffman and activator-inhibitor networks with respect to reliability, determinism, and efficiency, where reliability measures whether the coarse-graining result is correct (system attractors are preserved), determinism measures whether the result is useful (does posses few non-deterministic nodes), and efficiency measures the compression ratio of the coarse-grained versus the original system. In the simulations, network sizes have been kept small in order to allow for a systematic traversal of their state spaces.

Our analysis has revealed that activator-inhibitor networks can be coarse-grained significantly more reliable than Kauffman networks with equal topology (i.e. number of modules and nodes per module). At the same time, their coarse-grained projections posses a higher ratio of deterministic attractors, and their state space is slightly better compressed. The difference in performance with respect to these measures is mainly due to a "more orderly" dynamics in activator-inhibitor networks, i.e. they yield shorter attractors which are mainly confined within individual modules. As a result, information is exchanged relatively slow among modules, which is the requirement that motivated the design of the method (time scale separation). On the other hand, we find that networks without time scale separation, as is the case for a majority of Kauffman networks, are likely to be projected incorrect.

In activator-inhibitor networks, the vast majority (97%) of attractors can be projected correctly – practically independent of the size of the tested networks. The method is able to compress the state space of the system dramatically, e.g. by a factor of 1000 for networks with 20 nodes. Most of the compression results from the removal of transient nodes in the attractor landscape of the original network, rather than from reducing the set of attractor states. At the same time, more than 2/3 of the coarse-grained attractors are deterministic. When considering only non-deterministic components that are stable to asynchronous updates – which is a plausible assumption in case of biological networks that lack a central clock – the number of deterministic attractors increases to 99%, i.e. the amount of uncertainty introduced by the coarse-graining is vanishingly small.

The high reliability and determinism of the coarse-graining method for biologically plausible update functions implies that modules in biological systems are robust: the ability of our coarse-graining method to reliably identify macro-states and the fact that the majority of transitions among macro states is deterministic are consequences of the loose intermodular coupling in the first place. The comparison to Kauffman networks implies, that these networks owe their robustness not only to topological properties, but also to the specific mechanism of interactions (activator-inhibitor dynamics).

Acknowledgments

This work was partially supported by the Santa Fe Institute whose research and education programs are supported by core funding from the National Science Foundation and by gifts and grants from individuals, corporations, other foundations, and members of the Institute's Business Network for Complex Systems Research.

[1] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99:, 7821–7826, 2002.

[2] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297:1551–1555, 2002.

[3] A. W. Rives and T. Galitski. Modular organization of cellular networks. *Proc. Natl. Acad. Sci. USA*, 100(3):1128–1133, 2003.

[4] L. Danon, J. Duch, A. Diaz-Guilera, and A. Arenas. Comparing community structure identification. *J. Stat. Mech.*, page P09008, 2005.

[5] Jian jun Wu, Zi you Gao, and Hui jun Sun. Cascade and breakdown in scale-free networks with community structure. *Phys. Rev. E*, 74:066111, 2006.

[6] A. Arenas, A. Diaz-Guilera, and C. J. Perez-Vicente. Synchronization reveals topological scales in complex networks. *Phys. Rev. Lett.*, 96:114102, 2006.

[7] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.*, 22:437–467, 1969.

[8] K. C. Chen, A. Csikasz-Nagy, Gyorffy B., J. Val, B. Novak, and J. J. Tyson. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell*, 11:369–391, 2000.

[9] B. Novak, Z. Pataki, A. Ciliberto, and J.J. Tyson. Mathematical model of the cell division cycle of fission yeast. *Chaos*, 11(1):277–286, 2001.

[10] B. Novak and J.J. Tyson. Numerical analysis of a comprehensive model of m-phase control in xenopus oocyte extracts and intact embryos. *J. Cell Sci.*, 106:1153–1168, 1993.

[11] B. Novak and J. J. Tyson. Modeling the control of DNA replication in fission yeast. *Proc. Natl. Acad. Sci. USA*, 94:9147–9152, 1997.

[12] B. Novak and J. J. Tyson. A model for restriction point control of the mammalian cell cycle. *J. Theor. Biol.*, 230:563–579, 2004. doi:10.1016/j.jtbi.2004.04.039.

[13] J. J. Tyson, Csikasz-Nagy, and B. A., Novak. The dynamics of the cell-cycle regulation. *BioEssays*, 24:1095–1109, 2002.

[14] M. Davidich and S. Bornhold. Boolean network model predicts cell cycle sequence of fission yeast. *PLOS ONE*, 27(3):e1672, 2008.

[15] M. Nilsson Jacobi. Hierarchical dynamics. In Robert A. Meyers, editor, *Encyclopedia on Complex Systems*. Springer. in press.

[16] D. Givon, R. Kupferman, and A. Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, 17:R55R127, 2004.

[17] S. Boccaletti, M. Ivanchenko, V. Latora, A. Pluchino, , and A. Rapisarda. Detecting complex network modularity by dynamical clustering. *Phys. Rev. E*, 75:045102R, 2007.

[18] M. E. J. Newman. Detecting community structure in networks. *Eur. Phys. J. B*, 38:321330, 2004.

[19] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, 69:066133,

2004.

[20] B. Samuelsson and C. Troein. Superpolynomial growth of the number of attractors in Kauffman networks. *Acta Physica Polonica B*, 34(10), 2003.

[21] C. Gershenson. Classification of random boolean networks. In *Artificial Life VIII: Proceedings of the Eighth International Conference on Artificial Life*, pages 1–8. MIT Press, 2002.

[22] H. Mahmoudi, A. Pagnani, M. Weigt, and R. Zecchina. Propagation of external regulation and asynchronous dynamics in random boolean networks. *Chaos*, 17:026109, 2007.

[23] K. Klemm and S. Bornholdt. Stable and unstable attractors in Boolean networks. *Phys. Rev. E*, 72:055101(R), 2005.