

How should complexity scale with system size?

E. Olbrich^{1,a}, N. Bertschinger¹, N. Ay^{1,2}, and J. Jost^{1,2}

¹ Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany

² Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA

Received 31 August 2007 / Received in final form 17 January 2008

Published online 2 April 2008 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2008

Abstract. We study how statistical complexity depends on the system size and how the complexity of the whole system relates to the complexity of its subsystems. We study this size dependence for two well-known complexity measures, the excess entropy of Grassberger and the neural complexity introduced by Tononi, Sporns and Edelman. We compare these results to properties of complexity measures that one might wish to impose when seeking an axiomatic characterization. It turns out that those two measures do not satisfy all those requirements, but a renormalized version of the TSE-complexity behaves reasonably well.

PACS. 89.75.-k Complex systems – 89.70.+c Information theory and communication theory

1 Introduction

One would expect that the topic of measuring complexity is a central one in the field of complex systems. But this is not the case. Instead, the notion of complexity is very often used in an informal way with the hope that the readers will understand what the authors mean. Even though this is very often the case we believe that more care should be devoted to the use of the notion of complexity, at least in the field of complex systems itself and that a further development of complexity measures would help this a lot.

The quest for complexity measures has already produced sound results in its long history. On a general level one can distinguish between the complexity of describing and the complexity of interpreting [1]. The first refers to the problem of describing a system while the latter is related to the difficulty of the task to construct the system given the description. One can consider these two aspects as the encoding and the decoding aspect of complexity. In the following we will only consider descriptonal complexity. Moreover, we will not consider the complexity of a single object but the statistical complexity of an ensemble of objects, which allows us to exclude all features from the description, that are randomly distributed in the ensemble. In the case of a single object, the algorithmic complexity is the appropriate setting and the entropy itself is the appropriate statistical measure [2]. In case of the ensemble, we are not interested in every detail of the objects in our ensemble, but only in its structures, i.e. its non-random part. The corresponding measures are often called statistical or structural complexity [3,4].

Following the line of research started in [5], in this paper we study two measures of statistical complexity,

the excess entropy [6], originally introduced as effective measure complexity in a time series context [4], and the TSE-complexity [7]. We will show that both measures are closely related.

We study how both measures behave if the system size is increased and will discuss three special cases: (i) adding one independent element (ii) adding an independent system and (iii) adding an exact copy of the first system.

We will discuss how these results correspond to what we expect from a complexity measure and we propose a modification of the TSE-complexity based on these results. Furthermore, we hope that the present results will contribute to a better understanding and a specification of our information-geometric class of complexity measures that we introduced in [5].

2 Preliminaries from information theory

In this section, we recall some basic notions from information theory; a reference is [8]. We consider a set V of $1 \leq N < \infty$ nodes with state sets \mathcal{X}_v , $v \in V$. Given a finite subset $A \subseteq V$, we write \mathcal{X}_A instead of $\times_{v \in A} \mathcal{X}_v$, and the total configuration set is \mathcal{X}_V . $|\mathcal{X}_A|$ is the number of elements in \mathcal{X}_A , that is, the number of different states that can be attained on A . We have the natural projections

$$X_A : \mathcal{X}_V \rightarrow \mathcal{X}_A, \quad (x_v)_{v \in V} \mapsto (x_v)_{v \in A}.$$

Given a probability vector p on \mathcal{X}_V , the X_A become random variables. For three subsets A, B, C of V , we shall use the following information-theoretic quantities: The *entropy* of X_C is defined as

$$H(X_C) := - \sum_{z \in \mathcal{X}_C} \Pr(X_C = z) \log_2 (\Pr(X_C = z))$$

^a e-mail: olbrich@mis.mpg.de

This quantity is a natural measure of the uncertainty that one has about the outcome of measuring X_C , that is, the information one expects to gain by observing that outcome. Another interpretation of $H(X_C)$ is to consider it as the *variety* of X_C .

Knowing the state of B reduces the uncertainty that one has about the state of C . The remaining uncertainty is then quantified by the *conditional entropy* of X_C given X_B :

$$H(X_C|X_B) := - \sum_{\substack{y \in \mathcal{X}_B \\ z \in \mathcal{X}_C}} \Pr(X_B = y, X_C = z) \log_2 (\Pr(X_C = z|X_B = y)). \quad (1)$$

In terms of these entropy measures, the *mutual information* of X_C and X_B is given by

$$I(X_C : X_B) := H(X_C) - H(X_C|X_B) = H(X_C) + H(X_B) - H(X_C, X_B) \quad (2)$$

which measures the reduction of the uncertainty of the outcome of X_C given the outcome of X_B and vice versa.

The *conditional mutual information*, which is defined as

$$I(X_C : X_B|X_A) := H(X_C|X_A) - H(X_C|X_A, X_B), \quad (3)$$

quantifies the average reduction of the uncertainty of the outcome of measuring X_C knowing the state of B , if the state from A was already known.

The *multi-information* is a generalization of the mutual information for more than two random variables. The multi-information of the system X_V with respect to its nodes is defined as

$$I(X_V) := \sum_{v \in V} H(X_{\{v\}}) - H(X_V). \quad (4)$$

It is the difference between the sum of the variety of the elements and the variety of the system as a whole. It is often considered as a measure for the total statistical interdependence of the nodes with respect to the joint distribution p , but it remains unclear at this point what total statistical interdependence exactly means as long as one does not consider (4) itself as its definition.

In the literature it is also referred to as *integration* (see e.g. [7]). It becomes zero if and only if the probability distribution p has the product structure

$$p(x) = \prod_{v \in V} p_v(x_v), \quad (5)$$

where each p_v denotes the marginal distribution of the projection $X_{\{v\}}$. In particular, the multi-information vanishes in the case of complete randomness, given by the uniform distribution, and in the case of complete determinism, given by a distribution that is concentrated in one configuration.

3 Information-theoretic complexity measures

The information-theoretic quantities of Section 2 can be used to define complexity measures for random fields: we shall consider here two measures of structural complexity — the excess-entropy and the TSE complexity.

3.1 Excess entropy

One possibility to introduce a measure of statistical complexity is to ask to which extent the state of a subsystem remains uncertain if the state of the rest of the system is known. This uncertainty is given by the conditional entropy $H(X_A|X_{V \setminus A})$. It quantifies the amount to which the state cannot be explained by dependencies in the system and is therefore considered as random.

In particular, one can ask this question for any single element of our system. The excess entropy is then the difference between the uncertainty of the state of the whole system and the sum of the unreducible uncertainties of the state of the elements using all information available in the system

$$E(X_V) := H(X_V) - \sum_{v \in V} H(X_{\{v\}}|X_{V \setminus \{v\}}). \quad (6)$$

It quantifies the “explainable” part of the variety of the system. The excess entropy as the non-extensive part of the entropy was originally introduced as a complexity measure for time series [3,4,9] and provides in this context a lower bound for the amount of memory one needs for an optimal prediction. The excess entropy (6) is, however, not a direct generalization of the measures for time series, because in the latter case the conditioning in the second term is restricted to observables from the past only. Formula (6) can be rewritten using the average entropies of subsets of size k denoted by $H(k, N)$

$$H(k, N) = \binom{N}{k}^{-1} \sum_{\substack{Y \subseteq V \\ |Y|=k}} H(X_Y).$$

We get the following expression

$$E(X_V) = H(N, N) - N(H(N, N) - H(N - 1, N))$$

which is illustrated in Figure 1.

As a complexity measure for finite systems it was mentioned in passing in [10], see also [5] for a more comprehensive discussion of the relation between the time series case and the case of finite systems. The excess entropy (6) has the following properties (for all proofs see Appendix A):

1. The excess entropy is monotonically increasing with the system size because

$$E(X_{V \cup \{v'\}}) - E(X_V) = \sum_{i=1}^N I(X_{\{v_i\}} : X_{\{v'\}}|X_{V \setminus \{v_i\}}) \geq 0 \quad (7)$$

with v' denoting the additional element.

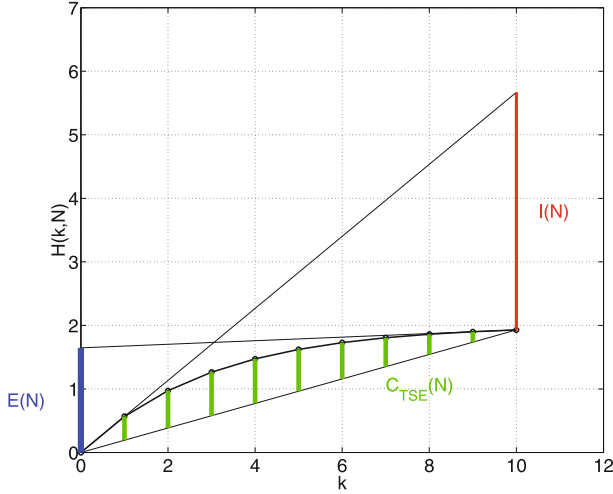


Fig. 1. (Color online) The figure shows schematically for a system with N elements the relationship between the mean entropy of subsystems of size k , $H(k, N)$, the integration $I(N)$ (4), the excess entropy $E(N)$ (6), and the TSE-complexity $C_{TSE}(N)$ (12).

2. The excess entropy of a system consisting of two subsystems A and B is always larger than the mutual information between these two subsystems:

$$E(X_{A \cup B}) \geq I(X_A : X_B).$$

3. The excess entropy of the union of two subsystems is always larger than the excess entropy of one of the subsystems.

$$E(X_{A \cup B}) \geq E(X_A) \quad E(X_{A \cup B}) \geq E(X_B)$$

4. In general the sum of the excess entropies of the subsystems is neither a lower nor an upper bound for the excess entropy of the whole system.

$$\begin{aligned} E(X_{A \cup B}) &= E(X_A) + E(X_B) \\ &+ \sum_{v \in A} I(X_{\{v\}} : X_B | X_{A \setminus \{v\}}) \\ &+ \sum_{v \in X_B} I(X_{\{v\}} : X_A | X_{B \setminus \{v\}}) \\ &- I(X_A : X_B). \end{aligned} \quad (8)$$

3.2 TSE complexity

Tononi et al. [7] introduced a complexity measure called “neural complexity”. It was motivated by the attempt to measure the potential ability of a neural system to produce consciousness in the framework of the information integration theory of consciousness and the dynamical core hypothesis¹ [11]. Starting from the intuition that conscious experience is very rich, they first required that a corresponding neural system should have a large number of

¹ In their more recent work (e.g. [12]) Tononi et al. used a different measure which also takes into account causal effects.

available states. On the other hand, consciousness is experienced as a unity. Tononi et al. translated this intuition into the requirement that the corresponding systems should have both high entropy and a high integration or multi-information (4) on the system level. In the following we will denote their complexity measure by C_{TSE} , where TSE stands for Tononi-Sporns-Edelman. They defined it as

$$C_{TSE}(X_V) := \sum_{k=1}^N \left(\frac{k}{N} I(X_V) - I(k, N) \right), \quad (9)$$

with the abbreviations $I(N) = I(X_V)$ for the multi-information of the whole system and

$$I(k, N) = \binom{N}{k}^{-1} \sum_{\substack{Y \subseteq V \\ |Y|=k}} I(X_Y)$$

for the average multi-information of subsystems of size k . This complexity is the higher the larger the increase of the integration with the size of the subsystems deviates from a linear one.

One can express C_{TSE} also using the mean entropies of the subsystems of size k or the mean mutual information for bipartitions into subsystems of size k and $N - k$. A particular interesting representation shows its relation to the excess entropy. If we denote the mean excess entropy of all subsystems with k elements by $E(k, N)$ we get for the TSE-complexity

$$C_{TSE}(X_V) = \frac{1}{2} \sum_{k=1}^N E(k, N). \quad (10)$$

For the proof see Appendix B. Using

$$E(k, N) = \frac{1}{\binom{N}{k}} \sum_{Y \subseteq X, |Y|=k} E(X_Y)$$

we can write the TSE-complexity also as a weighted sum over the excess entropy of all subsets

$$C_{TSE}(X_V) = \frac{1}{2} \sum_{Y \subseteq V} \frac{1}{\binom{N}{|Y|}} E(X_Y). \quad (11)$$

If we interpret the excess entropy as the complexity of a system with respect to its elements, the TSE-complexity measures the sum over the complexities on all levels with respect to its basic level. Maximizing the TSE complexity should therefore lead to systems which are “complex” on all levels.

How does the TSE-complexity grow with system size? To investigate this question it might be useful to start with the most basic representation using the entropies

$$C_{TSE}(X_V) = \sum_{k=1}^N \left(H(k, N) - \frac{k}{N} H(X_V) \right) \quad (12)$$

which is illustrated in Figure 1. After adding one element we not only have to replace N by $N+1$, but we also have to

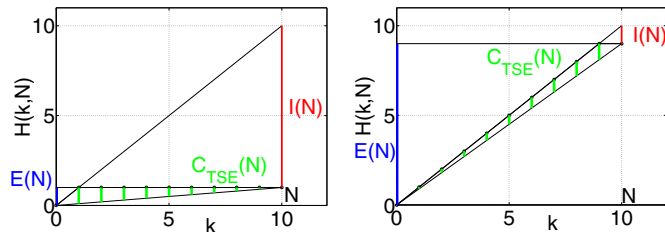


Fig. 2. Same as in Figure 1 for a system maximizing the integration (left) and the excess entropy (right).

take into account that the averages now include additional subsets. In fact, because $\binom{N}{k}$ is the number of subsets with k elements out of N elements, we have $\binom{N+1}{k} = \binom{N}{k} + \binom{N}{k-1}$, where the second term is the number of new subsystems containing v' . Thus with $H(k, N)$ denoting the mean entropy of subsystems of size k in the system with N elements

$$\binom{N+1}{k} H(k, N+1) = \binom{N}{k} H(k, N) + \binom{N}{k-1} H(k-1, N) + \sum_{\substack{Y \subseteq V \\ |Y|=k-1}} H(X_{\{v'\}} | X_Y). \quad (13)$$

Inserting this in equation (12) and rearranging the summation we get (see Appendix C):

$$C_{TSE}(X_{V \cup \{v'\}}) = \left(1 + \frac{1}{N+1}\right) C_{TSE}(X_V) + \sum_{k=1}^{N+1} \binom{N+1}{k}^{-1} \sum_{\substack{Y \subseteq V \\ |Y|=k-1}} I(X_{\{v'\}} : X_{V \setminus Y} | X_Y). \quad (14)$$

3.3 Maximizer for the the integration and the excess entropy

A representation similar to Figure 1 can also be used to understand the properties of the distributions that maximize the integration (4) and the excess entropy (6) for a system with a fixed number of binary elements ($x_v \in \{0, 1\}$). For a more detailed discussion of this topic see [5]. Figure 2 shows the behaviour of $H(k, N)$ as a function of k . If the integration is maximized, the entropy of the whole system is equal to the marginal entropy of a single element, which has to be maximal, i.e. 1 bit. The fact that the entropy does not increase further with increasing k implies that the state of the other elements is a function of the state of a single element. One can say that the system is synchronized.

On the other hand, the excess entropy is maximized by a distribution which looks like independent random variables up to subsets of size $N-1$, but the N th element is a function of the other $N-1$ elements. These conditions are fulfilled by the parity function $x_N = \left(\sum_{i=1}^{N-1} x_i\right) \bmod 2$.

A distribution that maximizes the TSE-complexity has to be somewhere between the two cases. One might expect

that for such maximizers $H(k, N)$ increases with the maximal possible slope of 1 up to a subsystem size of $N/2$ and then remains constant. Such a behaviour of $H(k, N)$ is, however, in general not achievable due to combinatorial constraints.

4 Special cases

Let us now consider three special cases of (i) adding an independent element, (ii) two independent subsystems, and (iii) two identical copies. What would we require for a reasonable complexity measure?

- (i) additional independent element: The element has no structure itself, so it has no own complexity. Because it is independent on the rest of the system the complexity should **not change**;
- (ii) union of two independent systems: Because there are no dependencies between the two systems the complexity of the union should be simply the **sum of the complexities** of the subsystems;
- (iii) union of two identical copies: Because there is no need for additional information to describe the second system one could argue that the complexity should be **equal to the complexity of one system**. One has, however, to include the fact in the description that the second system is a copy of the first one. At least this part should not be extensive with respect to the system size.

4.1 Adding an independent node

From (7) we get

$$E(X_{V \cup \{v'\}}) = E(X_V), \quad (15)$$

i.e. the excess entropy remains constant. Because the excess entropy of a single node is zero, adding an independent node leaves the excess entropy unchanged.

For the TSE-complexity we can use (Eq. (14)). The second term of the sum vanishes because of the independence.

$$C_{TSE}(X_{V \cup \{v'\}}) = \left(1 + \frac{1}{N+1}\right) C_{TSE}(X_V). \quad (16)$$

This shows that the TSE-complexity is increased by adding an independent element.

4.2 Two independent subsystems

For the excess entropy we have from (8) the general result that the excess entropy of a system composed of two independent subsystems is the sum of the excess entropies of the subsystems

$$E(X_{A \cup B}) = E(X_A) + E(X_B) \quad (17)$$

if $p_{A \cup B}(x_A, x_B) = p_A(x_A)p_B(x_B)$.

For the TSE-complexity this case is already a non-trivial one because the subsystems of the whole system can contain elements from both subsystems. But using the result (17), we have for an arbitrary subset $Y \subseteq A \cup B$

$$E(X_Y) = E(X_{Y \cap A}) + E(X_{Y \cap B}).$$

By using this property, we can show (see Appendix D) that

$$C_{TSE}(X_{A \cup B}) = \frac{N_A + N_B + 1}{N_A + 1} C_{TSE}(X_A) + \frac{N_A + N_B + 1}{N_B + 1} C_{TSE}(X_B) \quad (18)$$

with N_A denoting the number of elements in A and N_B the number of elements in B . Obviously, in contrast to the excess entropy the TSE-complexity of the unification of two independent compounds is not the sum of the TSE-complexities of the parts. But, we can re-establish this result for a renormalized version of the TSE-complexity

$$\tilde{C}_{TSE}(X_V) = \frac{1}{N + 1} C_{TSE}(X_V) \quad \text{with } N = |V|. \quad (19)$$

Note that this renormalization also restores the intuitive property that the complexity remains unchanged if a single independent node is added.

4.3 Two copies of the same system

Now we consider again two subsystems X_A and X_B but now X_B is an exact copy of X_A . Thus

$$p_{A \cup B}(x_A, x_B) = p_A(x_A) \delta_{x_A}(x_B).$$

It follows that $H(X_A, X_B) = H(X_A)$ and for any $v^A \in A$ and $v^B \in B$

$$H(X_{v^A} | X_{A \setminus \{v^A\}}, X_B) = H(X_{v^B} | X_{B \setminus \{v^B\}}, X_A) = 0.$$

Therefore we get for the excess entropy

$$E(X_{A \cup B}) = H(X_A) = H(X_B) = I(X_A : X_B). \quad (20)$$

The main problem with this result is that the ‘‘complexity’’ of two identical copies measured by the excess entropy is independent of the complexity of the single system which is clearly counterintuitive and shows a severe limitation of the excess entropy as a complexity measure for finite systems.

The situation for the TSE-complexity is more complicated. Using a reasoning similar to the case of the two independent subsystems (see Appendix E) one finally arrives at

$$C_{TSE}(X_{A \cup B}) = \frac{2N_A + 1}{N_A + 1} C_{TSE}(X_A) + \sum_{Y_A \subseteq A} \sum_{Y_B \subseteq B} \binom{N}{|Y_A| + |Y_B|}^{-1} H(X_{Y_B} | X_{Y_A}). \quad (21)$$

which leads to a very reasonable lower bound for the renormalized TSE-complexity

$$\tilde{C}_{TSE}(X_{A \cup B}) \geq \tilde{C}_{TSE}(X_A). \quad (22)$$

The single summands in the second term of (21) cannot be simplified further, one can, however, try to derive an upper bound by using

$$H(X_{Y_B} | X_{Y_A}) \leq H(X_{Y_B \setminus (Y_A \cap Y_B)}).$$

According to our requirements for a complexity measure the second term in (21) should grow at least more slowly than the system size. At the moment, however, it is an open problem under which conditions this applies.

5 Discussion

We studied how two established measures of statistical complexity behave if the system size is increased and compared the results with intuitive requirements for complexity measures in three special cases: (i) adding one element, (ii) adding an independent system and (iii) adding a perfect copy of the original system.

While the excess entropy behaves according to what one should expect in the cases (i) and (ii), the result in case (iii) did not meet our intuitions: the excess entropy of the two copies does not depend on the complexity of the original system at all. This indicates a possible limitation of the excess entropy as a general complexity measure, because it only relates two levels to each other: the level of the elements and the system level, but ignores the intermediate levels.

This disadvantage might be avoided in the TSE-complexity because it considers all possible subsets of a given system. We could show this by writing the TSE-complexity as a sum over all subsystem sizes of the mean excess entropies of subsystems of this size. The further analysis of the behavior of the TSE-complexity revealed however, that it grows even if one adds independent elements which again is a counterintuitive and therefore unwanted property. The detailed analysis of the case of two independent systems showed that this unwanted property can be avoided by dividing the TSE-complexity by $N + 1$ if the system has N elements. Therefore we argue that the normalized version (19) provides a better measure of statistical complexity than the original TSE-complexity. It is however, not clear, to which extent the normalized TSE-complexity avoids the problem of the excess entropy in the case of two identical copies, because we found no simple interpretation of the second term in (21).

We hope that the insights obtained in this paper will also help to advance our geometric approach to complexity presented in [5]. This, however, requires more sophisticated mathematical considerations and will therefore be presented elsewhere.

Funded by Volkswagen Foundation.

Appendix A: Excess entropy of two subsystems

The excess entropy of a random variable $X_V = X_{A \cup B}$ is defined as

$$E(X_{A \cup B}) = H(X_{A \cup B}) - \sum_{v \in X_{A \cup B}} H(X_{\{v\}} | X_{(A \cup B) \setminus \{v\}}). \quad (23)$$

Now let us label the elements of A in an arbitrary order as v_j^A , $j = 1, \dots, N_A$ and the elements in B as v_k^B , $k = 1, \dots, N_B$. Moreover, be A_k the set containing v_j^A with $1 \leq j \leq k$ and B_k the set containing v_j^B with $1 \leq j \leq k$. Using

$$H(X_{A \cup B}) = I(X_A : X_B) + H(X_A | X_B) + H(X_B | X_A) \quad (24)$$

with

$$H(X_A | X_B) = \sum_{j=1}^{N_A} H(X_{\{v_j^A\}} | X_{A_{j-1}}, X_B) \quad (25)$$

and

$$H(X_B | X_A) = \sum_{j=1}^{N_B} H(X_{\{v_j^B\}} | X_{B_{j-1}}, X_A) \quad (26)$$

we can rewrite the excess entropy as

$$\begin{aligned} E(X_{A \cup B}) &= I(X_A : X_B) \\ &+ \sum_{j=1}^{N_A} I(X_{\{v_j^A\}} : X_{A \setminus A_j} | X_{A_{j-1}}, X_B) \\ &+ \sum_{j=1}^{N_B} I(X_{\{v_j^B\}} : X_{B \setminus B_j} | X_{B_{j-1}}, X_A). \end{aligned} \quad (27)$$

The conditional mutual information is always larger than zero and therefore

$$E(X_{A \cup B}) \geq I(X_A : X_B). \quad (28)$$

To compare the excess entropy of the whole system with the excess entropy of one of its parts we have using the same notation

$$\begin{aligned} E(X_{A \cup B}) &= H(X_A) + H(X_B | X_A) \\ &- \sum_{v \in A} H(X_{\{v\}} | X_{A \setminus \{v\}}) - \sum_{v \in B} H(X_{\{v\}} | X_{A \cup B \setminus \{v\}}) \\ &+ \sum_{v \in A} I(X_{\{v\}} : X_B | X_{A \setminus \{v\}}). \end{aligned}$$

By writing $H(X_B | X_A)$ as a sum according to (26) we can express the right hand side as a sum of $E(X_A)$ and a sum of conditional mutual informations:

$$\begin{aligned} E(X_{A \cup B}) &= E(X_A) + \sum_{v \in A} I(X_{\{v\}} : X_B | X_{A \setminus \{v\}}) \\ &+ \sum_{j=1}^{N_B} I(X_{\{v_j^B\}} : X_{B \setminus B_j} | X_{B_{j-1}}, X_A) \end{aligned} \quad (29)$$

Thus $E(X_{A \cup B}) \geq E(X_A)$ and similarly $E(X_{A \cup B}) \geq E(X_B)$. By combining equations (27) and (29) we can also relate the excess entropy of the system to the sum of the excess entropy of its subsystems

$$\begin{aligned} E(X_{A \cup B}) &= E(X_A) + E(X_B) + \sum_{v \in A} I(X_{\{v\}} : X_B | X_{A \setminus \{v\}}) \\ &+ \sum_{v \in B} I(X_{\{v\}} : X_B | X_{A \setminus \{v\}}) - I(X_A : X_B). \end{aligned}$$

Appendix B: TSE-complexity and excess entropy

It was already noted in [10] that the last summand in (12) is equal to the excess entropy. But it is also possible to express the whole sum as a sum over excess entropies:

$$C_{TSE}(X_V) = \frac{1}{2} \sum_{k=1}^N E(k, N).$$

$E(k, N)$ and $H(k, N)$ denote the average excess entropy or entropy, respectively, of subsystems with k elements of a system with N elements. Moreover we use the abbreviations $E(N) = E(N, N)$ and $H(N) = H(N, N)$. First we rewrite the excess entropy as

$$\begin{aligned} E(N) &= H(N) - N(H(N) - H(N-1, N)) \\ &= NH(N-1, N) - (N-1)H(N). \end{aligned}$$

which leads to

$$H(N) = \frac{N}{N-1} H(N-1, N) - \frac{1}{N-1} E(N).$$

Now we can apply this formula again for all subsystems of size $N-1$

$$\begin{aligned} H(N) &= \frac{N}{N-2} H(N-2, N) \\ &- \frac{N}{(N-1)(N-2)} E(N-1, N) - \frac{1}{N-1} E(N). \end{aligned}$$

Here we used that for $k \leq N-k'$

$$\begin{aligned} H(k, N) &= \binom{N}{k}^{-1} \sum_{\substack{A \subseteq V \\ |A|=k}} H(X_A) \\ &= \binom{N}{N-k'}^{-1} \sum_{\substack{V' \subseteq V \\ |V'|=N-k'}} \binom{N-k'}{k}^{-1} \sum_{\substack{A \subseteq V' \\ |A|=k}} H(X_A) \end{aligned}$$

meaning that averaging the entropy of subsystems of size k over the whole system gives the same result than averaging first for a subsystem of size $N-k' \geq k$ and then averaging over all these subsystems. Repeating the recursion we get

$$H(N) = \frac{N}{N-k} H(N-k, N) + \sum_{j=1}^k \frac{N \cdot E(N-j+1, N)}{(N-j+1)(N-j)}$$

or with $k' = N - k$

$$H(k', N) - \frac{k'}{N} = \sum_{j=1}^{N-k'} \frac{k' \cdot E(N-j+1, N)}{(N-j+1)(N-j)}.$$

and finally

$$C_{TSE}(X_V) = \sum_{k'=1}^{N-1} \sum_{j=1}^{N-k'} \frac{k' \cdot E(N-j+1, N)}{(N-j+1)(N-j)}.$$

Rearranging the sums and summing up leads to equation (10)

$$C_{TSE}(X_V) = \frac{1}{2} \sum_{k=1}^N E(k, N).$$

Appendix C: TSE-complexity with an additional element

Starting point is again the TSE-complexity expressed by the average subset entropies (12)

$$C_{TSE}(X_V) = \sum_{k=1}^N \left(H(k, N) - \frac{k}{N} H(X_V) \right).$$

After increasing the system by one additional element we have

$$C_{TSE}(X_{V \cup \{v'\}}) = \sum_{k=1}^{N+1} \left(H(k, N+1) - \frac{k}{N+1} H(X_{V \cup \{v'\}}) \right).$$

Now we use that

$$\binom{N+1}{k} H(k, N+1) = \binom{N}{k} H(k, N) + \binom{N}{k-1} H(k-1, N) + \sum_{\substack{A \subseteq V \\ |A|=k-1}} H(X_{\{v'\}} | X_A)$$

and therefore

$$H(k, N+1) = \left(1 - \frac{k}{N+1}\right) H(k, N) + \frac{k}{N+1} H(k-1, N) + \binom{N+1}{k}^{-1} \sum_{\substack{A \subseteq V \\ |A|=k-1}} H(X_{\{v'\}} | X_A). \quad (30)$$

Now we can evaluate the sum over k

$$\begin{aligned} \sum_{k=1}^{N+1} H(k, N+1) &= \sum_{k=1}^N H(k, N) \\ &+ \sum_{k=1}^N \left(\frac{k+1}{N+1} - \frac{k}{N+1} \right) H(k, N) \\ &+ \sum_{k=1}^{N+1} \binom{N+1}{k}^{-1} \sum_{\substack{A \subseteq V \\ |A|=k-1}} H(X_{\{v'\}} | X_A) \\ \sum_{k=1}^{N+1} H(k, N+1) &= \left(1 + \frac{1}{N+1}\right) \sum_{k=1}^N H(k, N) \\ &+ \sum_{k=1}^{N+1} \binom{N+1}{k}^{-1} \sum_{\substack{A \subseteq V \\ |A|=k-1}} H(X_{\{v'\}} | X_A). \end{aligned}$$

Moreover

$$\begin{aligned} \sum_{k=1}^{N+1} \frac{k}{N+1} H(X_{V \cup \{v'\}}) &= \frac{N+2}{2} (H(X_V) \\ &+ H(X_{\{v'\}} | X_V)) = \left(1 + \frac{1}{N+1}\right) \frac{N+1}{2} H(X_V) \\ &+ \sum_{k=1}^{N+1} \frac{k}{N+1} H(X_{\{v'\}} | X_V). \end{aligned}$$

Now we can introduce for any subset Y of size $k-1$ the conditional mutual information

$$I(X_{\{v'\}} : X_{V \setminus Y} | X_Y) = H(X_{\{v'\}} | X_Y) - H(X_{\{v'\}} | X_V)$$

and get the result (14)

$$\begin{aligned} C_{TSE}(X_{V \cup \{v'\}}) &= \left(1 + \frac{1}{N+1}\right) C_{TSE}(X_V) \\ &+ \sum_{k=1}^{N+1} \binom{N+1}{k}^{-1} \sum_{\substack{Y \subseteq V \\ |Y|=k-1}} I(X_{\{v'\}} : X_{V \setminus Y} | X_Y) \end{aligned}$$

for the behavior of TSE-complexity after adding one element. Here we used that there are $\binom{N}{k-1}$ subsets from the original system containing $k-1$ elements and that $\binom{N}{k-1} \frac{N+1}{k} = \binom{N+1}{k}$.

Appendix D: TSE-complexity of the sum of two independent systems

How does the complexity measure of Tononi, Sporns and Edelman decompose if a system V is composed of two independent subsystems A and B , i.e. $V = A \cup B$ and $I(X_A : X_B) = 0$? In the following the number of elements in V will be denoted by N . For the subsystems N_A and N_B will be used respectively.

Starting from

$$C_{TSE}(X_V) = \frac{1}{2} \sum_{k=1}^N E(k, X)$$

where $E(k, X) = \frac{1}{\binom{N}{k}} \sum_{Y \subseteq X, |Y|=k} E(X_Y)$ is the average excess entropy over all subsets of cardinality k , one obtains

$$C_{TSE}(X_V) = \frac{1}{2} \sum_{Y \subseteq X} \frac{1}{\binom{N}{|Y|}} E(X_Y)$$

. Using the fact that the excess entropy is additive for independent subsystems, the TSE-complexity can be decomposed as follows:

$$\begin{aligned} C_{TSE}(X_V) &= \frac{1}{2} \sum_{Y \subseteq A \cup B} \frac{1}{\binom{N}{|Y \cap A| + |Y \cap B|}} (E(X_{Y \cap A}) + E(X_{Y \cap B})) \\ &= \frac{1}{2} \sum_{Y_A \subseteq A, Y_B \subseteq B} \frac{1}{\binom{N}{|Y_A| + |Y_B|}} (E(X_{Y_A}) + E(X_{Y_B})) \\ &= \frac{1}{2} \sum_{Y_A \subseteq A} \sum_{Y_B \subseteq B} \frac{1}{\binom{N}{|Y_A| + |Y_B|}} (E(X_{Y_A}) + E(X_{Y_B})) \\ &= \frac{1}{2} \sum_{Y_A \subseteq A} E(X_{Y_A}) \sum_{Y_B \subseteq B} \frac{1}{\binom{N}{|Y_A| + |Y_B|}} \\ &\quad + \frac{1}{2} \sum_{Y_B \subseteq B} E(X_{Y_B}) \sum_{Y_A \subseteq A} \frac{1}{\binom{N}{|Y_A| + |Y_B|}}. \end{aligned}$$

Comparing this to $C_{TSE}(X_A) = \frac{1}{2} \sum_{Y_A \subseteq A} \frac{1}{\binom{N_A}{|Y_A|}} E(X_{Y_A})$ shows that the two expressions differ in their weights in the sum by a factor² of

$$\sum_{Y_B \subseteq B} \frac{\binom{N_A}{k}}{\binom{N}{k + |Y_B|}} = \sum_{l=0}^{N_B} \frac{\binom{N_B}{l} \binom{N_A}{k}}{\binom{N_A + N_B}{k + l}}. \quad (31)$$

If this factor is independent of k the TSE-complexity of X can be written in terms of the TSE-complexities of A and B . We can show (see Appendix F) that the sum is equal to $\frac{N_A + N_B + 1}{N_A + 1}$. Then the TSE-complexity can be decomposed as follows:

$$\begin{aligned} C_{TSE}(X_{A \cup B}) &= \frac{1}{2} \sum_{Y_A \subseteq A} \frac{1}{\binom{N_A}{|Y_A|}} E(X_{Y_A}) \frac{N_A + N_B + 1}{N_A + 1} \\ &\quad + \frac{1}{2} \sum_{Y_B \subseteq B} \frac{1}{\binom{N_B}{|Y_B|}} E(X_{Y_B}) \frac{N_A + N_B + 1}{N_B + 1} \\ &= \frac{N_A + N_B + 1}{N_A + 1} C_{TSE}(X_A) + \frac{N_A + N_B + 1}{N_B + 1} C_{TSE}(X_B). \end{aligned}$$

² Since this factor just depends on the cardinality of the specific subset, k will be used in the following to denote the cardinality of Y_A .

Appendix E: TSE-complexity of two identical copies

Now consider the situation that the system consists of two identical copies of one smaller system, i.e. $V = A \cup B$ with $p(x_A, x_B) = p(x_A) \delta_{x_A}(x_B)$. In this case each subset Y of V can be considered to be composed of two subsets from $Y_A \subseteq A$ and $Y_B \subseteq B$ respectively. Furthermore each subset Y_B of B corresponds to a subset Y_B^A of A . Since V is a disjoint union of A and B , one has $|Y| = |Y_A| + |Y_B|$. Due to B being a copy of A the entropy of Y is given by

$$H(X_Y) = H(X_{Y_A \cup Y_B^A}) = H(X_{Y_A}) + H(X_{Y_B^A} | X_{Y_A}).$$

The TSE-complexity of X can therefore be decomposed as follows³:

$$\begin{aligned} C_{TSE}(X_V) &= \sum_{k=1}^N \left(\sum_{Y \subseteq X, |Y|=k} \frac{1}{\binom{N}{k}} H(X_Y) - \frac{k}{N} H(X_V) \right) \\ &= \sum_{Y \subseteq X} \frac{1}{\binom{N}{|Y|}} H(X_Y) - \frac{N+1}{2} H(X_V) \\ &= \sum_{Y_A \subseteq A} \sum_{Y_B \subseteq A} \frac{1}{\binom{N}{|Y_A| + |Y_B|}} H(X_{Y_A \cup Y_B}) - \frac{N+1}{2} H(X_V) \\ &= \sum_{Y_A \subseteq A} \sum_{Y_B \subseteq A} \frac{1}{\binom{N}{|Y_A| + |Y_B|}} H(Y_A) - \frac{N+1}{2} H(X_A) \\ &\quad + \sum_{Y_A \subseteq A} \sum_{Y_B \subseteq A} \frac{1}{\binom{N}{|Y_A| + |Y_B|}} H(X_{Y_B} | X_{Y_A}) \\ &= \sum_{Y_A \subseteq A} \frac{1}{\binom{N_A}{|Y_A|}} H(X_{Y_A}) \sum_{Y_B \subseteq A} \frac{\binom{N_A}{|Y_A|}}{\binom{N}{|Y_A| + |Y_B|}} \\ &\quad - \frac{2N_A + 1}{2} H(X_A) \\ &\quad + \sum_{Y_A \subseteq A} \sum_{Y_B \subseteq A} \frac{1}{\binom{N}{|Y_A| + |Y_B|}} H(X_{Y_B} | X_{Y_A}) \\ &= \sum_{Y_A \subseteq A} \frac{1}{\binom{N_A}{|Y_A|}} H(X_{Y_A}) \sum_{l=0}^{N_A} \frac{\binom{N_A}{|Y_A|} \binom{N_A}{l}}{\binom{N}{|Y_A| + l}} \\ &\quad - \frac{2N_A + 1}{2} H(X_A) \\ &\quad + \sum_{Y_A \subseteq A} \sum_{Y_B \subseteq A} \frac{1}{\binom{N}{|Y_A| + |Y_B|}} H(X_{Y_B} | X_{Y_A}) \\ &= \frac{2N_A + 1}{N_A + 1} C_{TSE}(X_A) + \sum_{Y_A \subseteq A} \sum_{Y_B \subseteq A} \frac{1}{\binom{N}{|Y_A| + |Y_B|}} \\ &\quad \times H(X_{Y_B} | X_{Y_A}). \end{aligned}$$

³ The cardinalities of V and A are denoted by N and N_A respectively. Due to V being composed of two copies of A one also has that $N = 2N_A$.

$$\begin{aligned}
 \sum_{l=0}^{N+1} \frac{\binom{N+1}{l}}{\binom{N_A+N+1}{k+l}} &= \sum_{l=0}^N \frac{\binom{N}{l}}{\binom{N_A+N+1}{k+l}} + \sum_{l=0}^N \frac{\binom{N}{l-1}}{\binom{N_A+N+1}{k+l}} + \binom{N_A+N+1}{N_A-k}^{-1} \\
 &= \sum_{l=0}^N \frac{\binom{N}{l}}{\binom{N_A+N+1}{k+l}} + \sum_{l'=-1}^{N-1} \frac{\binom{N}{l'}}{\binom{N_A+N+1}{k-1+l'+1}} + \binom{N_A+N+1}{N_A-k}^{-1} = S(N, N_A+1, k) + S(N, N_A+1, k-1) \\
 &\quad - \binom{N_A+N+1}{N_A-k}^{-1} + \binom{N_A+N+1}{N_A-k}^{-1} = \binom{N_A+1}{k}^{-1} \frac{N_A+1+N+1}{N_A+2} \\
 &\quad + \binom{N_A+1}{k-1}^{-1} \frac{N_A+1+N+1}{N_A+2} = \left(\frac{1}{\binom{N_A+1}{k}} + \frac{1}{\binom{N_A+1}{k-1}} \right) \frac{N_A+(N+1)+1}{N_A+2} \\
 &= \frac{1}{\binom{N_A}{k}} \frac{N_A+2}{N_A+1} \frac{N_A+(N+1)+1}{N_A+2} = \binom{N_A}{k}^{-1} \frac{N_A+(N+1)+1}{N_A+1}. \quad \square
 \end{aligned}$$

Appendix F: Evaluation of the sum (31)

In the following we want to show by induction that

$$\sum_{l=0}^{N_B} \frac{\binom{N_B}{l}}{\binom{N_A+N_B}{k+l}} = S(N_B, N_A, k) = \binom{N_A}{k}^{-1} \frac{N_A+N_B+1}{N_A+1} \tag{32}$$

Proof. First we note that for $N_B = 0$ the equation holds for all N_A and k .

Second we assume that it holds for $N_B = N$ and arbitrary N_A and k , i.e. we have

$$S(N, N_A, k) = \binom{N_A}{k}^{-1} \frac{N_A+N+1}{N_A+1}. \tag{33}$$

What happens for $N_B = N + 1$? Using that $\binom{N+1}{k} = \binom{N}{k} + \binom{N}{k-1}$ we get

See equation above.

References

1. L. Löfgren, *Int. J. Gen. Syst.* **3**, 197 (1977)
2. M. Li, P.M.B. Vitányi, *An Introduction to Kolmogorov Complexity and its Applications*, 2nd edn. (Springer, 1997)
3. J.P. Crutchfield, K. Young, *Phys. Rev. Lett.* **63**, 105 (1989)
4. P. Grassberger, *Int. J. Theor. Phys.* **25**, 907 (1986)
5. N. Ay, E. Olbrich, N. Bertschinger, J. Jost, *A unifying framework for complexity measures of finite systems, Working Paper 06-08-028*, Santa Fe Institute, Proceedings of ECCS'06 (2006)
6. J.P. Crutchfield, D.P. Feldman, *Chaos* **13**, 25 (2003)
7. G. Tononi, O. Sporns, G.M. Edelman, *PNAS* **91**, 5033 (1994)
8. T.M. Cover, J.A. Thomas, *Elements of Information Theory, Wiley series in telecommunications* (Wiley, New York, 1991)
9. W. Bialek, I. Nemenman, N. Tishby, *Neural Computation* **13**, 2409 (2001)
10. G. Tononi, O. Sporns, G.M. Edelman, *PNAS* **96**, 3257 (1999)
11. G. Tononi, G.M. Edelman, *Science* **282**, 1846 (1998)
12. G. Tononi, *BMC Neurosci* **5**, 42 (2004)