

Exploring the organisation of complex systems through the dynamical interactions among their relevant subsets

Alessandro Filisetti^{1,2}, Marco Villani⁴, Andrea Roli⁵, Marco Fiorucci³ and Roberto Serra⁴

¹European Centre for Living Technology, Venice

²Explora s.r.l, Rome, Italy

³Department of Environmental Sciences (DAIS), Università Ca' Foscari

⁴Department of Physics, Informatics and Mathematics, Università di Modena e Reggio Emilia

⁵Department of Computer Science and Engineering (DISI), Università di Bologna

a.filisetti@explora-biotech.com

Abstract

Complex systems often show forms of organisation where a clear-cut hierarchy of levels with a well-defined direction of information flow cannot be found. In this paper we propose an information-theoretic method aimed at identifying the dynamically relevant parts of a system along with their relationships, interpreting in such a way the system's dynamical organisation. The analysis is quite general and can be applied to many dynamical systems. We show here its application to two relevant biological examples, the case of mammalian cell cycle network and of Mitogen Activated Protein Kinase (MAPK) cascade. The result of our analysis shows that the elements of the mammalian cell cycle network act as a single compact group, whereas the MAPK system can be decomposed into two dynamically distinct parts, with asymmetric information flows.

Introduction

Complex systems often show forms of organisation that cannot be understood by referring to simple hierarchical models (like e.g. a tree). In most interesting cases one is faced with complicated situations, sometimes referred to as “tangled hierarchies”, where a clear-cut hierarchy of levels, with a unique well-defined direction of information flow cannot be found.

In previous work (Villani et al., 2013; Villani and Serra, 2013; Villani et al., 2015) we have introduced the *Dynamical Cluster Index method* (shortly DCI¹), which makes use of a measure based on information theory that appears well suited to explore the organisation of such complex systems. The DCI makes it possible to identify Candidate Relevant Subsets of variables (CRS) that show an integrated behaviour, while interacting more weakly with the rest of the system. The DCI has been applied with interesting results to several systems: some of them were artificially designed in order to test the effectiveness of the technique, while others refer to interesting chemical or biological systems (Filisetti et al., 2011; Sarma and Ghosh, 2012; Chaos et al., 2006; Farmer and Kauffman, 1986). Moreover, ongoing and yet

¹When clear from the context, the same acronym ‘DCI’ will be used either to denote the index or the method using it.

unpublished work is also showing promising results in the study of social systems such as the detection of dynamical communities in regional innovation projects studied within the “Emergence by Design”² European project, the Tuscany innovation system and the UK consumer price indices.

In many cases, though, the identified CRS have an intricate nested structure, so that it might not be clear which groups of variables are really important. To overcome this problem, a filtering algorithm is needed, which should be able to select those CRS responsible for the dynamics of the system. To this aim, in this paper we introduce a sieving procedure able to extract, from the set of candidate CRS, a smaller number of disjoint or partially overlapping sets, that capture the basic features of the groups of integrated variables, thus reducing the need for user intervention in the interpretation of the outcomes of the method.

In this work we also introduce a further improvement by considering the flow of information among the identified subsets of variables (i.e. relevant subsets RS) that is estimated on the basis of transfer entropy (Schreiber, 2000). In this way it is possible to determine whether there is an information flow from set *A* to set *B*, and vice-versa. Moreover, also the direction of the net information flow is defined. We use this method to analyse some artificially designed cases and also two biological models.

The paper is structured as in the following. In the subsequent sections we summarise the theoretical background and describe the sieving procedure and the technique used to detect the information flow among RS. Then, several case studies are presented, while in the final section conclusions and further developments are discussed.

The Dynamical Cluster Index method

In this section we summarise the main features of the DCI, as presented in (Villani et al., 2013, 2015). The DCI is an extension of the Functional Cluster Index (CI) introduced by Edelman and Tononi in 1998 (Tononi et al., 1998) and aimed at detecting functional clusters in brain regions. In our work, we relaxed the stationary constraint and extended the CI to

²EU grant agreement n. 284625

actual dynamical systems, so as to apply it to a wide range of system classes, from abstract models to biological and socio-economic models.

The DCI is an information theoretical measure based on the well-know Shannon Entropy (Cover and Thomas, 2006). The entropy $H(X)$ of a random variable X is computed as:

$$H(X) = - \sum_x p(x) \log p(x) \quad (1)$$

The joint entropy of a pair of variables $H(X, Y)$ as:

$$H(X, Y) = - \sum_x \sum_y p(x, y) \log p(x, y) \quad (2)$$

where Y assumes values y . Note that eq. 2 can be naturally extended to sets of k elements. As in our work we rely on observational data, probabilities are estimated by the relative frequencies of the values observed for each variable. Let us now consider a system U composed of K variables (e.g. agents, chemicals, genes, artificial entities) and suppose that S_k is a subset composed of k elements, with $k < K$. The value $DCI(S_k)$ is defined as the ratio between two measures based on the aforementioned entropy: the *integration* (I) and the *Mutual Information* (MI). $I(S_k)$ measures the statistical independence of the k elements in S_k and is defined as:

$$I(S_k) = \sum_{s \in S_k} H(s) - H(S_k) \quad (3)$$

Then mutual information $MI(S_k; U \setminus S_k)$ measures the mutual dependence between subset S_k and the rest of the system $U \setminus S_k$ and it is defined as:

$$MI(S_k; U \setminus S_k) = H(S_k) + H(U \setminus S_k) - H(S_k, U \setminus S_k) \quad (4)$$

hence, the DCI,

$$DCI(S_k) = \frac{I(S_k)}{MI(S_k; U \setminus S_k)} \quad (5)$$

The value of DCI is not defined if $MI(S_k; U \setminus S_k) = 0$. However, a vanishing mutual information is a sign of separation of the investigated subset from the rest of the system, and therefore the subset must be studied separately. It is worth noting that the DCI scales with the subset size. In Viliani et al. (2015) we show a procedure to normalize it, nevertheless a better approach is that of assessing the statistical significance of the DCI of S_k by means of a statistical significance index:³

$$t_c(S_k) = \frac{DCI(S_k) - \langle DCI_h \rangle}{\sigma(DCI_h)} = \frac{\nu DCI - \nu \langle DCI_h \rangle}{\nu \sigma(DCI_h)} \quad (6)$$

³Introduced in (Tononi et al., 1998).

where $\langle DCI_h \rangle$ and $\sigma(DCI_h)$ are respectively the average and the standard deviation of the DCI of a sample of subsets of size k extracted from a reference system U_h randomly generated according to the frequency of each single state in U and $\nu = \langle MI_h \rangle / \langle I_h \rangle$ is the normalisation constant. It is worth noting that the aim of the reference system is that of quantify the finite size effects affecting the information theoretical measures on a random instance of a system with finite dimensions.

A list of CRS can be in principle obtained by computing the DCI of every possible subset of variables in U and ranking the subsets by DCI values. The subsets occupying the first positions are most likely to play a relevant role in system dynamics. For large-size systems, exhaustive enumeration is computationally impractical as it requires to enumerate the power set of U . In this case, we resort to heuristic algorithms, such as genetic algorithms. However, since in this work we are interested in uncovering some properties concerning the information exchanged among RS, only small systems suitable for an exhaustive assessment will be tackled, postponing to a future contribution the presentation of the application of the DCI to large-size systems.

Finally, random perturbations are applied to the dynamical system in order to observe their dynamical feedbacks also in case of stable situations.

Sieving the candidate subsets

The list of CRS can be ranked according to the significance of their DCI. In previous works we directly used this ranking to identify by hand the CRS relevant for the dynamics of the system. Nevertheless, in many cases this analysis might return a huge list of entangled CRS, so that a direct inspection is required for explaining their relevance. To this aim, we present a DCI analysis post-processing sieving algorithm to reduce the overall number of CRS to manually tackle. The main assumption of the algorithm is that if a CRS A is a proper subset of CRS B , i.e. $A \subset B$, then only the subset with the higher DCI is maintained between the two. Thus, only disjoint or partially overlapping CRSs are retained: the used assumption implies that the remaining CRSs are not further decomposable, forming in such a way the “building blocks” of the dynamical organisation of the system. The pseudo-code of the algorithm is presented in Algorithm 1.

Assessing the temporal correlation among the subsets: the D index

Although by the application of the DCI, CRS are detected, this measure does not provide indications neither on the quantity nor on the direction of the entropy, i.e. the information flowing among subsets. To this aim we applied the directionality index proposed by Lautier and Raynaud (2014).

Let X and Y be two random variables—or, equivalently, two sets of variables. We can define the **entropy rate** of X as the average number of bits needed to encode a successive

Algorithm 1 Sieving algorithm

Input: The array C of all the CRS ranked by their $t_c(\text{DCI})$
Output: A subset $RS \in CRS$
 $RS \leftarrow \emptyset$
 $n \leftarrow |CRS|$
Initialize auxiliary array $Del[k] \leftarrow \{0 \text{ for } k \text{ in } 1 \dots n\}$
for $i = 1$ to $n - 1$ **do**
 for $j = i + 1$ to n **do**
 if $Del[i] \neq 1 \wedge Del[j] \neq 1$ **then**
 if $C[i] \subset C[j] \vee C[j] \subset C[i]$ **then**
 $Del[j] \leftarrow 1$
 end if
 end if
 end for
end for
for $i = 1$ to n **do**
 if $Del[i] = 0$ **then**
 $RS \leftarrow RS \cup \{C[i]\}$
 end if
end for

state of X if all the previous states are known, considering that the value of X at $t + 1$ depends either on X and Y at time t , eq. 7a, or just on the value of X at time t , eq. 7b.⁴

$$h_1 = - \sum_{X,Y} p(x^{t+1}, x^t, y^t) \log p(x^{t+1} | x^t, y^t) \quad (7a)$$

$$h_2 = - \sum_{X,Y} p(x^{t+1}, x^t, y^t) \log p(x^{t+1} | x^t) \quad (7b)$$

then, the transfer entropy T is defined as the difference between the aforementioned entropy rates. T describes how the uncertainty of X is reduced by knowing the previous states of Y and X itself, eq. 8.

$$\begin{aligned} T_{Y \rightarrow X} &= h_2 - h_1 \\ &= \sum_{X,Y} p(x^{t+1}, x^t, y^t) \log \frac{p(x^{t+1} | x^t, y^t)}{p(x^{t+1} | x^t)}, \end{aligned} \quad (8)$$

Thus, $T_{Y \rightarrow X}$ can be described in term of entropy as:

$$T_{Y \rightarrow X} = H(X^{t+1} | X^t) - H(X^{t+1} | Y^t, X^t), \quad (9)$$

⁴Note that the temporal dependency is not necessarily of unitary lag, i.e. $t - 1 \rightarrow t$. For a complete assessment of the statistical dependency of X on Y one should sum over $t - 1, t - 2, \dots, t - T$, where T is the observation time limit. Nevertheless, note that (i) in this paper we are analysing Markovian systems, whose behaviour depends only from the immediately previous step and (ii) although TE is not a direct measure of causal effect, the use of short history length alters the character of the measure towards inferring causal effect (Lizier and Prokopenko, 2010).

and, since the $T_{Y \rightarrow X}$ describes the information moving from Y to X , and the **transfer entropy is not symmetric**, the information from X to Y is computed as well, eq. 10.

$$T_{X \rightarrow Y} = H(Y^{t+1} | Y^t) - H(Y^{t+1} | X^t, Y^t), \quad (10)$$

Once that $T_{Y \rightarrow X}$ and $T_{X \rightarrow Y}$ are known, the directionality D of the information flow between X and Y can be measured as:

$$D_{X \rightarrow Y} = \begin{cases} 0, & \text{if } T_{X \rightarrow Y} = T_{Y \rightarrow X} \\ \frac{T_{X \rightarrow Y} - T_{Y \rightarrow X}}{T_{X \rightarrow Y} + T_{Y \rightarrow X}} \in [-1, 1], & \text{otherwise} \end{cases} \quad (11)$$

where $D_{X \rightarrow Y} = 1$ indicates that all the information moves from X to Y , i.e. absence of information flow from Y to X and, conversely, $D_{X \rightarrow Y} = -1$ indicates that X , with respect to Y , is just an information receiver and not an information sender.

It is worthwhile to notice that D does not provide any indication about the amount of information exchanged between the variables, but it only provides suitable indications on the direction of the information flow.

As introduced above, these considerations are also valid for groups of variables; however, the number of available observations limits the number of joint variables that may be analyzed (e.g. see Kraskov et al. (2004)): this is because spurious correlations are more frequently “detected” as this limit is approached, increasing the error in measurement. In order to avoid as much as possible this effect, each perturbation of the following analyses is introduced after the system has recovered a stable dynamical condition. Finally, the small size of the relevant subsets presented in this work allows a direct use of transfer entropy; in case of bigger subsets it is nevertheless possible to use sampling techniques, as in Lizier et al. (2011).

Analysis on test cases

In order to test the presented methodology, in the following we analyse several cases whose dynamical mechanisms are precisely known. We consider here boolean networks, a modelling framework that despite its apparent simplicity has obtained remarkable results in simulating real gene regulatory networks (Serra et al., 2004; Shmulevich et al., 2005; Serra et al., 2007; Villani et al., 2011). In particular, we present a system composed of 12 boolean nodes updated on the basis to either a boolean function or a random boolean value generator. Nodes update their state in parallel and synchronously. We illustrate the results of five instances of this network, defined in Table 1.⁵

⁵Note that the size of these systems allows for an exhaustive enumeration of all the possible groups, allowing their complete assessment. It is worth remarking that each perturbation is introduced after the system has recovered a stable dynamical situation.

The five instances share a common structure but differ in specific dynamical organizations of some nodes. In *case 1*, we consider two integrated groups of three nodes (namely, *group A* and *group B*), by assigning at each node the *XOR* function of the other two nodes in the group. In this case the *sieving* algorithm filtered the 94% of the evaluated CRS, making it possible to easily identify the subsets responsible for the dynamical organization of the system. Only the two correct CRSs have high t_c values, whereas all the other ones have t_c values lower of more than 2 orders of magnitude. Moreover, no information exchange takes place between *group A* and *group B*, as they are structurally independent.

Case 2 derives from *case 1* by introducing in the first node of *group B* a further dependence from the last node of *group A*, hence introducing information transfer from *group A* to *group B*. The combination of the DCI analysis and the sieving algorithm correctly recalls the dynamical organization of the system—i.e. *group A* influences the behaviour of (a part of) *group B*. We observe that the whole *group B* (Figure 1b) was anyway ranked high w.r.t. its DCI significance, but it was discarded by the sieving algorithm because the dynamics of its first node is influenced also by *group A* and so the assessment of the whole *group B* is weakened. In general, the amount of this difference depends on the strength of the forces that influence the interface nodes and the elements interfacing different CRS can be detected by a simple comparison between the DCI analysis and the sieving algorithm outputs.

Case 3 derives from *case 2* by introducing a further dependence of the first node of *group A* from the last node of *group B*: again, the combination of the DCI and the sieving algorithm detects the interface nodes and the underlining dynamical system organization. Note that the asymmetry in transfer entropies (and therefore in D index) are due to differences in the initial conditions of the boolean network trajectories: a shift in initial conditions can change the direction of this asymmetry.⁶

In *case 4* five heterogeneously linked nodes are influenced by a triplet identical to that of *group A*. The combination of the dynamical rules of the nodes and their initial condition makes the dynamical behaviour of the sixth node always in phase with the triplet, so our analysis individuates this quartet as the most relevant CRS. The other dynamical relations are not sufficiently strong to coordinate all the 8 nodes, nevertheless the method detects their influence by returning some masks⁷ having high t_c values, partially

⁶The interesting study of this effect is out of the scope of this paper and is subject of future work; in any case, a brief exploration of a limited set of initial conditions supports this preliminary claim (data not shown).

⁷CRSs can be represented by rows, where entries corresponding

overlapped with the leading quartet. The overlap of these masks indirectly indicates the presence of a greater group with respect to the winning quartet, but having a less evident dynamical presence (see Figure 1b).

Case 5 derives from *case 1* by adding two nodes whose dynamical behaviour directly depends on nodes of both *group A* and *group B*: these 8 nodes form therefore a group clearly different from the remaining 4 random nodes, as they are interdependent and ruled by deterministic functions. This group is identified by the plain DCI method, but the combination of DCI and sieving algorithm strikingly enlightens the interpretation of two triplets directly influencing a couple of nodes (Figure 1a).

Gene regulatory networks

In this section we show the application of our method to two models of regulatory networks: a model of mammalian cell cycle network (MCC in the following), as “booleanized” in Fauré et al. (2006)—see Table 2 for the chosen boolean model—and a model of one of the major cellular signal transduction pathways, known as the Mitogen Activated Protein Kinase (MAPK) cascade (Widmann et al., 1999). In Fauré et al. (2006) the authors provides a boolean dynamical model of the mammalian cell cycle, able to reproduce the main characteristics of the succession of molecular events leading to the reproduction of the genome of a cell and its division into two daughter cells.

Mammalian cell division must be coordinated with the overall growth of the organism; this coordination is achieved through extra-cellular signals whose balance decides whether a cell will divide or remain in a resting state. The positive signals or growth factors ultimately elicit the activation of Cyclin D (CycD) in the cell. In the proposed model CycD thus represents the input and its activity is considered constant. By pointing the interested reader to (Fauré et al., 2006) for the details, for now it is enough to say that in absence of CycD the system presents a unique stable attractor where only Rb, p27 and Cdh1 are active, whereas in its presence E2F, CycE, CycA, Cdc20, Cdh1, UbcH10 and CycB cycle with a period of length 7. We perturbed both asymptotic states, obtaining in each case only one group (composed of E2F, CycE, CycA, Cdc20, Cdh1, UbcH10 and CycB in the first case, and of Rb, E2F, p27, Cdc20, UbcH10 and CycB in the second case). The leading groups of the two situations are different, but in each case the other CRS identified overlap with these groups and their sum cover the whole system, indicating the presence of a single coordinated pattern. So, the analysis indicates that the elements of the mammalian cell cycle network act as a single compact engine, see Figure 2a.

to variables belonging to CRS are marked in black (“masks” in the following)

Node	Node Rule				
	Case 1	Case 2	Case 3	Case 4	Case 5
N01	Random(0.5)	Random(0.5)	Random(0.5)	Random(0.5)	Random(0.5)
N02	Random(0.5)	Random(0.5)	Random(0.5)	Random(0.5)	Random(0.5)
N03	(N04 \oplus N05)	(N04 \oplus N05)	N10 \wedge (N04 \oplus N05)	(N04 \oplus N05)	(N04 \oplus N05)
N04	(N03 \oplus N05)	(N03 \oplus N05)	(N03 \oplus N05)	(N03 \oplus N05)	(N03 \oplus N05)
N05	(N03 \oplus N04)	(N03 \oplus N04)	(N03 \oplus N04)	(N03 \oplus N04)	(N03 \oplus N04)
N06	Random(0.5)	Random(0.5)	Random(0.5)	(N05 \oplus N08)	(N05 \oplus N08)
N07	Random(0.5)	Random(0.5)	Random(0.5)	(N07+N08+N09+N10) \geq 2	\neg (N05 \oplus N08)
N08	(N09 \oplus N10)	N05 \wedge (N09 \oplus N10)	N05 \wedge (N09 \oplus N10)	N03 \oplus N05	N09 \oplus N10
N09	(N08 \oplus N10)	(N08 \oplus N10)	(N08 \oplus N10)	(N04+N05+N07+N08) \leq 2	(N08 \oplus N10)
N10	(N08 \oplus N09)	(N08 \oplus N09)	(N08 \oplus N09)	N06 \wedge (N05 \oplus N09)	(N08 \oplus N09)
N11	Random(0.5)	Random(0.5)	Random(0.5)	Random(0.5)	Random(0.5)
N12	Random(0.5)	Random(0.5)	Random(0.5)	Random(0.5)	Random(0.5)

Table 1: The update rules of the boolean networks discussed on the text. Random(0.5) denotes a Bernoulli distribution with probability 0.5.

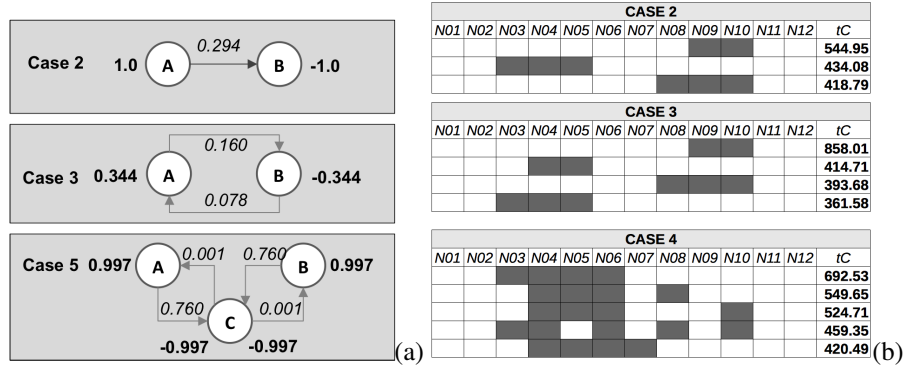


Figure 1: (a) The dynamical organization of the systems described in the text; on the edges, the Transfer Entropy, and on the nodes the D index. Group C in case 5 has two values for the D index, each value depicting the group C role in the relation with group A and group B, respectively. (b) The CRS identified by the DCI only in some systems described in the text. Each row denotes one CRS, composed of nodes whose entries are marked in black; on the right, the value of t_c . Note that in case 2 the third row is the correct identification of group B, and that in case 3 the third and fourth rows are group A and group B. For case 4 the output of the combination of the DCI analysis and sieving algorithm are presented; note that besides the correct group formed by nodes N03–N06 other CRS have high t_c values, highlighting the presence of a larger but less evident dynamical group.

Product	Logical Rule leading to an activity of product	Legend
<i>CycD</i>	$CycD$	Cyclin D
<i>Rb</i>	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{CycD} \wedge \overline{CycB})$	Retinoblastoma Protein
<i>E2F</i>	$(rB \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge rB \wedge \overline{CycB})$	Transcription factors
<i>CycE</i>	$(E2F \wedge rB)$	Cyclin E
<i>CycA</i>	$(E2F \wedge rB \wedge \overline{Cdc20} \wedge \overline{Cdh1} \wedge \overline{Ubc}) \vee (CycA \wedge rB \wedge \overline{Cdc20} \wedge \overline{Cdh1} \wedge \overline{Ubc})$	Cyclin A
<i>p27</i>	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB} \wedge \overline{CycD})$	<i>p27</i> enzyme inhibitor
<i>Cdc20</i>	$CycB$	Activators of the Anaphase Promoting Complex
<i>Cdh1</i>	$(\overline{CycA} \wedge \overline{CycB}) \vee (Cdc20) \vee (p27 \wedge \overline{CycB})$	Promoting Complex
<i>Ubc</i>	$(\overline{Cdh1}) \vee (Cdh1 \wedge Ubc \wedge (Cdc20 \vee CycA \vee CycB))$	E2 ubiquitin conjugating enzyme
<i>CycB</i>	$(\overline{Cdc20} \wedge \overline{Cdh1})$	Cyclin B

Table 2: Adapted from Fauré et al. (2006), the boolean regulatory network of mammalian cell cycle network and a short description of each node of the system.

Metabolic pathway MAPK

The MAPK pathway (evolutionarily conserved from yeasts to humans) responds to a wide range of external stimuli, triggering growth, cell division and proliferation (Sarma and Ghosh, 2012). Sarma and Ghosh also introduce the models considered in our analysis. The basic model is composed of reactions that are quite well-established from an experimental viewpoint, and it has the hierarchical structure shown in Figure 3a. The three chemicals identified as the core of this three-layered system are the *MAPKKK*, *MAPKK* and *MAPK* kinases (respectively *M3K*, *M2K* and *MK* for short) (Widmann et al., 1999). *M3K* is activated by means of a single phosphorylation whereas *M2K* and *MK* are both activated by double phosphorylation. Parallel to the phosphorylation by kinases, phosphatases present in the cellular volume can dephosphorylate the phosphorylated kinases (Figure 3a shows the schema of the *MAPK* cascade where each layer of the cascade is dephosphorylated by a specific phosphatase). Note that superimposed on the three-layered structure of substrates-product reactions there is the properly called *MAPK* signalling cascade, a linear chain of catalysis (dashed lines in Figure 3a) that transmit the external signal from *M3K** to *MK***.⁸

When the external signal and the concentrations of the phosphatases are kept constant, a top-down reaction scheme as the one described in Figure 3a leads to fixed-point solutions. On the other hand, oscillations have been reported in the MAPK cascade (Shin et al., 2009) and, in order to account for them, Sarma and Ghosh adopt a models with feedback, one of which is described in Figure 3b. This variant (called S2 in the following) is characterized by a configuration of the activating and inhibiting interactions among layers that alters the “layered” structure of the basic model, which is no longer strictly hierarchical. This alternative model is grounded on experimental data; we will not enter here a discussion about the merits and limits of the model, referring the interested reader to the original paper, but we will take it “for granted” and we will apply our method to test whether it can discover significant dynamical organization, without any prior knowledge of the interactions, but on the sole basis of the dynamics of concentrations. We simulated the Sarma and Ghosh models with the CellDesigner software (Funahashi et al., 2008, 2003), keeping the P1, P2 and P3 phosphatases as constant (as they do) obtaining the same asymptotic states shown by the authors. In order to apply our method we perturbed the asymptotic states of these models: in particular, we focused our analysis on kinases perturbations.⁹ The stable situations that are reached can

⁸The symbol “*” indicates the phosphorylated version of the molecule.

⁹In particular we performed 10 perturbation cycles, each cycle involving the perturbation of each single kinase and the successive relaxing to a stable situation before the subsequent perturbation, see Villani et al. (2015) for details.

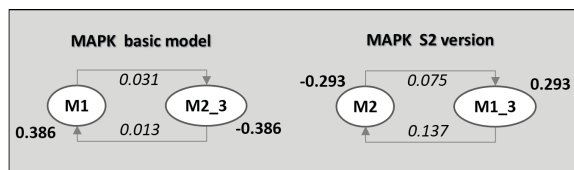


Figure 4: The dynamical organization of MAPK system (basic model and S2 version). In italic the Transfer Entropy, in bold the D index associated to the interested group within the relation. M1 group involves the first layer of 3a; M2 group involves the second layer, whereas groups M2_3 and M1_3 involve respectively the layers 2 and 3, and 1 and 3.

show both oscillating (S2 system) or constant concentrations (basic system). Concentration changes are more significant than their absolute values (it is important to monitor the variables that are changing in coordinate way); therefore the continuous concentration values are represented according to a three levels code related to the sign of the time derivatives at time t (“decreasing concentration”, “no significant change”, “increasing concentration”).

The combination of DCI and sieving algorithm applied to the basic MAPK model detects two dynamical groups: the first including the first layer of Figure 3a and the second including the other two layers. The two groups exchange information, the second transmitting more information to the first one (see Figure 4). The introduction of the feedbacks changes system dynamics: there are still two dynamically relevant groups, now composed of the second layer and by the other two layers, respectively. The analysis therefore suggests that the MAPK system may be decomposed in two dynamically distinct parts.

Conclusions

In this paper we have presented an improvement of a method based on information-theoretical measures able to identify the most relevant groups of variables that impact the dynamics of a system. We first introduced a sieving algorithm which makes it possible to identify the groups of variable that are responsible for system dynamics. Moreover, we show that it is possible to recover the direction of information flow among these groups, thus characterising the dynamical organisation of the system.

With respect to this, we noted that better normalisation methods can be applied to improve the algorithms efficacy: these methods are subject of ongoing work.

Finally, the effectiveness of this combination of techniques has been validated on test cases and subsequently applied to two prominent biological models, i.e. the mammalian cell cycle network and of Mitogen Activated Protein Kinase (MAPK) cascade.

MAPK basic model								
M3K	M3K*	M2K	M2K*	M2K**	MK	MK*	MK**	tC
								763.35
								511.38

CycD Inactive										
CycD	rB	E2F	CycE	CycA	p27	Cdc2	Cdh1	Ubc	CycB	tC
										7034.13
										148.08
										147.8

MAPK basic model								
M3K	M3K*	M2K	M2K*	M2K**	MK	MK*	MK**	tC
								763.35
								511.38

CycD Active										
CycD	rB	E2F	CycE	CycA	p27	Cdc2	Cdh1	Ubc	CycB	tC
										1393.7
										278.52
										278.52
										278.52

(a)

(b)

Figure 2: The masks identify by the combination of the DCI and the sieving algorithm for (a) the Mitogen Activated Protein Kinase (MAPK) cascade and (b) the two dynamical conditions of the mammalian cell cycle network. In each situation only the masks having significantly high tC values are represented.

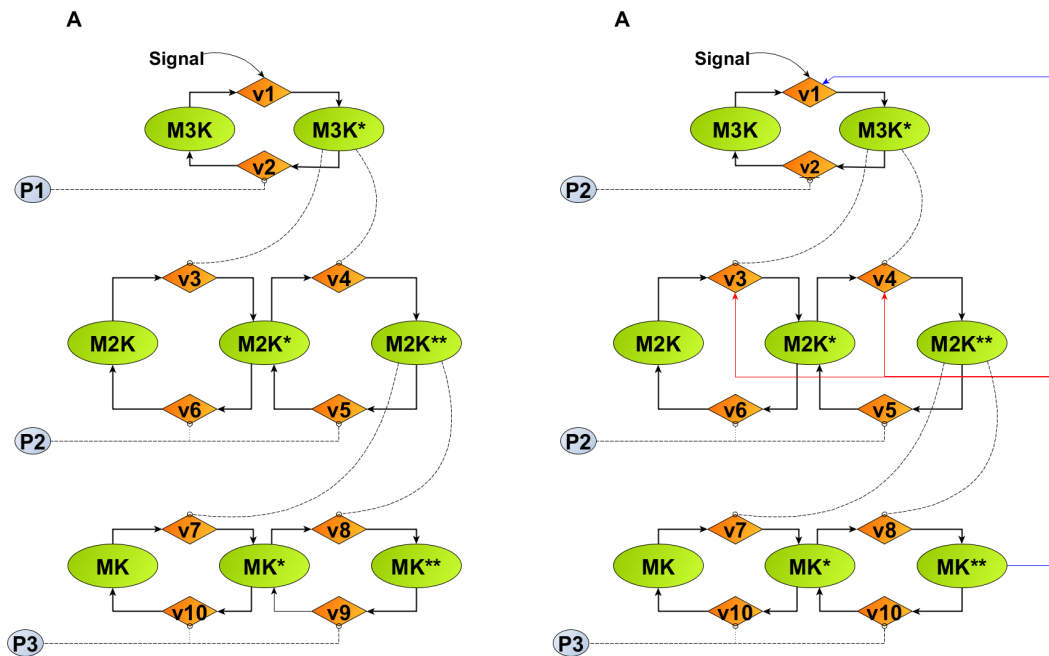


Figure 3: (A) Basic model: the scheme of the three layers MAPK cascade reaction pathway is represented (“*” stands for the phosphorylation). The signal catalyzes the phosphorylation of M3K to M3K* that in turn catalyzes the phosphorylation of M2K to M2K* and the successive phosphorylation of M2K* to M2K**. Finally M2K** performs the double phosphorylation of MK in MK** that is the final output of the MAPK cascade. P1, P2 and P3 dephosphorylate M3K, M2K and MK kinases respectively. V1-V10 stand for the involved reactions. Dashed lines with circle head represent catalysis; the figure highlights the presence of the three “layers” described on the text. (B) Two distinct positive and negative feedbacks are added to the basic model: the negative feedback goes from MK** to the second layer (M2K, M2K* and M2K**) while the positive feedback goes from MK** to the first layer (M3, M3K*).

Acknowledgements

This article has been partially funded by the UE projects “MD, Emergence by Design”, Pr.ref. 284625 under the 7th FWP - FET programme. For their useful comments and suggestions, reviewers are sincerely acknowledged.

References

- Chaos, A., Aldana, M., Espinosa-Soto, C., León, B. G. P., Arroyo, A. G., and Alvarez-Buylla, E. R. (2006). From Genes to Flower Patterns and Evolution: Dynamic Models of Gene Regulatory Networks. *Journal of Plant Growth Regulation*, 25(4):278–289.
- Cover, T. and Thomas, J. (2006). *Elements of Information Theory*.
- Farmer, J. and Kauffman, S. (1986). Autocatalytic replication of polymers. *Physica D: Nonlinear Phenomena*, 220:50–67.
- Fauré, A., Naldi, A., Chaouiya, C., and Thieffry, D. (2006). Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14):124–131.
- Filisetti, A., Graudenzi, A., Serra, R., Villani, M., Füchslin, R. M., Packard, N., Kauffman, S. a., and Poli, I. (2011). A stochastic model of autocatalytic reaction networks. *Theory in biosciences = Theorie in den Biowissenschaften*, pages 1–9.
- Funahashi, A., Matsuoka, Y., Jouraku, A., Morohashi, M., Kikuchi, N., and Kitano, H. (2008). CellDesigner 3.5: A versatile modeling tool for biochemical networks. *Proceedings of the IEEE*, 96:1254–1265.
- Funahashi, A., Morohashi, M., Kitano, H., and Tanimura, N. (2003). CellDesigner: a process diagram editor for gene-regulatory and biochemical networks.
- Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 69(6 2).
- Lautier, D. and Raynaud, F. (2014). Information flows in the term structure of commodity prices. Available at SSRN 2430168.
- Lizier, J. T., Heinzle, J., Horstmann, A., Haynes, J. D., and Prokopenko, M. (2011). Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *Journal of Computational Neuroscience*, 30(1):85–107.
- Lizier, J. T. and Prokopenko, M. (2010). Differentiating information transfer and causal effect. *European Physical Journal B*, 73(4):605–615.
- Sarma, U. and Ghosh, I. (2012). Oscillations in MAPK cascade triggered by two distinct designs of coupled positive and negative feedback loops. *BMC research notes*, 5(1):287.
- Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, 85(2):461–4.
- Serra, R., Villani, M., Graudenzi, A., and Kauffman, S. A. (2007). Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data. *J Theor Biol*, 246:449–460.
- Serra, R., Villani, M., and Semeria, A. (2004). Genetic network models and statistical properties of gene expression data in knock-out experiments. *J Theor Biol*, 227:149–157.
- Shin, S.-Y., Rath, O., Choo, S.-M., Fee, F., McFerran, B., Kolch, W., and Cho, K.-H. (2009). Positive- and negative-feedback regulations coordinate the dynamic behavior of the Ras-Raf-MEK-ERK signal transduction pathway. *Journal of cell science*, 122:425–435.
- Shmulevich, I., Kauffman, S. A., and Aldana, M. (2005). Eukaryotic cells are dynamically ordered or critical but not chaotic. *Proceedings of the National Academy of Sciences of the United States of America*, 102:13439–13444.
- Tononi, G., McIntosh, a. R., Russell, D. P., and Edelman, G. M. (1998). Functional clustering: identifying strongly interactive brain regions in neuroimaging data. *NeuroImage*, 7(2):133–49.
- Villani, M., Barbieri, A., and Serra, R. (2011). A dynamical model of genetic networks for cell differentiation. *PLoS ONE*, 6.
- Villani, M., Filisetti, A., Benedettini, S., Roli, A., Lane, D., and Serra, R. (2013). The detection of intermediate-level emergent structures and patterns. In *Advances in Artificial Life, ECAL 2013*, volume 12, pages 372–378. MIT Press.
- Villani, M., Roli, A., Filisetti, A., Fiorucci, M., Poli, I., and Serra, R. (2015). The search for candidate relevant subsets of variables in complex systems. *In Press on Artificial Life*.
- Villani, M. and Serra, R. (2013). On the dynamical properties of a model of cell differentiation. *EURASIP journal on bioinformatics & systems biology*, 2013:4.
- Widmann, C., Gibson, S., Jarpe, M. B., and Johnson, G. L. (1999). Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiological reviews*, 79(1):143–80.
- Alessandro Filisetti, Marco Villani, Andrea Roli, Marco Fiorucci, Roberto Serra (2015) Exploring the organisation of complex systems through the dynamical interactions among their relevant subsets. *Proceedings of the European Conference on Artificial Life 2015*, pp. 286-293