entropy

MDPI

Article

Asymptotic Information-Theoretic Detection of Dynamical Organization in Complex Systems

Gianluca D'Addese [1], Laura Sani [2], Luca La Rocca [1], Roberto Serra [1,3,4] and Marco Villani [1,3,*]

[1] Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, 41125 Modena, Italy; gianluca.daddese@unimore.it (G.D.); luca.larocca@unimore.it (L.L.R.); rserra@unimore.it (R.S.)
[2] Department of Engineering and Architecture, University of Parma, 43124 Parma, Italy; laura.sani@unipr.it
[3] European Centre for Living Technology, 30123 Venice, Italy
[4] Institute for Advanced Studies, University of Amsterdam, 1012 GC Amsterdam, The Netherlands
* Correspondence: marco.villani@unimore.it

Abstract: The identification of emergent structures in complex dynamical systems is a formidable challenge. We propose a computationally efficient methodology to address such a challenge, based on modeling the state of the system as a set of random variables. Specifically, we present a sieving algorithm to navigate the huge space of all subsets of variables and compare them in terms of a simple index that can be computed without resorting to simulations. We obtain such a simple index by studying the asymptotic distribution of an information-theoretic measure of coordination among variables, when there is no coordination at all, which allows us to fairly compare subsets of variables having different cardinalities. We show that increasing the number of observations allows the identification of larger and larger subsets. As an example of relevant application, we make use of a paradigmatic case regarding the identification of groups in autocatalytic sets of reactions, a chemical situation related to the origin of life problem.

Keywords: chi-squared approximation; cluster index; integration; mutual information; relevance index; relevant subset

Citation: D'Addese, G.; Sani, L.; La Rocca, L.; Serra, R.; Villani, M. Asymptotic Information-Theoretic Detection of Dynamical Organization in Complex Systems. *Entropy* **2021**, *23*, 398. https://doi.org/10.3390/e23040398

Academic Editors: Irad E. Ben-Gal and Amichai Painsky

Received: 15 February 2021
Accepted: 22 March 2021
Published: 27 March 2021

1. Introduction

The identification of emergent structures in complex dynamical systems is a very difficult task with broad applications. In particular, the formation of intermediate-level dynamical structures is of high interest for what concerns biological as well as artificial systems. This phenomenon is among the most intriguing ones in natural as well as in artificial systems, and a fascinating aspect is its "sandwiched" nature [1]. While past emergence examples were focused on bottom-up emergence in two-level systems, like, for instance, Benard–Marangoni convection cells emerging from the interaction of water molecules under the influence of temperature gradients [2], more recent work points out that in many interesting cases the new entities and levels emerge between preexisting ones [1,3,4]. The paradigmatic example may be that of organs and tissues in multicellular organisms: both the lower (cellular) level and the upper (organism) level predate the appearance of the intermediate structures. Other examples come from the physical world (e.g., mesolevel structures in climate), social systems (e.g., factions within political parties and the parties themselves), or socio-technical systems (e.g., communities in social networks). Very often artificial (and sometimes social) architectures have been devised precisely to stimulate the formation of these mesolevel structures, but here we are concerned with structures that come into being by spontaneous processes, even though their formation may be eased or hindered by an external design.

A central question is then that of identifying the emerging "things": these may be either static entities or dynamical patterns, or some mixture of the two. The identification

of these configurations is seldom simple, because of the more-than-binary relationships among variables, the multiple memberships of system entities or the fuzziness of boundaries among groups. In Network Science [5,6], static emergent structures take the form of topological features, like, e.g., motifs in genetic networks or communities in a broader context; in particular, in the case of socio-technical systems there is an extensive literature on community detection [7,8]. However, most methods are based on static features (such as link distributions or topologies), whereas the system's elements may work in a coordinated manner even though they are not directly linked, because of the effects of the dynamical laws. If the topology were regular, these nodes might be identified by visual inspection, but in the case of irregular topologies this approach seems hopeless.

Building on the information-theoretic approach proposed by Tononi [9,10], this paper presents a methodology for the identification of mesolevel objects in a system. As these objects may have a topological as well as a dynamical nature, and they typically play a central role in the system's dynamics, we will refer to them as *relevant subsets*, and we will refer to the indices used for their identification as *relevance indices*. The identification of relevant subsets is a hard task, as discussed above, so we will show here the general schema of a promising approach, compare some indices, and show some results. In particular, in this work we apply our methodology to a relevant case, in which we look for the dynamic organizations (here recognized as groups of variables in dynamic relationship) responsible for the observed behaviors. Specifically, we analyze the results of an experiment frequently performed to examine a system: the observation of its responses when it is subject to solicitations—in our case, distinct groups of autocatalytic chemical reactions in which silencing actions are performed one chemical species at a time. To each solicitation (a silencing) follows a response of the system (a set of chemical species change their activity). We will show that from the observations of the spread of changes it is possible to infer the presence and the composition of dynamically coordinated groups of chemicals, while remarking that there are still open questions to answer.

Our approach will be presented in Section 3, after the necessary background material has been introduced in Section 2, and it will be assessed in Section 4, both from simulation and empirical perspectives; some conclusions will be drawn in Section 5.

2. Methodological Background

Section 2.1 sets our mathematical notation and introduces the information-theoretic concepts used in this paper; see Cover and Thomas [11] for a proper introduction to information theory. Section 2.2 presents the Cluster Index (CI) proposed by Tononi et al. [9] in the context of neuroimaging.

2.1. Information-Theoretic Preliminaries

Let $X_S = \{X_u\}_{u \in S}$ be a random vector indexed by a finite set of elements $S = \{u_1, u_2, \ldots, u_k\}$ in some system of interest. We assume $\emptyset \subset S \subset U$, where $U = \{1, \ldots, q\}$ is the set of all elements in the system and $\emptyset$ denotes the empty set. As a special case, if $S = \{u_1\}$ is a singleton, $X_{\{u_1\}} = X_{u_1}$ is a random variable corresponding to an individual element in the system. The idea is that the vector $X_U = \{X_u\}_{u \in U}$ models a noisy reading of the system's state and by considering X_S we focus on the subsystem formed by the elements in S.

The *entropy* of X_S, assuming a finite alphabet (set of possible values) for all variables, is defined as

$$\mathcal{H}(S) = -\sum_{x_1 \in \mathcal{X}_1} \cdots \sum_{x_k \in \mathcal{X}_k} p_S(x_1, \ldots, x_k) \log p_S(x_1, \ldots, x_k), \tag{1}$$

where $\mathcal{X}_j$ is the alphabet of variable X_{u_j}, for $j = 1, \ldots, k$, and

$$p_S(x_1, \ldots, x_k) = \mathbb{P}\{X_{u_1} = x_1, \ldots, X_{u_k} = x_k\}, \quad x_1 \in \mathcal{X}_1, \ldots, x_k \in \mathcal{X}_k, \tag{2}$$

is the probability mass function of X_S. The base of the logarithm used in (1) determines the unit of information: a *bit*, if binary logarithms are used, or a *nat*, if natural logarithms are used. If $\mathcal{H}_2$ denotes the entropy in bits, the change of logarithm base formula gives $\mathcal{H}_e = \ln(2)\mathcal{H}_2$ for the entropy in nats; more generally, if $\mathcal{H}_b$ is the entropy with logarithm base b, then $\mathcal{H}_b = \log_b(e)\mathcal{H}_e$. In the following, we use nats (unless noted otherwise).

Following Tononi et al. [10], when $S = \{u_1, \ldots, u_k\}$ with $k \geq 2$, we define the *integration* of X_S as

$$\mathcal{I}(S) = -\mathcal{H}(S) + \sum_{j=1}^{k} \mathcal{H}(u_j), \tag{3}$$

where $\mathcal{H}(u_j)$ is the entropy of the individual variable X_{u_j}. It can be shown, by means of the chain rule of probability, that $\mathcal{I}(S)$ is always positive and vanishes if $X_{u_1}, \ldots X_{u_k}$ are stochastically independent. Watanabe [12] introduced (3) as a measure of "total correlation" among the variables indexed by S. In the special case $k = 2$, the integration of $X_{\{u_1,u_2\}}$ reduces to the *mutual information* between X_{u_1} and X_{u_2}; see Cover and Thomas ([11], Ch. 2).

Now let $X_S^{(1)}, \ldots, X_S^{(n)}$ be a random sample of observations from the unknown distribution of X_S. We estimate the distribution of X_S by means of the *empirical distribution* of $X_S^{(1)}, \ldots, X_S^{(n)}$, whose probability mass function $\hat{p}_S(x_1, \ldots, x_k)$, $x_1 \in \mathcal{X}_1, \ldots, x_k \in \mathcal{X}_k$, is given by the relative frequencies of the possible values of X_S in the sample. Using this distribution in (1), we obtain the *empirical entropy* of $X_S^{(1)}, \ldots, X_S^{(n)}$, which we denote by $\mathcal{H}_n(S)$. Similarly, using empirical entropies in (3), we define the *empirical integration* $\mathcal{I}_n(S)$ of $X_S^{(1)}, \ldots, X_S^{(n)}$. More elaborate entropy estimators are available [13–15], but in this work we focus on making the most out of the simplest one.

It has been known since Miller and Madow [16] that, on average, the empirical entropy $\mathcal{H}_n(S)$ underestimates its theoretical counterpart $\mathcal{H}(S)$:

$$\mathbb{E}[\mathcal{H}_n(S)] = \mathcal{H}(S) - \frac{c(S) - 1}{2n} + O\left(\frac{1}{n^2}\right), \tag{4}$$

where $c(S) = \prod_{j=1}^{k} |\mathcal{X}_j| > 1$ is the number of cells in the table indexed by S and it is understood that $p_S(x_1, \ldots, x_k) > 0$ for all $x_1 \in \mathcal{X}_1, \ldots x_k \in \mathcal{X}_k$; note that $|\mathcal{X}_j|$ denotes the cardinality of $\mathcal{X}_j$. It follows that the empirical integration overestimates its theoretical counterpart:

$$\mathbb{E}[\mathcal{I}_n(S)] = \mathcal{I}(S) + \frac{d_k}{2n} + O\left(\frac{1}{n^2}\right), \tag{5}$$

where

$$d_k = c(S) - 1 + k - \sum_{j=1}^{k} |\mathcal{X}_j| \tag{6}$$

is a strictly positive integer. In the special case $k = 2$, first studied by Miller [17], the quantity in (6) can be written as $d_2 = (|\mathcal{X}_1| - 1)(|\mathcal{X}_2| - 1)$; see Luce [18].

2.2. Cluster Index

The Cluster Index (CI) is an information-theoretical measure proposed in the 1990s by Tononi and Edelman [9,10], within researches on human brain processes, whose purpose is the identification of subsets of variables in a dynamical system that behave in a coordinated way, while having a relatively limited exchange of information with the rest of the system; these subsets can then be used to describe the whole system organization. Given a subset $S = \{u_1, \ldots, u_k\}$ of elements in a system indexed by $U = \{1, \ldots, q\}$, where $1 < k < q$, Tononi et al. [10] defined the integration of S as in (3) and considered it as a proxy for the degree of coordination among the k variables indexed by S. Then, Tononi et al. [9] considered the mutual information between S and $U \setminus S$

$$\mathcal{M}(S) = \mathcal{H}(S) + \mathcal{H}(U \setminus S) - \mathcal{H}(U) \tag{7}$$

as a measure of the mutual dependence between the subset S and the rest of the system $U \setminus S$, and defined the CI as the ratio between the integration of S and the mutual information between S and $U \setminus S$:

$$\mathrm{CI}(S) = \frac{\mathcal{I}(S)}{\mathcal{M}(S)}. \tag{8}$$

In practice, analyses resort to the empirical version $\mathrm{CI}_n(S) = \mathcal{I}_n(S)/\mathcal{M}_n(S)$, based on the relative frequencies observed in a sample of system states.

High values of (8) correspond to subsets of U where the internal coordination exceeds the exchange of information with the rest of the system, allowing in such a way the identification of interesting groups of variables. As the value of the CI depends on the size of the group under examination, as well as on the size of the system, this index should be normalized before groups of different size can be compared. Tononi et al. [10] proposed as normalizing constants for the numerator and denominator of (8) the averages of integration and mutual information over groups of equal size embedded in a system, called the *homogeneous system*, with the same number of variables and no dynamical organization:

$$\mathrm{CI}^{\star}_n(S) = \frac{\mathcal{I}_n(S)/\langle \mathcal{I}_n(k) \rangle_0}{\mathcal{M}_n(S)/\langle \mathcal{M}_n(k) \rangle_0}. \tag{9}$$

Eventually, the z-score of the normalized CI (numerically identical to the z-score of the CI) can be used as an evidence index:

$$z\mathrm{CI}_n(S) = \frac{\mathrm{CI}^{\star}_n(S) - \langle \mathrm{CI}^{\star}_n(k) \rangle_0}{\mathrm{sd}_0(\mathrm{CI}^{\star}_n(k))} = \frac{\mathrm{CI}_n(S) - \langle \mathrm{CI}_n(k) \rangle_0}{\mathrm{sd}_0(\mathrm{CI}_n(k))}; \tag{10}$$

see Villani et al. [19]. Finally, and interestingly, the zCI index is related to the identification of dynamical criticality in complex systems, see Roli et al. [20,21].

3. Proposed Methodology

In Section 3.1, we discuss some limitations of the Cluster Index, and we propose a methodology to overcome these limitations. The methodology requires the use of indices to evaluate groups of interest: the identification and comment of these indices is the focus of the current work. In Section 3.2, we show that in homogeneous systems the empirical integration follows an asymptotic chi-squared distribution: this knowledge can be used in different ways, generating different indices. In Section 4, these indices will be applied in different situations, which will allow us to identify the z-score of the integration as the most effective one.

3.1. Searching for Relevant Subsets

3.1.1. Challenges

Implementing the CI index presents two main challenges:

(a) The cardinality of the set of all possible subsets of a set is gigantic. However, even beforehand the computational effort needed to deal with this amount of data, it is noteworthy that this wide set contains many groups included in others and a huge number of partially overlapping groups. All these situations require further analyses to assess their actual relevance or independence. Indeed, a high index value is not sufficient to characterize a relevant subset, because such a value might result from the presence of a smaller subset characterized by a higher coordination among variables. Conversely, a set having a high index value might reach an even higher value, if some other relevant variables are added to it.

(b) It is burdensome to compute the averages of integration and mutual information on a suitable homogeneous system. Even though simulations from the homogeneous system are straightforward, they have to be repeated for all subsets of interest, which results in very long computing times. Furthermore, a specific homogeneous system

has to be selected for the simulations, which introduces an unwelcome degree of arbitrariness in the analysis.

We address challenges (a) and (b) by using the integration alone (through its z-score, zI in the following) within an iteration scheme like the one presented by Villani et al. [22]. This results in an efficient framework for the identification and subsequent enlargement of dynamically interesting groups.

Use of integration alone enables us to deal with challenge (b) by means of an asymptotic approximation that holds for all systems with independent variables and does not require to simulate from any of them: as shown in Section 3.2, for large n, the integration, multiplied by twice the number of observations, approximately follows a chi-squared distribution with degrees of freedom depending on the number of variables belonging to the analyzed subgroup and on the cardinality of their alphabets; such an approximation can be used to obtain z-scores with negligible computational effort.

Using the integration alone also deals with a problematic aspect of the division by mutual information: low values of $\mathcal{M}(S)$ can derive from a low information exchange between the subgroup S and every element of the rest of the system, or from a high exchange of information between S and a small part of the rest of the system, while the other parts are not involved in the exchange. The (z-score of the) CI index does not distinguish between the two situations [23].

3.1.2. The Iterative Sieving Method

Challenge (a) requires a procedure to compare different subgroups. We suggested [24,25] a comparison procedure (sieve, or sieving algorithm, in the following) based on the consideration that if a set A is a proper subset of a set B and ranks higher than B, then A should be considered more relevant than B. Therefore, the sieve keeps only those sets that are not included in, nor include, any other set with higher zI. The comparison procedure is therefore composed by two basic steps: (i) detection of relevant variable sets based on the computation of the zI metric and (ii) application of the sieving algorithm, which refines the results. This approach allows one to identify a plausible organization of the system in terms of non-overlapping groups of variables [24,25].

In order to analyze the hierarchical organization [4] of the system under examination, we propose an iterative version of the sieving method that groups one or more sets into a single entity to derive a hierarchy. The simplest, yet effective, way to do so consists in iteratively running the sieving algorithm on the same data, each time using a new representation in which the top-ranked relevant subset of the previous iteration, in terms of zI values, is considered as atomic and is substituted by a single variable (group variable). Each iteration produces therefore a new atomic group of variables: the iterations end when the zI of the top-ranked relevant subset falls below a preset threshold, usually equal to 3.0, that is, three standard deviations from the reference condition of variable independence [22]; see Appendix A for details.

The iterative sieving approach highlights the organization of a dynamical system by partitioning it into sets of variables detected at different iterations of the sieve. At the same time, the process of aggregation, by adding new elements to the already existing groups, allows the procedure to identify the variables of "the rest of the system" that exchange information with the subset originally under examination, discriminating in such a way between the two problematic situations of low mutual information describe above. Iteration of the "sieve and subsequent aggregation of variables" process thus allows to identify the parts of the system that can be aggregated.

Last, but not least, we note that we are not interested in analyzing all the subsets that can be drawn from the system in question, but rather we want to identify the subsets having the maximum values of the chosen index. It is therefore possible to use optimization procedures, which have the aim of finding the best values without having to go through a complete enumeration. For this purpose, we have used several heuristics in the past, includ-

ing suitable variants of genetic algorithms [25,26]. Finally, it is possible to (at least partially) deal with the system's curse of dimensionality by using parallelization strategies [27].

3.2. Asymptotic Null Distribution of the Empirical Integration

As anticipated, we are interested in the computation of the distribution of the empirical integration $\mathcal{I}_n(S)$ when the variables indexed by S are stochastically independent (null distribution, or homogeneous system distribution). We will find an asymptotic (large sample) approximation that does not depend on the marginal distributions of $X_{u_1}, \dots X_{u_k}$, that is, on the specific homogeneous system chosen as a benchmark for the lack of coordination among variables.

The key observation is that the counts $n_S(x_1, \dots, x_k) = n\hat{p}_S(x_1, \dots, x_k)$ form a multinomial random vector with size n and class probabilities $p_S(x_1, \dots, x_k)$. This gives rise to the likelihood

$$\mathcal{L}_n(p_S) \quad \propto \quad \prod_{x_1 \in \mathcal{X}_1} \cdots \prod_{x_k \in \mathcal{X}_k} p_S(x_1, \dots, x_k)^{n_S(x_1, \dots, x_k)}, \tag{11}$$

where the omitted proportionality constant is the multinomial coefficient that counts the ways to group n observations into $c(S)$ cells with $n_S(x_1, \dots, x_k)$ observations in cell $x_1, \dots, x_k$.

If the probabilities $p_S(x_1, \dots, x_k)$ are free to vary in the standard simplex, the maximizer of (11) consists of the relative frequencies $\hat{p}_S(x_1, \dots, x_k)$; see, for instance, Held and Sabanés Bové ([28], Ch. 5). It follows that the empirical entropy $\mathcal{H}_n(S)$ equals, up to an additive constant, the negative maximized average log-likelihood $-\bar{\ell}_n(\hat{p}_S) = -n^{-1} \log \mathcal{L}_n(\hat{p}_S)$:

$$\bar{\ell}_n(p_S) \quad \doteq \quad \sum_{x_1 \in \mathcal{X}_1} \cdots \sum_{x_k \in \mathcal{X}_k} \hat{p}_S(x_1, \dots, x_k) \log p_S(x_1, \dots, x_k), \tag{12}$$

where $\doteq$ denotes equality up to an additive constant.

On the other hand, if the constraint $p_S(x_1, \dots, x_k) = \prod_{j=1}^{k} p_{u_j}(x_j)$ is introduced, where $p_{u_j}(x_j) = \mathbb{P}\{X_{u_j} = x_j\}$, that is, the variables indexed by S are assumed to be stochastically independent, the likelihood (11) can be written as

$$\mathcal{L}^0_n(p_{u_1}, \dots, p_{u_k}) \quad \propto \quad \prod_{x_1 \in \mathcal{X}_1} p_{u_1}(x_1)^{n_{u_1}(x_1)} \cdots \prod_{x_k \in \mathcal{X}_k} p_{u_k}(x_k)^{n_{u_k}(x_k)}, \tag{13}$$

where $n_{u_j}(x_j)$ is the count for the value x_j in the marginal sample $X^{(1)}_{u_j}, \dots X^{(n)}_{u_j}$, for $j = 1, \dots, k$. The maximizer of (13) clearly consists of the marginal relative frequencies $\hat{p}_{u_j}(x_j) = n_{u_j}(x_j)/n$, and it is apparent that $\sum_{j=1}^{k} \mathcal{H}_n(u_j)$ equals $-\bar{\ell}^0_n(\hat{p}_{u_1}, \dots \hat{p}_{u_k}) = -n^{-1} \log \mathcal{L}^0_n(\hat{p}_{u_1}, \dots, \hat{p}_{u_k})$ up to the same additive constant as before.

By virtue of a classical theorem of mathematical statistics, due to Wilks [29], assuming natural logarithms are used, the likelihood-ratio test statistic

$$\Lambda_n = 2n(\bar{\ell}_n(\hat{p}_S) - \bar{\ell}^0_n(\hat{p}_{u_1}, \dots \hat{p}_{u_k})),$$

for large n, approximately follows a chi-squared distribution with degrees of freedom given by the difference in dimensions between the unconstrained and the independence statistical models, that is, equal to d_k in (6); see ([28], Ch. 5) for a modern presentation of this result from an applied viewpoint. As $\mathcal{I}_n(S) = \bar{\ell}_n(\hat{p}_S) - \bar{\ell}^0_n(\hat{p}_{u_1}, \dots \hat{p}_{u_k})$, the likelihood-ratio statistic can be written as $2n\mathcal{I}_n(S)$ and we have an asymptotic distribution for the empirical integration. Note that, by construction, the likelihood-ratio test statistic is positive and vanishes when $\hat{p}_S(x_1, \dots, x_k) = \prod_{j=1}^{k} \hat{p}_{u_j}(x_j)$, which confirms the same properties for $\mathcal{I}_n(S)$.

In the special case of mutual information between two variables, where the degrees of freedom can be written as $d_2 = (|\mathcal{X}_1| - 1)(|\mathcal{X}_2| - 1)$, the result presented above dates back to Luce [18] with a direct justification in pioneering work by Wilks [30]. If bi-

nary logarithms are used, the asymptotic chi-squared distribution applies to the statistic $2n\ln(2)\,\mathcal{I}_n(S) \simeq 1.3863n\,\mathcal{I}_n(S)$; Wilks [18] uses the bit as unit of information and states the result in this form. In general, if b is the base used for logarithms, we can write $2n\ln(b)\,\mathcal{I}_n(S) \approx \text{Chisq}(d_k)$, where $\approx$ means that the left hand side has the right hand side as asymptotic distribution (approximate distribution for large n).

As a chi-squared distribution with d_k degrees of freedom has mean d_k and variance $2d_k$, a standardized version of the empirical integration is given by

$$z\mathrm{I}_n(S) \quad = \quad \frac{2n\ln(b)\,\mathcal{I}_n(S) - d_k}{\sqrt{2d_k}}, \tag{14}$$

where $b = e$ for nats and $b = 2$ for bits. Note that the approximate z-score in (14) is consistent with the bias assessment in (5), but it also deals with sample variability; it accounts for the sample size n and, through d_k, for the number variables indexed by S and their alphabet sizes. Alternatively, one can compare subsets of different dimension using the simple normalized index $2nI/d = 2n\mathcal{I}_n(S)/d_k$, or one minus the chi-squared p-value of $2n\mathcal{I}_n(S)$, denoted by ChSq. In Section 4, we will compare the z-score in (14), $2nI/d$, and ChSq to the Cluster Index.

4. Results

In order to assess the correctness of our approach on the one hand, and to acquire new knowledge on the organization of dynamic systems on the other, we apply our relevance index methodology, or RI methodology in the following, to two very different cases. The first situation, considered in Section 4.1, involves systems in which there is no dynamic organization: each variable completes its own trajectory (in this case a random trajectory) regardless of the behavior of the other variables. By contrast, in Section 4.2, the second set of observations comes from a highly organized system, which presents two peculiar dynamic structures, capable of providing (within certain limits) self-maintenance.

4.1. Dynamically Homogeneous Systems

In this section, we show the results of the analysis of a set of trajectories, having different lengths, extracted from a homogeneous system composed of 21 binary variables whose two symbols have equal probability of appearing. This provides a paradigmatic example of the behavior of the RI methodology when it is applied to systems having no dynamic organization.

Figure 1 displays the average integration by group size for trajectories having different lengths. For intermediate group sizes it was not feasible to consider all possible groups: if the number of groups with a given size exceeded the threshold 10,000, we performed a sampling and used only 10,000 randomly extracted groups (for each analyzed size). We can observe that shorter trajectories, as well as larger groups, lead to higher average integration, which reflects the bias term $d_k/2n$ in (5). Indeed, as illustrated in Figure 2, the average values of $2n\mathcal{I}_n(S)$ do not depend on n for small group sizes, which shows that in such cases the chosen trajectory lengths are sufficient to provide a reliable estimate of integration. Furthermore, it can be seen in Figure 3 that the average values of $2n\mathcal{I}_n(S)/d_k$ are close to 1.0 for small groups sizes, which confirms the reliability of such estimates. Note that the maximum group size, such that the estimated integration is reliable, grows with the number of observations, from $k = 7$ with $n = 50$ observations to $k = 14$ with n = 10,000 observations, which is consistent with the fact that d_k can be seen as the average of $2n\mathcal{I}_n(S)$ with infinite observations. Note as well that the width of the error bars in Figure 3 is greater for smaller group sizes, where there are few distinct groups with little overlap between them, while it drops monotonically with the increasing length of the trajectories.

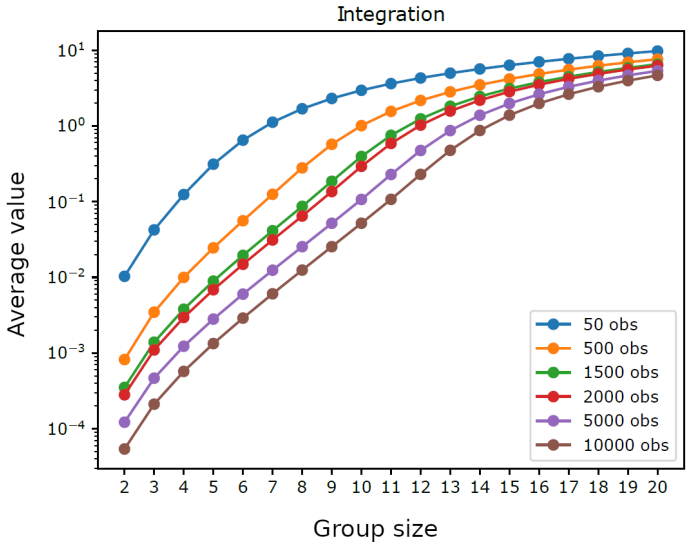


Figure 1. Homogeneous system. Average integration by analyzed group size ($k = 2, 3, \ldots, 19, 20$) with varying trajectory length ($n = 50$, 500, 1500, 2000, 5000, and 10,000).

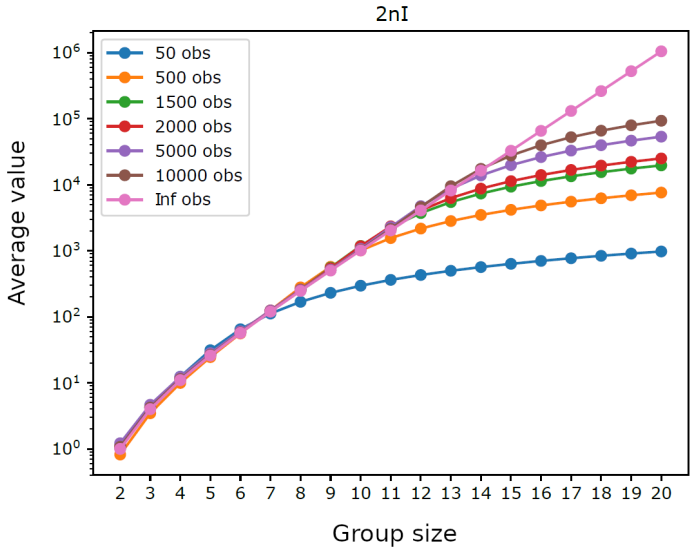


Figure 2. Homogeneous system. Average of $2n\mathrm{I} = 2n\mathcal{I}_n(S)$, by group size $k = |S|$, for six different trajectory lengths, compared with the mean $d = d_k$ of the chi-squared distribution (average with infinite observations).

Figure 3 shows that the average of $2n\mathcal{I}_n(S)/d_k$ quickly drops below the value 1.0 for large groups. This fact indicates that, in the trajectories examined, the values of $2n\mathcal{I}_n(S)/d_k$ for these groups are small compared to what they should be in order to thoroughly observe large groups: large size groups are not very evident, or, in other words, they are difficult to detect. Indeed, the same limit holds for groups formed by a few variables having a large number of symbols. Actually, systems with a number of degrees of freedom greater than the number of observations necessarily cannot be fully observed, which results in a fundamental limit regarding the reliability of their identification. As as special case, this limit holds also for the group variables formed by the aggregation of simpler variables happening during to the iteration of the sieving algorithm, a process which creates new group variables with a number of symbols adequate to maintain all the information carried by the original ones. In such a way, groups formed by few variables with a large number of internal levels maintain the same number of degrees of freedom as groups formed by a large number of simpler variables, proposing again the critical situation. To sum up,

however, this limitation depends on the number of performed observations rather than on the method used for the detection of groups.

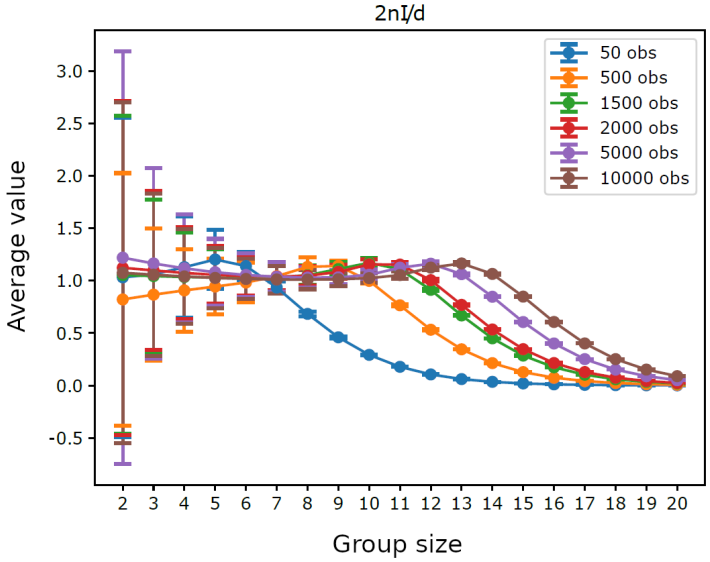


Figure 3. Homogeneous system. Average of $2n\mathrm{I}/d = 2n\mathcal{I}_n(S)/d_k$, by group size $k = |S|$, for six different trajectory lengths, with error bars at three times the standard deviation of observations.

The above-described limit puts a natural end to the process of iterative grouping described in Section 3.1. However, even before that limit is reached, the process should be stopped if larger interacting groups are not present. This will be done on the basis of a method-specific assessment of the probability of making a mistake during the progress of the agglomeration process. To this aim, in the following, we study the performance of our proposed relevance index, the approximate z-score of integration (zI) given by (14), in comparison with three other relevance indices, which can replace it in the sieve: the z-score of Tononi's Cluster Index (zCI), the simple normalized index $2n\mathrm{I}/d = 2n\mathcal{I}_n(S)/d_k$, and one minus the chi-squared p-value of $2n\mathcal{I}_n(S)$, denoted by ChSq.

Figure 4 shows the maximum values of the four indices for each group size, sorted by group size. Basically, in a homogeneous system, there should not be any relevant subset, although some will spuriously show up due to the finiteness of the number of observations. A high threshold on the relevance index should be able to limit the frequency of such spurious appearances. In this respect, the plot in Figure 4 suggests that ChSq is unfit to the task, because it gets very close to its upper bound for all interesting group sizes. Furthermore, it can be seen from Figure 4 that the variability of $2n\mathrm{I}/d$ decreases with group size, which makes the emergence of smaller spurious groups more frequent (for a given threshold); this is clearly an undesirable feature. On the other hand, it appears that zCI and zI have comparable maxima across interesting group sizes; recall, however, that the computation of zCI is onerous, whereas that of zI is not, which makes zI preferable to zCI in practice.

Figure 5 focuses on the distribution of zCI and zI in the smallest groups. It can be seen that each group size presents several extreme values, which indicates that several groups could spuriously exhibit some coordination activity. Remarkably, within each group size, the two indices provide the same ranking (convey the same information). We shall see in the next section that this is not necessarily the case for non-homogeneous systems.

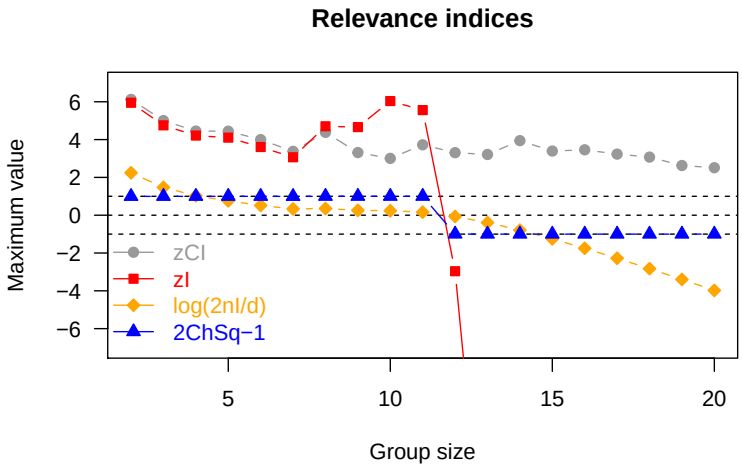


Figure 4. Homogeneous system. Maximum values of the four relevance indices under comparison (*z*CI, *z*I, $2nI/d$, ChSq). The indices $2nI/d$ and ChSq are transformed so that they have zero as the null value (like *z*CI and *z*I). The horizontal dashed lines mark the lower and upper bounds for ChSq, and the null level. The dramatic drop in the indices based on the chi-squared distribution after dimension 12 indicates that the approximation used (that of assuming infinite observations) is no longer valid. Such knowledge is not available in the case of the *z*CI index, which we calculated in each dimension resorting to a very onerous bootstraps procedure.

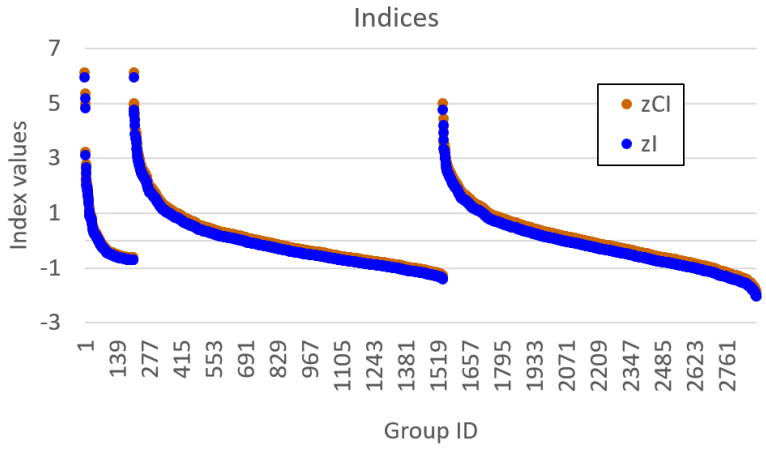


Figure 5. Homogeneous system. Values of *z*CI and *z*I for each single group of up to four variables, sorted by group size and then by *z*I.

4.2. Dynamically Organized Systems

In the previous section, we used data from a homogeneous system to verify the effectiveness of the chi-squared distribution and to make some general remarks on the application of our relevance index to the evaluation of the degree of dynamic organization of subgroups of variables. Our goal, however, is to look for groups that are dynamically relevant (relevant subsets), whose identification can facilitate the understanding of the system's dynamic organization. To this aim, the homogeneous system plays the role of a yardstick, while we are interested in analyzing dynamically non-homogeneous systems.

There are two different ways of searching for relationships present within a dynamical system. The first strategy consists in juxtaposing several separate instances of the same organization; the other strategy is that of observing one particular trajectory, possibly disturbing it from time to time. Indeed, as we are adopting entropic measures, we do not actually make use of the hypothesis of having all states successor of one another: changing

the order of the observations does not influence the frequency of each state. We can therefore use the same framework for both situations. Indeed, we analyzed several systems with strong dynamical organization, by juxtaposing the states belonging to different asymptotic behaviors of the same system (different attractors of a genetic regulatory network [20,31,32] and patients affected by the same kind of disease [33]) or by observing the trajectory of a single system (a socio-economic system [34]), sometimes perturbing it (metabolic networks [24] and autocatalytic systems [19,35]). The performed RI analyses show some common characteristics, so in this paper we choose to expose them by commenting in detail a particular system: an autocatalytic reaction network introduced in [35].

The situation concerns the formation of groups of molecules able to collectively catalyze and self-replicate, a process that is thought to be fundamental for the origin of life [36–42] and is likely to play an important role also in future biotechnological systems [43]. Indeed, currently living beings are based on self-replicating chemical structures, where the presence of enzymes (biological catalyzers) plays an essential role. A useful representation of such systems is based on Reflexive Autocatalytic Food (RAF)-generated sets [44,45], a sophisticated description recently utilized in biochemical contexts [44,46–48] or in protocell architectures [49] to characterize structures with different kind of interactions (production and catalysis).

In this paper, we deterministically simulate two particular instances of RAFs: a linear chain of reactions, having the root with existence guaranteed from the outside, and a ring, in which the substances produced are catalysts for at least one of the other substance to be produced. The two structures are immersed in a Continuous-flow Stirred-Tank Reactor (CSTR) [50] featuring a constant influx of feed molecules, constantly present in the incoming flow of the CSTR and therefore playing the role of the "food" species at the base of RAF arrangements, and a continuous outgoing flux of all the molecular species proportional to their concentration. The simulations are based on a relatively simple system inspired by a model used in [51–53] and originally proposed by Kauffman [41,54]. The scheme simulated in this paper, represented in Figure 6, involves only enzymatic condensations, decomposed in three steps: the first two steps create (and destroy in an overall reversible process) a temporary complex, composed by one of the two substrates (the "first substrate") and the catalyst, which is combined in the third step with the other substrate to release the catalyst and the final product; see in [49] for a more accurate description. The dynamics of the systems are simulated adopting a deterministic approach: the reaction scheme is translated into a set of ordinary differential equations ruled by the mass action law [49,55] and integrated by means of an Euler method with step-size control. Figure 6 represents the simulated system and Figure 7 illustrates its dynamic behavior.

The asymptotic behavior of this kind of system is a single fixed point [56], which does not provide any useful observation for identifying the underlining dynamic structure; in order to apply our analysis, indeed we need to observe the feedbacks in action. We then follow a perturbative approach, consisting in disturbing the asymptotic behavior and recording the consequent transient: we temporarily lower, one by one, by two orders of magnitude, the input concentrations of the food species (green ellipses in Figure 6) after the system has reached its stationary state. In order to analyze the system response to perturbations, we use a three-level coding where, for each species, the digits 0, 1, and 2 stand for "concentration decreasing", "no change", and "concentration increasing", respectively. Specifically, in this experiment, we consider the concentration of a chemical species as being constant if it has not changed by more than 1% in a given time period (ten seconds in the example of Figure 7).

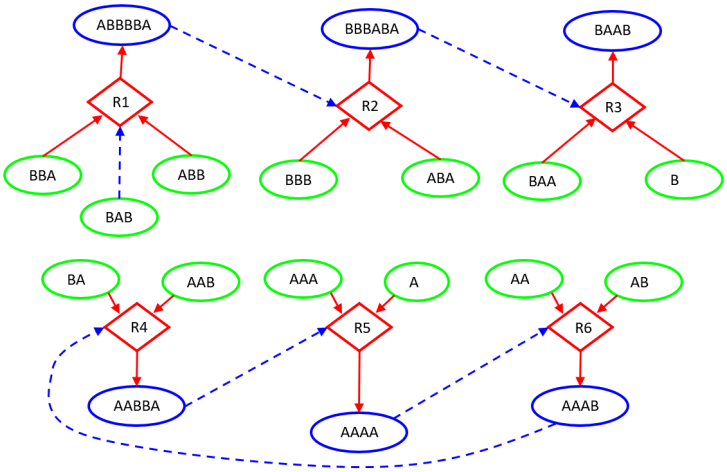


Figure 6. CSTR26 system. The chemical system under analysis. Circular nodes depict chemical species: the green ones stand for those injected into the Continuous-flow Stirred-Tank Reactor (CSTR) (food species) and the blue ones represent the more complex species built by specific concatenations of the food species. Diamond shapes represent reactions, where incoming arrows go from substrates to reactions and outgoing arrows go from reactions to products. Dashed lines indicate the catalytic role of a particular molecular species within the specific reaction context. For instance, thanks to the catalyst BAB, reaction R1 combines the food species ABB and BBA into the complex ABBBBA, while reaction R3 combines the food species BAA and B into the complex BAAB, when the catalyst BBBABA is present.

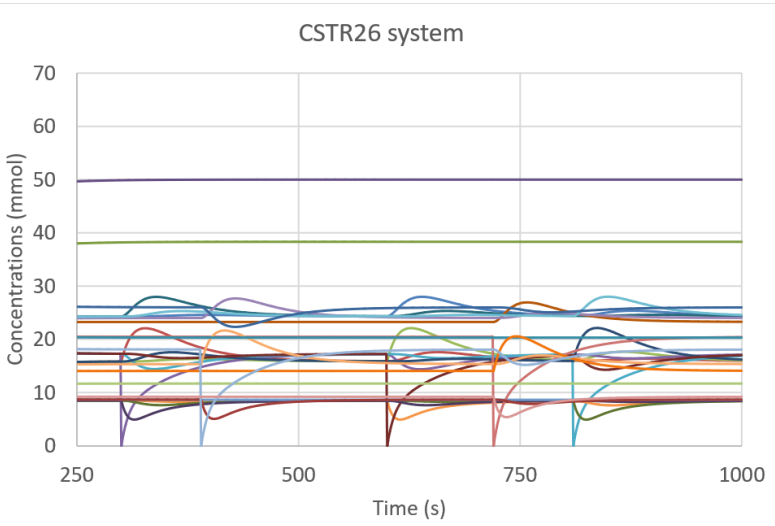


Figure 7. CSTR26 system. The system behavior within the time interval under analysis, including the externally generated perturbations in order to stimulate the dynamic response of the system. The kinetic constants of all present reactions have the same value $k = 0.0025\ \mathrm{s}^{-1}\ \mathrm{mol}^{-1}$; the incoming concentration of each food species is 0.001 M, while each second 2% of the CSTR volume is renewed.

Let us now analyze the findings of our algorithm. Figure 8 represents the distribution of zI, zCI, $2nI/d$, and ChSq in the smallest groups of variables. The indices $2nI/d$ and ChSq show no discriminatory power. The indices zI and zCI no longer provide the same information, as it happened in the homogeneous case. In particular, a detailed group by group analysis has shown that the presence of numerous dynamically organized assemblies that interact with each other does not allow the zCI index to correctly identify the relevant parts. As an example of such interactions, in Figure 6, the couple of variables BBB and ABA is in strong association (with different intensities) with the variation of concentrations of

ABBBBA and BBBABA, or with the variation of concentrations of BAA or BAAB. In the present case, as previously commented, low values of mutual information can derive from a low information exchange between the subgroup and every element of the rest of the system or from a high exchange of information between the subgroup and a small part of the rest of the system, while the other parts are not involved in the exchange. The two situations have to be discriminated, but the mutual information is not providing the correct information to satisfy the need. On the other hand, if we consider the iterated sieve guided by the values of zI, we observe that the process of growing already existing strong groups makes it possible to gradually identify the variables to be aggregated, avoiding the difficulty and discriminating in such a way the situations otherwise giving almost similar and low mutual information values. For these reasons, the index zI appears to be preferable to the index zCI.

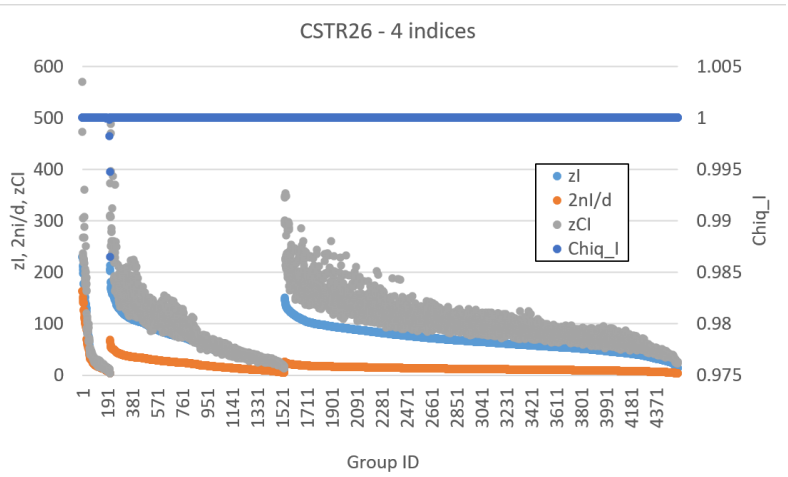


Figure 8. CSTR26 system. Values of zCI, zI, $2nI/d$, and ChSq (right vertical axis) for each single group of up to four variables, sorted by group size and then by zI.

Figures 9 and 10 illustrate the analysis of the CSTR system using the zI index. Remarkably, the final groups (the relevant sets object of the research) correctly identify the two dynamic organizations we have included in the system (Figure 10). The three variables—BBA, BAB, and ABB—were not significantly involved in any perturbation event, and they are correctly outside any group. The presence of subgroups within the relevant sets indicates the presence of a hierarchy: the identified dynamic organizations are composed of smaller parts. In the case of an unknown system, the search for relations between the parts will be investigated by experts in the field. In our case the ground truth is known, and we can appreciate the order in which the algorithm evaluates the coordination evidences. First, as shown in Figure 9, the analysis merged the catalyst–"first substrate" pairs. Recall that, in the system under examination, the catalytic action is carried out through the formation of a short-lived active complex, composed by the catalyst and one of the substrates (the "first substrate"). If later on the complex meets the other substrate (the "second substrate"), the reaction proceeds releasing the catalyst and producing the final product; otherwise, the complex dissociates releasing the two species of which it is composed.) (actually having very strong relationships within the system) and the terminal pair of the linear chain (each perturbation of a chemical species belonging to this system actually producing a coordinated signal in this pair). After that, the "second substrate" was added to each group with catalyst, and finally further mergers were made, until the analysis reached the threshold for zI and stopped in the situation of Figure 10.

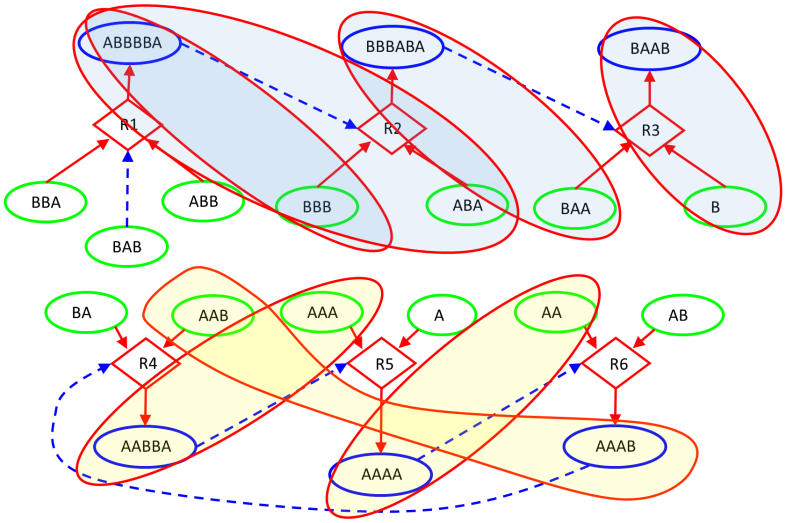


Figure 9. CSTR26 system. The first groups identified by our RI analysis: catalyst-first substrate pairs, and the terminal pair of chemicals within the linear chain of reactions.

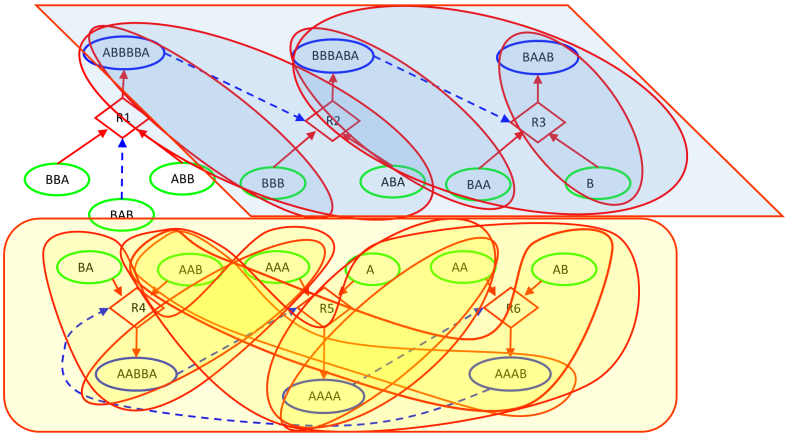


Figure 10. CSTR26 system. The final situation identified by our RI analysis, in which the action of the iterated sieve induces the formation of a hierarchy of encapsulated groups.

5. Conclusions

A formidable challenge in Complexity Science is that of identifying emergent organizations in complex dynamical systems, a theme with broad applications. A central question is then that of detecting the emerging structures: these may be either static entities or dynamical patterns, or some mixture of the two. Identifying these configurations is seldom simple, because of the more-than-binary relationships among variables, the possible multiple memberships of system entities, or the fuzziness of boundaries among groups. A large part of the current approaches aiming at the detection of these objects is based on network representations and static features, such as link distributions or topologies, whereas the system's elements may work in a coordinated manner even though they are not directly linked, because of the dynamical laws governing the system.

In this paper, we presented a methodology for the identification of mesolevel objects, which we call relevant subsets, based on entropic measures, which may involve dynamical aspects [19,24,27,34] or juxtapose different realizations within a population of individuals sharing the same common organization [20,31,33]. We identified an entropic measure useful for the detection of relevant subsets and studied its theoretical distribution, a fact that helps in the interpretation of the results and allows to avoid the excessively onerous bootstrap calculations from a homogeneous system that are needed to compare groups

of different size. Finally, we showed that the increase in the number of observations allows the identification of larger and larger groups, up to the asymptotic case of complete observability in the case of infinite observations. As an example of application, we analyzed a paradigmatic case regarding the identification of autocatalytic sets of reactions, a chemical situation related to the origin of life problem.

The general schema for the identification of relevant subsets presented in this paper is a promising approach for a difficult task: we showed some interesting results, while remarking that there are still open questions to answer. A delicate aspect concerns the number of observations necessary to identify large groups, which makes it useful to search for statistical corrections in case of few of them. In this regard, more elaborate entropy estimators [13–15] could be helpful. Furthermore, it should be observed that the current version of our method is based on the classical measure introduced in information theory by Claude Shannon (and sometimes referred to as the Boltzmann–Gibbs–Shannon entropy). It is well known that different definitions of entropy have been proposed including those of Tsallis and Renyi, which have interesting features, while lacking the additivity of Shannon's; see, e.g., in [57] for a recent review. We think that a generalization of the RI method based on these nonadditive entropies might lead to interesting results, in particular in complex systems with long-range interactions. Finally, it can be observed that the iterated sieving algorithm presupposes an at least partial decomposability of the dynamical organization into separate parts. On the other hand, the iterated and progressive recomposition decreases, at each iteration, the number of degrees of freedom of the model representing the system, and if correct it could allow to identify more easily (or with fewer observations) large groups. We think that these questions deserve further investigation in future works.

Author Contributions: Conceptualization, G.D., L.S., L.L.R., R.S. and M.V.; methodology, G.D., L.L.R., R.S. and M.V.; software, G.D., L.S. and M.V.; validation, G.D., L.S. and M.V.; formal analysis, L.L.R., R.S. and M.V.; data curation G.D., L.S. and M.V.; writing—original draft preparation, G.D., L.S., L.L.R., R.S. and M.V.; writing—review and editing, G.D., L.S., L.L.R., R.S. and M.V.; visualization, G.D., L.S., L.L.R. and M.V.; supervision, L.L.R., R.S. and M.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Università degli Studi di Modena e Reggio Emilia (FAR2019 project of the Department of Physics, Informatics and Mathematics).

Data Availability Statement: The data analyzed in this study are publicly available and can be found here: http://morespace.unimore.it/marcovillani/software/, accessed on 26 March 2021.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Appendix A.1. The Sieving Method

The collection of sets returned by our proposed analysis is likely to contain groups included in others or partially overlapping groups, a fact that requires further analyses to assess their actual relevance. Having a high index value is not sufficient to characterize a set, because such ranking might result from the inclusion of a smaller set of variables characterized by an even higher zI, i.e., the set under consideration could be a superset of a more relevant one. In this case, the only relevant set would be the latter. On the other hand, a set having a high zI value might reach an even higher value, if some other relevant variables are added to it, i.e., the set under consideration could be a subset of a more relevant one. In this case, we would consider only the larger set as relevant.

In order to tackle this problem, we proposed a postprocessing sieving algorithm to reduce the overall number of subsets. The main assumption of the procedure is that if A is a proper subset of set B, that is, $A \subset B$, then only the higher value subset is taken into

consideration, see Figure A1 [22]. Therefore, only disjoint or partially overlapping subsets are kept by the sieving algorithm.

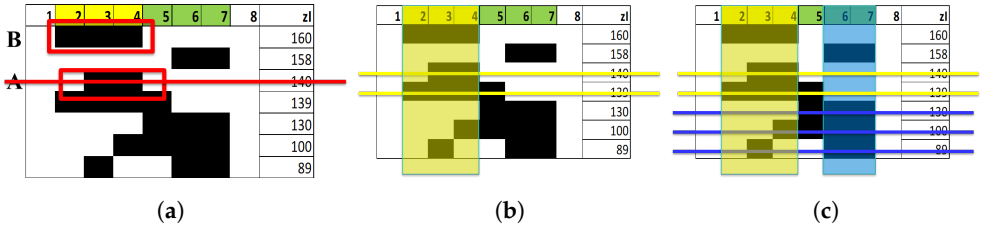


Figure A1. Sieving method. (**a**) If A is a subset (or a superset) of B, and B ranks higher than A, then we can think that A is an epiphenomenon, and B a more fundamental element than A (only the first part of the list of all possible groups is shown in the figure). (**b**) The sieving algorithm, applied to the strongest set, eliminates some groups (yellow lines). (**c**) The procedure is then repeated starting from the strongest remaining set and iterated until there are no more subsets or supersets to be eliminated. In the figure, the two remaining sets of the visible part of the list are those corresponding to index values 160 and 158

After the sieve has been applied, the remaining subsets cannot be decomposed any further, and thus they represent the building blocks of the dynamical organization of the system [24,25]. We refer to these building blocks as Candidate-Relevant Sets (CRS) in the following pseudocode of the sieving method (Algorithm A1).

Algorithm A1 The Sieving Method

Input: The array S of all the subsets, ranked by their index value in descending order
Output: CRS, a subset of S
CRS $\Leftarrow \emptyset$
$n \Leftarrow |S|$
Initialize auxliary array Aux[k] $\Leftarrow 0$ for k in 1...n
for $i = 1$ to $n - 1$ **do**
  for $j = i + 1$ to n **do**
    if Aux[i] $\neq 1$ **and** Aux[j] $\neq 1$ **then**
      if S[i] $\subset$ S[j] **or** S[j] $\subset$ S[i] **then**
        Aux[j] $\Leftarrow 1$
for $i = 1$ to n **do**
  if Aux[i] $= 0$ **then**
    CRS $\Leftarrow$ CRS $\cup$ {S[i]}

Appendix A.2. The Iterative Sieving Method

The method previously described allows one to identify a plausible organization of the system in terms of its lowest level, possibly overlapping, subsets of variables. Nevertheless, as complex systems have often a hierarchical structure, one may want to be able to make hypotheses on aggregated relations among the dynamic building blocks thus identified. To this aim, we devised an iterative version of the sieving method, which acts on the data by iteratively grouping the variables in one or more building blocks into a single entity. There are several ways to do so, but the simplest one, yet quite effective, consists in iteratively running the sieving algorithm on the same data, each time using a new representation in which the top-ranked building block of the previous iteration is considered as atomic and substituted by a single variable (henceforth called a group variable). In this way, each run produces a new atomic group of variables composed of both single variables and group variables introduced in previous iterations [22].

Let us consider a synthetic example—of which we therefore know the connection topology. The system includes eight variables, denoted by $1, \ldots, 8$, and suppose that the group $\{2, 3, 4\}$ is the most relevant set detected by the first iteration of the algorithm. The

second iteration will then analyze the dynamics of a system comprising the six variables 1, $\{2,3,4\}$, 5, 6, 7, and 8; the third iteration will analyze, for instance, the dynamics of a system comprising the five variables 1, $\{2,3,4\}$, 5, $\{6,7\}$, and 8; and so on until the index value of the most relevant set detected falls below a predetermined threshold, which we usually set equal to 3.0, see Figure A2.

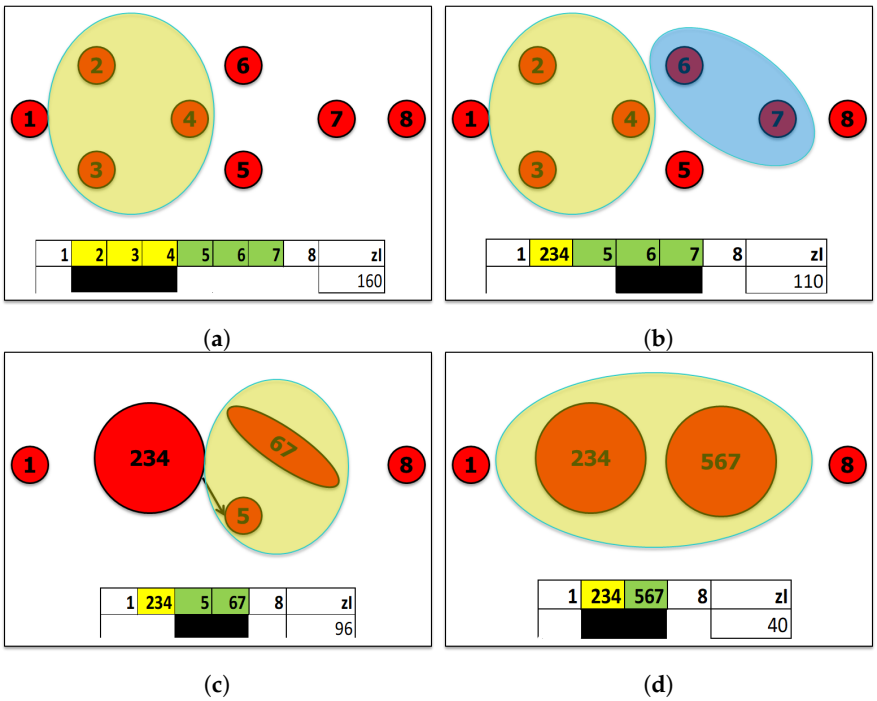
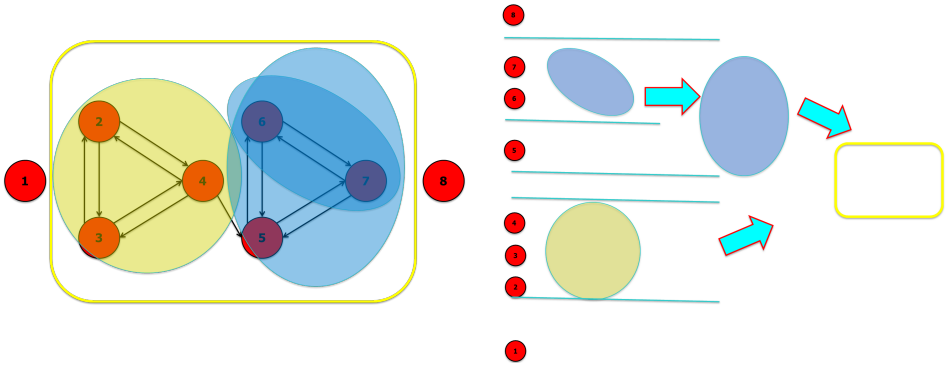


Figure A2. Iterative sieving method. (**a**) The first iteration of the sieving algorithm identifies the three variables 2, 3, and 4 as the top-ranked building block and compacts them into a single group variable. The new variable has a number of levels equal to or greater than the maximum number of levels of its components, but typically (much) less than their Cartesian product, because the variables are dynamically dependent. The new variable is inserted in the observations in place of the three from which it is composed. (**b**) The subsequent analysis identifies the group composed of the variables 6 and 7, which will then be compacted and replaced by the group variable $\{6,7\}$. (**c**) The variables $\{6,7\}$ and 5 are further merged into the group variable $\{5,6,7\}$. (**d**) The variables $\{2,3,4\}$ and $\{5,6,7\}$ are merged into the variable $\{2,3,4,5,6,7\}$. The algorithm will not accept further mergers, because the groups $\{1,2,3,4,5,6,7\}$, $\{2,3,4,5,6,7,8\}$, and $\{1,2,3,4,5,6,7,8\}$ have a zI lower than the chosen threshold (situation not shown). Note that only one group is compacted at a time, and so it is possible to avoid calculating the eliminations due to the simple sieving algorithm; see also the final paragraph of this Appendix A.

The succession of mergers performed by the iterated sieving algorithm allows to observe a hierarchy of nested groups: the final groups are the largest possible groupings, which constitute our Relevant Sets, see Figure A3.



(**a**) (**b**)

Figure A3. Iterative sieving method. (**a**) The nesting of the observed dynamic groups (superimposed on the system variables). The variables 1 and 8 are in fact random variables separated from the system at work (the largest group identified, highlighted here by the yellow box). Each variable performs the XOR of its input variables, except for node 5 that sets node 4 in AND with the XOR of nodes 6 and 7. (**b**) The hierarchy identified by the succession of groupings carried out by the *z*I analysis.

The above-described version of the iterative sieve is quite effective. As only one group is compacted at a time, there is no need to calculate the eliminations due to the simple sieve. Nevertheless, other variants may be implemented, which may produce more than one group variable per iteration: instead of considering just the top-ranked building block as a new group variable, one may possibly transform the first q sets into group variables, with q chosen according to some empirical criterion.

References

1. Lane, D.; Pumain, D.; van der Leeuw, S.E.; West, G. (Eds). *Complexity Perspectives in Innovation and Social Change*; Springer: Dordrecht, The Netherlands, 2009.
2. Haken, H. *Synergetics: Introduction and Advanced Topics*; Springer: Berlin, Germany, 2004.
3. Emmeche, C.; Køppe, S.; Stjernfelt, F. Explaining emergence: Towards an ontology of levels. *J. Gen. Philos. Sci.* **1997**, *28*, 83–117. [CrossRef]
4. Lane, D. Hierarchy, complexity, society. In *Hierarchy in Natural and Social Sciences*; Pumain, D., Ed.; Springer: Dordrecht, The Netherlands, 2006; pp. 81–119.
5. Barabási, A.L. *Network Science*; Cambridge University Press: Cambridge, UK, 2016.
6. Lewis, T.G. *Network Science: Theory and Application*; Wiley: Hoboken, NJ, USA, 2009.
7. Fortunato, S. Community detection in graphs. *Phys. Rep.* **2010**, *486*, 75–174. [CrossRef]
8. Scott, J.G. *Social Network Analysis: A Handbook*, 2nd ed.; SAGE: London, UK, 2000.
9. Tononi, G.; McIntosh, A.R.; Russel, D.P.; Edelman, G.M. Functional clustering: Identifying strongly interactive brain regions in neuroimaging data. *Neuroimage* **1998**, *7*, 133–149. [CrossRef]
10. Tononi, G.; Sporns, O.; Edelman, G.M. A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 5033–5037. [CrossRef] [PubMed]
11. Cover, T.M.; Thomas, A. *Elements of Information Theory*, 2nd ed.; Wiley: New York, NY, USA, 2006.
12. Watanabe, S. Information theoretical analysis of multivariate correlation. *IBM J. Res. Dev.* **1960**, *4*, 66–82. [CrossRef]
13. Nemenman, I.; Shafee, F.; Bialek, W. Entropy and inference, revisited. In Proceedings of the 14th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–8 December 2001; MIT Press: Cambridge, MA, USA, 2002; pp. 471–478.
14. Hausser, J.; Strimmer, K. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *J. Mach. Learn. Res.* **2009**, *10*, 1469–1484.
15. Archer, E.; Park, I.M.; Pillow, J.W. Bayesian and quasi-Bayesian estimators for mutual information from discrete data. *Entropy* **2013**, *15*, 1738–1755. [CrossRef]
16. Miller, G.A.; Madow, W.G. *On the Maximum Likelihood Estimate of the Shannon-Wiener Measure of Information*; Technical Report 54-75; Air Force Cambridge Research Center: Dayton, OH, USA, 1954.
17. Miller, G.A. Note on the bias of information estimates. In *Information Theory in Psychology*; Quastler, H., Ed.; Free Press: Glencoe, IL, USA, 1955; pp. 95–100.

18. Luce, R.D. The theory of selective information and some of its behavioral applications. In *Developments in Mathematical Psychology*; Luce, R.D., Ed.; Free Press: Glencoe, IL, USA, 1960; pp. 5–119.
19. Villani, M.; Roli, A.; Filisetti, A.; Fiorucci, M.; Poli, I.; Serra, R. The search for candidate relevant subsets of variables in complex systems. *Artif. Life* **2015**, *21*, 412–431. [CrossRef] [PubMed]
20. Roli, A.; Villani, M.; Caprari, R.; Serra, R. Identifying critical states through the relevance index. *Entropy* **2017**, *19*, 73. [CrossRef]
21. Roli, A.; Villani, M.; Filisetti, A.; Serra, R. Dynamical criticality: Overview and open questions. *J. Syst. Sci. Complex.* **2018**, *31*, 647–663. [CrossRef]
22. Villani, M.; Sani, L.; Pecori, R.; Amoretti, M.; Roli, A.; Mordonini, M.; Serra, R.; Cagnoni, S. An iterative information-theoretic approach to the detection of structures in complex systems. *Complexity* **2018**, *2018*, 3687839. [CrossRef]
23. Filisetti, A.; Villani, M.; Roli, A.; Fiorucci, M.; Poli, I.; Serra, R. On some properties of information theoretical measures for the study of complex systems. In *Advances in Artificial Life and Evolutionary Computation*, *Proceedings of the WIVACE 2014, Vietri sul Mare (SA), Italy, 14–15 May 2014*; Pizzuti, C.; Spezzano, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 140–150.
24. Filisetti, A.; Villani, M.; Roli, A.; Fiorucci, M.; Serra, R. Exploring the organisation of complex systems through the dynamical interactions among their relevant subsets. In Proceedings of the European Conference on Artificial Life 2015, York, UK, 20–24 July 2015; Andrews, P., Caves, L., Doursat, R., Hickinbotham, S., Polack, F., Stepney, S., Tim, T., Jon, T., Eds.; MIT Press: Cambridge, MA, USA, 2015; pp. 286–293.
25. Sani, L.; Amoretti, M.; Vicari, E.; Mordonini, M.; Pecori, R.; Roli, A.; Villani, M.; Cagnoni, S.; Serra, R. Efficient search of relevant structures in complex systems. In *AI*IA 2016 Advances in Artificial Intelligence*; Adorni, G., Cagnoni, S., Gori, M., Maratea, M., Eds.; Springer: Cham, Switzerland, 2016; pp. 35–48.
26. Silvestri, G.; Sani, L.; Amoretti, M.; Pecori, R.; Vicari, E.; Mordonini, M.; Cagnoni, S. Searching Relevant Variable Subsets in Complex Systems Using K-Means PSO. In *Artificial Life and Evolutionary Computation*, *Proceedings of the WIVACE 2017, Venice, Italy, 19–21 September 2017*; Pelillo, M., Poli, I., Roli, A., Serra, R., Slanzi, D., Villani, M., Eds.; Springer: Cham, Switzerland, 2018; pp. 308–321.
27. Vicari, E.; Amoretti, M.; Sani, L.; Mordonini, M.; Pecori, R.; Roli, A.; Villani, M.; Cagnoni, S.; Serra, R. GPU-based parallel search of relevant variable sets in complex systems. In *Advances in Artificial Life, Evolutionary Computation, and Systems Chemistry, Proceedings of the WIVACE 2016, Fisciano, Italy, 4–6 October 2016*; Rossi, F., Piotto, S., Concilio, S., Eds.; Springer: Cham, Switzerland, 2017; pp. 14–25.
28. Held, L.; Sabanés Bové, D. *Applied Statistical Inference*; Springer: Berlin, Germany, 2014.
29. Wilks, S.S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* **1938**, *9*, 60–62. [CrossRef]
30. Wilks, S.S. The likelihood test of independence in contingency tables. *Ann. Math. Stat.* **1935**, *6*, 190–196. [CrossRef]
31. Villani, M.; Campioli, D.; Damiani, C.; Roli, A.; Filisetti, A.; Serra, R. Dynamical regimes in non-ergodic random Boolean networks. *Nat. Comput.* **2017**, *16*, 353–363. [CrossRef]
32. Villani, M.; Sani, L.; Amoretti, M.; Vicari, E.; Pecori, R.; Mordonini, M.; Cagnoni, S.; Serra, R. A Relevance Index Method to Infer Global Properties of Biological Networks. In *Artificial Life and Evolutionary Computation*, *Proceedings of the WIVACE 2017, Venice, Italy, 19–21 September 2017*; Pelillo, M., Poli, I., Roli, A., Serra, R., Slanzi, D.; Villani, M., Eds.; Springer: Cham, Switzerland, 2018; pp. 129–141.
33. Sani, L.; D'Addese, G.; Graudenzi, A.; Villani, M. The detection of dynamical organization in cancer evolution models. In *Advances in Artificial Life and Evolutionary Computation*, *Proceedings of the WIVACE 2019, Rende, Italy, 18–20 September 2019*; Pizzuti, C., Spezzano, G., Eds.; Springer: Cham, Switzerland, 2020; pp. 49–61.
34. Righi, R.; Roli, A.; Russo, M.; Serra, R.; Villani, M. New paths for the application of DCI in social sciences: theoretical issues regarding an empirical analysis. In *Advances in Artificial Life, Evolutionary Computation, and Systems Chemistry, Proceedings of the WIVACE 2016, Fisciano, Italy, 4–6 October 2016*; Rossi, F., Piotto, S., Concilio, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 42–52.
35. Villani, M.; Filisetti, A.; Benedettini, S.; Roli, A.; Lane, D.; Serra, R. The detection of intermediate-level emergent structures and patterns. In *Advances in Artificial Life, Proceedings of the ECAL 2013, Sicily, Italy, 2–6 September 2013*; Liò, P., Miglino, O., Nicosia, G., Nolfi, S., Pavone, M., Eds.; MIT Press: Boston, MA, USA, 2013; pp. 372–378.
36. Dyson, F.J. *Origins of Life*; Cambridge Univesity Press: Cambridge, UK, 1985.
37. Eigen, M.; Schuster, P. A principle of natural self-organization. *Naturwissenschaften* **1977**, *64*, 541–565. [CrossRef] [PubMed]
38. Eigen, M.; Schuster, P. The hypercycle. *Naturwissenschaften* **1978**, *65*, 7–41. [CrossRef]
39. Filisetti, A.; Serra, R.; Carletti, T.; Villani, M.; Poli, I. Non-linear protocell models: Synchronization and chaos. *Eur. Phys. J. B* **2010**, *77*, 249–256. [CrossRef]
40. Jain, S.; Krishna, S. Autocatalytic sets and the growth of complexity in an evolutionary model. *Phys. Rev. Lett.* **1998**, *81*, 5684–5687. [CrossRef]
41. Kauffman, S.A. *The Origins of Order*; Oxford University Press: Oxford, UK, 1993.
42. Ruiz-Mirazo, K.; Briones, C.; de la Escosura, A. Prebiotic systems chemistry: New perspectives for the origins of life. *Chem. Rev.* **2014**, *114*, 285–366. [CrossRef]
43. Solé, R.V.; Munteanu, A.; Rodriguez-Caso, C.; Macía, J. Synthetic protocell biology: From reproduction to computation. *Philos. Trans. R. Soc. B Biol. Sci.* **2007**, *362*, 1727–1739. [CrossRef]

44. Hordijk, W.; Hein, J.; Steel, M. Autocatalytic sets and the origin of life. *Entropy* **2010**, *12*, 1733–1742. [CrossRef]
45. Hordijk, W.; Steel, M. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J. Theor. Biol.* **2004**, *227*, 451–461. [CrossRef]
46. Filisetti, A.; Villani, M.; Damiani, C.; Graudenzi, A.; Roli, A.; Hordijk, W.; Serra, R. On RAF sets and autocatalytic cycles in random reaction networks. In *Advances in Artificial Life and Evolutionary Computation, Proceedings of the WIVACE 2014, Vietri sul Mare, Italy, 14–15 May 2014*; Pizzuti, C., Spezzano, G., Eds.; Springer: Cham, Switzerland, 2014; pp. 113–126.
47. Hordijk, W.; Steel, M. A formal model of autocatalytic sets emerging in an RNA replicator system. *J. Syst. Chem.* **2013**, *4*, 3. [CrossRef]
48. Vasas, V.; Fernando, C.; Santos, M.; Kauffman, S.A.; Szathmáry, E. Evolution before genes. *Biol. Direct* **2012**, *7*, 1. [CrossRef] [PubMed]
49. Serra, R.; Villani, M. *Modelling Protocells: The Emergent Synchronization of Reproduction and Molecular Replication*; Springer: Dordrecht, The Netherlands, 2017.
50. Perry, R.; Green, D. *Perry's Chemical Engineer's Handbook*, 8th ed.; Mc-Graw Hill: New York, NY, USA, 2007.
51. Farmer, J.D.; Kauffman, S.A.; Packard, N.H. Autocatalytic replication of polymers. *Phys. D Nonlinear Phenom.* **1986**, *22*, 50–67. [CrossRef]
52. Filisetti, A.; Graudenzi, A.; Serra, R.; Villani, M.; De Lucrezia, D.; Füchslin, R.M.; Kauffman, S.A.; Packard, N.; Poli, I. A stochastic model of the emergence of autocatalytic cycles. *J. Syst. Chem.* **2011**, *2*, 2. [CrossRef]
53. Filisetti, A.; Graudenzi, A.; Serra, R.; Villani, M.; Füchslin, R.M.; Packard, N.; Kauffman, S.A.; Poli, I. A stochastic model of autocatalytic reaction networks. *Theory Biosci.* **2012**, *131*, 85–93. [CrossRef] [PubMed]
54. Kauffman, S.A. *At Home in the Universe*; Oxford University Press: Oxford, UK, 1995.
55. Serra, R.; Villani, M. Sustainable growth and synchronization in protocell models. *Life* **2019**, *9*, 68. [CrossRef]
56. Arkin, A.; Shen, P.; Ross, J. A test case of correlation metric construction of a reaction pathway from measurements. *Science* **1997**, *277*, 1275–1279. [CrossRef]
57. Amigó, J.M.; Balogh, S.G.; Hernández, S. A brief review of generalized entropies. *Entropy* **2018**, *20*, 813. [CrossRef]