

Measuring Self-Organization via Observers

Daniel Polani

Adaptive Systems Research Group
Department of Computer Science
University of Hertfordshire
d.polani@herts.ac.uk

Abstract. We introduce *organization information*, an information-theoretic characterization for the phenomenon of self-organization. This notion, which requires the specification of an *observer*, is discussed in the paradigmatic context of the Self-Organizing Map and its behaviour is compared to that of other information-theoretic measures. We show that it is sensitive to the presence and absence of “self-organization” (in the intuitive sense) in cases where conventional measures fail.

1 Introduction

Shalizi [21] tracks back the first use of the notion of “self-organizing systems” to Ashby [2]. The bottom-up cybernetic approach of early artificial intelligence [25, 14] devoted considerable interest and attention to the area of self-organizing systems; many of the questions and methods considered relevant today have been appropriately identified almost half a century ago [e.g. 26].

The notion of *self-organization*, together with the related notion of *emergence*, are of central importance in the sciences of complexity and Artificial Life¹. These phenomena form the backbone of those types of dynamics that lead to climbing the ladder of complexity which, as it is believed, lies ultimately behind the appearance of life-like phenomena studied in Artificial Life. Notwithstanding the importance and frequent use of these notions in the relevant literature, a both precise and useful mathematical definition remains elusive. While there is a high degree of intuitive consensus on what type of phenomena should be called “self-organizing” or “emergent”, the prevailing strategy of characterization is along the line of “I know it when I see it”. For a given system, the presence of self-organization or emergence is typically determined by explicit inspection. The present paper will concentrate on the discussion of how to characterize self-organization.

Specialized formal literature often does not go beyond pragmatic characterizations; e.g. Jetschke [10] defines a system as undergoing a self-organizing transition if the symmetry group of its dynamics changes to a less symmetrical one (e.g. a subgroup of the original symmetry group), typically occurring at

¹ To avoid possible misunderstandings, though they often co-occur and are mentioned together, the present paper construes “self-organization” and “emergence” to be two distinctly different phenomena, but intricately intertwined, see also Sec. 4.

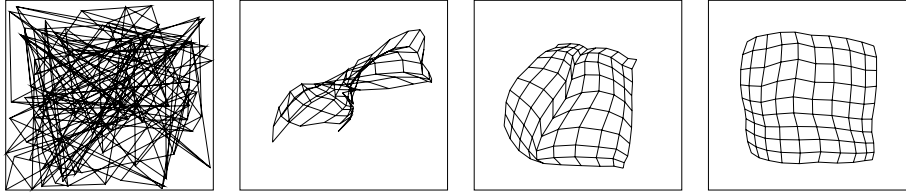


Fig. 1. An example for the training process for a Self-Organizing Map. The probability distribution used for generating the training samples is the equidistribution on the unit square $[0, 1]^2 \subseteq \mathbb{R}^2$. The sequence of plots shows the SOM weights X_i where two weights X_i and X_j are connected by a line if they belong to units i and j neighbouring each other in the grid at training steps 0, 10, 100, 1000.

phase transitions [19]. This view relates self-organization to phase transitions. However, there are several reasons to approach the definition of self-organization in a different way. The typical complex, living or artificial life system is not in thermodynamic equilibrium [see also 17]. One possible extension of the formalism is towards nonequilibrium thermodynamics, identifying phase transitions by *order parameters*. These are quantities that characterize the “deviation” of the system in a more “organized” state in the sense of Jetschke from the system in a less organized state, measured by absence or presence of symmetries. Order parameters have to be constructed by explicit inspection of the system since a generic approach is not available. Also, in Alife systems, one can not expect the a priori existence of any symmetry to act as indicators for self-organization.

2 Self-Organization

To discuss approaches to quantify self-organization, it is useful to introduce a paradigmatic system. We choose a system which is not among the typical Alife systems, but exemplifies the principles we need to develop. These principles are, however, not restricted to that model and generalize immediately to any stochastic dynamic system and will be applied to Alife systems in future. Here, our paradigm system is Kohonen’s Self-Organizing Map (SOM) [11, 20, 12].

For lack of space, we do not go into details about the SOM dynamics and give only give a brief outline. A SOM is a set of units i each of which carries a weight $X_i \in \mathbb{R}^n$. The units i are located on the nodes of a (typically square) grid. The SOM is now being *trained* with samples V drawn from a probability distribution on \mathbb{R}^n , during which the weights X_i reorganize in such a fashion that, if possible, neighbouring features in \mathbb{R}^n are represented by weights w_i and w_j belonging to units i and j which are neighbours in the grid and vice versa.

Figure 1 shows a typical training process for a SOM. The “self-organization” property is reflected by the fact that during the training the SOM representation changes from an “unorganized” to the “organized” final state. This “self-organization” property is immediately evident to a human observer, but far from

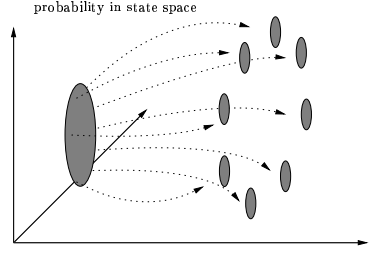


Fig. 2. Schematic representations of the evolution of probability distribution during the SOM training. The set of random initial configurations covers a larger proportion of the state space. During training the dynamics separates into distinct paths each concentrating along one of the eight probable organized solutions.

obvious to quantify [5, 24, 6, 15]. Many SOM organization measures are system-specific and therefore inadequate as universal notions of self-organization.

To alleviate this problem, Spitzner and Polani [22] attempted to apply Haken’s synergetics formalism [7] to the SOM. This formalism only requires a dynamical system structure. They converted the stochastic SOM training rule to its Kushner-Clark deterministic continuous counterpart [13] and applied the synergetics formalism to it. Unfortunately, the synergetics formalism has to be applied close to fixed points of the system under which conditions it turns out to be insensitive to SOM self-organization. However, the view as a dynamical system proves fertile and opens up a selection of possible alternative characterization approaches. Figure 2 shows schematically the evolution of the state probability distribution during SOM training for the example from Fig. 1. The initial state is a random configuration of an entire vector (X_1, X_2, \dots, X_k) of initial random weights $X_i \in [0, 1]^2$, $k = 1, 2, \dots, k$ covering large parts of state space. During training, the SOM will stabilize itself along one of the 8 organized “square” solutions (rightmost plot in Fig. 1), concentrating the state probability distribution around one of the 8 organized solutions. This phenomenology will serve as a basis for the following discussions.

Due to space limitations, the formalization of the exposition is restricted to a minimum. Consider a random variable X assuming values $x \in \mathcal{X}$, \mathcal{X} the set of possible values for X . For simplicity, assume that \mathcal{X} is finite. Define the *entropy* of X by $H(X) := -\sum_{x \in \mathcal{X}} p(x) \log p(x)$, the *conditional entropy* of Y as $H(Y|X) := \sum_{x \in \mathcal{X}} p(x) H(Y|X = x)$ with $H(Y|X = x) := -\sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x)$ for $x \in \mathcal{X}$. The *joint entropy* of X and Y is the entropy of the random variable (X, Y) . The *mutual information* of random variables X and Y is $I(X; Y) := H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y)$. A generalization is the *intrinsic information*: for random variables X_1, \dots, X_k , the intrinsic information is $I(X_1; \dots; X_k) := \sum_{i=1}^k H(X_i) - H(X_1, \dots, X_k)$. This notion is also known e.g. as *integration* in [23]. Similar to the mutual information, it is a measure for the degree of dependence between the different X_i .

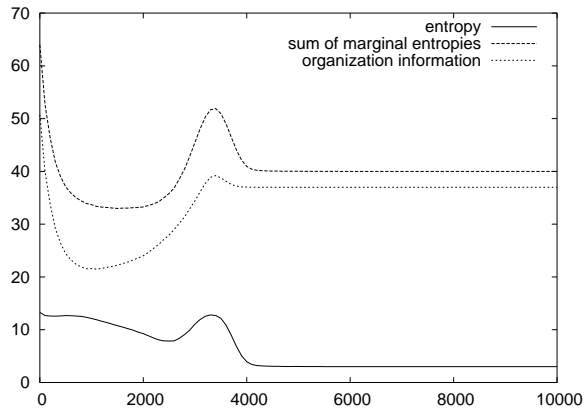


Fig. 3. Estimated entropy and information quantities for the SOM training process. For details see text (marginal entropies and organization information will be introduced in Secs. 3 and 3.)

We apply these notions to the discussion from Sec. 2. To obtain quantitative statements, we run 10000 SOM training runs of 10000 training steps each, using a SOM with units on a grid of 4×4 (as opposed to the 10×10 SOM used in Fig. 1). The entries for the total state vector variable $S = (X_{1,1}, X_{1,2}, \dots, X_{4,3}, X_{4,4}) \in \mathbb{R}^{2 \times 4 \times 4}$ (the 2 appears because weights are two-dimensional) are boxed into 4 possible values in each component. Via frequency statistics we compute empirical estimates for the different entropies. Starting with a SOM with random initial weight vectors, one aspect of the self-organization is the concentration of the probability distribution around the organized configurations during training. If the complete state configuration of the SOM at time t is given by $S^{(t)}$, then quantify this process by the *information gain* $H(S^{(t=0)}) - H(S^{(t=t_{\text{final}})})$. The information gain is related to Ashby's redundancy measure of self-organization $(H_{\text{max}} - H(S^{(t=t_{\text{final}})})) / H_{\text{max}}$ where H_{max} is the maximum possible entropy of the system [9]. In our case, the initial entropy of S is given by $H(S^{(t=0)}) = 4 \times 4 \log_2 4^2 = 64$ bit since there are 4×4 units, the weight of each can be in one of the 4^2 quantization boxes² The final entropy is $H(S^{(t=t_{\text{final}})}) = 3$ bit matching well the empirical values from Fig. 3, reflecting the logarithm of the 8 organized states.

The information gain only takes into account the probabilities at the beginning and the end of the training. It fails to detect the symmetry-breaking whereby the initial probability distribution splits into subprobabilities as in

² The initial value of ≈ 13.3 bit for the entropy in Fig. 3 at $t = 0$ is a strong underestimate, since the equidistribution in the high-dimensional state space is undersampled by the 10000 runs used. Only at $t = 2000 - 3000$ the probability distributions sufficiently concentrate for the empirical estimate of the entropies to become more accurate.

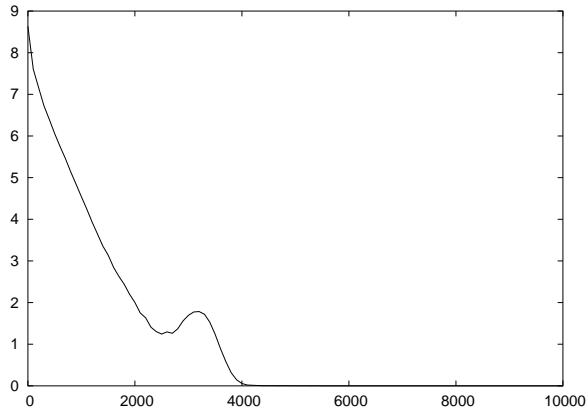


Fig. 4. Prediction entropy for the training process. For 1000 individual training runs, the single-step prediction entropy $H(S^{(t+1)}|S^{(t)})$ has been calculated for $S^{(t)}$ at different t . For a given instance $s^{(t)}$ of a state, 1000 individual single step probes were performed, obtaining the single-step prediction entropy for $H(S^{(t+1)}|s^{(t)})$. These were then averaged over all 1000 training runs.

Fig. 2. This split cannot be easily detected on the information-theoretic level unless trajectory history is brought into consideration³. If the system, after entering one of the subprobabilities regions, remains there with high probability, these can be considered stochastic attractor regions. This restriction can be detected by the *prediction entropy* $H(S^{(t+1)}|S^{(t)})$ which, for advanced t one expects to be smaller for systems with a history component than for systems whose history is quickly lost (e.g. for a system which jumps from branch to branch). Fig. 4 shows how the prediction entropy drops during training. The prediction entropy reflects the “freezing” of history. The bump at $t = 3000$ is due to the fact that in the initial part of the training process, the SOM contracts strongly, reducing the entropy, before beginning to spread its weights over $\mathcal{R} = [0, 1]^2$.

The prediction entropy is not yet sufficient to capture the whole essence of self-organization in the SOM and thus to be a promising candidate for general measures of self-organization. To see that, let us carry out a thought experiment. Modify the SOM training as to cause the SOM to always freeze in the same configuration at the end of the training. For the argument it is irrelevant whether this configuration is “organized” or a fixed random state $s^* \in [0, 1]^{2 \times 4 \times 4}$. Neither information gain nor prediction entropy are able to distinguish it from a training process generating the split probability distribution in Fig. 2.

³ This problem is also reflected by Ashby’s redundancy measure; that measure has the further problem that changed restrictions of the system induce a modification of H_{\max} .

3 Organization via Observers

This seems to indicate that self-organization cannot be defined via the intrinsic dynamics of the system and that one needs to assume additional structure for the system. Here, this additional structure will be an *observer*. The observer concept plays a role in obtaining entropies for physical systems [1]. It has been emphasized in the past that the observer concept may be of central importance to characterize self-organization or emergence [4, 8]. In [3, 18] the concept of emergence is being defined via observatory structures. Their approach uses the language of category theory and provides a highly general meta-mathematical framework to formulate the phenomenon of emergence. The generality of that approach, however, makes it harder to construct the pertinent notions and to evaluate them in a concrete system. It was felt that the less, but still sufficiently general information-theoretic perspective would be more directly applicable to practical purposes. Also, information theory offers a chance to provide *intrinsic* notions which do not require the introduction of additional structural language and which, at some point, might be derived as consequence directly from the system dynamics itself, which emphatically is not the case with the notions from [3, 18]. Here, we now combine the information-theoretic approach with the observer concept to characterize self-organization.

A (*perfect*) *observer* of a (random) variable S is a collection S_1, S_2, \dots, S_k of random variables allowing full reconstruction of S , i.e. for which $H(S|S_1, S_2, \dots, S_k)$ vanishes. Let the (*observed*) *organization information* be the intrinsic information $I(S_1; \dots; S_k)$. Call a system *self-organizing* if the (*observed*) *organization information increase* is positive during the progress of its dynamics. $I(S_1; \dots; S_k)$ quantifies the dependency between the observer variables. In some respect it is related to Ashby's redundancy measure, however taking into account only the "effectively used" state space.

For the SOM, a natural choice for an observer is to set S_1, S_2, \dots, S_k to X_1, X_2, \dots, X_k of the k individual SOM units. Consider three cases: First, a modified dynamics that lets the training begin with a random state $S^{(t=0)}$ and end with a random state $S^{(t=t_{\text{final}})}$ independent from $S^{(t=0)}$ but with the same distribution. In this case, both the information gain and the organization information will vanish for start and end state. Both will correctly identify this system as not self-organizing. Assuming random sampling in each step, the prediction entropy will always yield the full entropy $H(S^{(t=0)})$ of the initial state with no decrease indicating an attractor; using the quantization from the example from Sec. 2 and Fig. 3 gives a prediction entropy of 64bit. In the second case, let the dynamics start the SOM training with a random state and end up with a unique final state, as in the last example of Sec. 2. The information gain will be $H(S^{(t=0)}) - 0$ bit (64 bit in our above model). However, the organization information for both $S^{(t=0)}$ and $S^{(t=t_{\text{final}})}$ vanishes, as in the unique final state all the entropies vanish. Such a system with a single attractor point will not be regarded as self-organizing by the organization information. This is plausible since such a dynamics is doing nothing else than "freezing" the system. Finally, use the original SOM dynamics. The information gain is 64 bit $- 3$ bit = 61 bit, large, but

smaller than for the case of a unique final state, since there are 8 ordered final states. Thus, from the “information gain” point of view, the standard dynamics is less “self-organizing” than that freezing into a a unique state. The organization information $\sum_i H(S_i^{(t=0)}) - H(S^{(t=0)})$ vanishes in the random initial state $S^{(t=0)}$, however, since the individual (marginal) entropies $H(S_i^{(t=0)})$, $i = 1, \dots, k$ are $\log 4^2 = 4$ bit each and the total entropy is $H(S^{(t=0)}) = 64$ bit. Summarizing, while the information gain “sees” only the attractor structure of the dynamics and can not distinguish between a structureless point attractor and a rich attractor landscape inducing organization in its systems, and while the prediction entropy can only quantify the “historicity” of a state class, the organization information measure is able to identify just the organizational aspect of the development of the dynamics. Final remarks on the plots: due to the underestimation of $H(S^{(t=0)})$ in the simulation runs — see footnote 2 — the value for the organization information does not seem to vanish for $t = 0$ in Fig. 3. Thus, one has to keep in mind that the value of the organization information begins with value 0 bit at $t = 0$, growing to the point where the plot becomes more accurate after $t \lesssim 2000$.

The question how the notion of organization information depends on a given observer is still under research. At the present point a transformation formula that trans the organization information of a fine-grained observer to that of a coarse-grained observer is known. It can not be given here due to space limitations.

4 Summary and Current Work

We have introduced a notion of *organization information* to quantify a process of self-organization. It is defined for a given stochastic dynamical system for which an observer is specified. The properties of this measure have been compared with other information-theoretic measures and discussed using the Self-Organizing Map as scenario. The notion is, however, general and in its application not limited to the SOM. It is natural and appears to be versatile and sensitive to precisely the relevant self-organization processes.

On a practical level, it is envisaged to improve the numerical calculation of the quantities involved and to apply the notion to other systems of interest to validate its power. On a methodological level, the study of the influence of generalized observer change as well as the existence of “canonical”, i.e. natural observers is of particular interest. This interest is fueled by strong indications that it is possible to define natural observers via recently introduced information-theoretic notions of emergence [16]. If that proves feasible, this would lead not only to an intrinsic characterization of self-organization (requiring no externally given observers), but also clarify some of the deep structural relations between the notions of self-organization and emergence.

Acknowledgements The author would like to thank Peter Dauscher and the anonymous reviewers for helpful comments on the paper.

Bibliography

- [1] Adami, C., [1998]. *Introduction to Artificial Life*. Springer.
- [2] Ashby, W. R., [1947]. Principles of the self-organizing dynamic system. *J. Gen. Psychol.*, 37:125–128.
- [3] Baas, N. A., and Emmeche, C., [1997]. On Emergence and Explanation. *Intellectica*, 2(25):67–83.
- [4] Crutchfield, J. P., [1994]. The Calculi of Emergence: Computation, Dynamics, and Induction. *Physica D*, 11–54.
- [5] Erwin, E., Obermayer, K., and Schulten, K., [1992]. Self-Organizing Maps: ordering, convergence properties and energy functions. *Biol. Cybern.*, 67:47–55.
- [6] Goodhill, G. J., and Sejnowski, T. J., [1997]. A unifying objective function for topographic mappings. *Neural Computation*, 9:1291–1304.
- [7] Haken, H., [1983]. *Advanced synergetics*. Berlin: Springer-Verlag.
- [8] Harvey, I., [2000]. The 3 Es of Artificial Life: Emergence, Embodiment and Evolution. Invited talk at Artificial Life VII, 1.-6. August, Portland.
- [9] Heylighen, F., Joslyn, C., and Turchin, V., [2003]. Principia Cybernetica Web. <http://pespmc1.vub.ac.be>, March 2003
- [10] Jetschke, G., [1989]. *Mathematik der Selbstorganisation*. Braunschweig: Vieweg.
- [11] Kohonen, T., [1989]. *Self-Organization and Associative Memory*, vol. 8 of *Springer Series in Information Sciences*. Berlin, Heidelberg, New York: Springer-Verlag. Third edition.
- [12] Kohonen, T., [1997]. *Self-Organizing Maps*. Springer. Second edition.
- [13] Kushner, H. J., and Clark, D. S., [1978]. *Stochastic Approximation for Constrained and Unconstrained Systems*, vol. 26 of *Applied Math. Science Series*. Springer.
- [14] Pask, G., [1960]. The Natural History of Networks. In [26].
- [15] Polani, D., [2001]. Measures for the Organization of Self-Organizing Maps. In Jain, L., and Seiffert, U., editors, *Self-Organizing Neural Networks. Recent Advances and Applications*. Springer.
- [16] Polani, D., [2002]. On Individuality, Emergence and Information Preservation. In Nehaniv, C. L., and te Boekhorst, R., editors, *Proceedings of the Symposium on Evolvability and Individuality, 18-20 September 2002, St. Albans*. University of Hertfordshire.
- [17] Prigogine, I., and Nicolis, G., [1977]. *Self-Organization in Non-Equilibrium Systems: From Dissipative Structures to Order Through Fluctuations*. New York: J. Wiley & Sons.
- [18] Rasmussen, S., Baas, N., Mayer, B., Nilsson, M., and Olesen, M. W., [2001]. Ansatz for Dynamical Hierarchies. *Artificial Life*, 7:329–353.
- [19] Reichl, L., [1980]. *A Modern Course in Statistical Physics*. Austin: University of Texas Press.
- [20] Ritter, H., Martinetz, T., and Schulten, K., [1994]. *Neuronale Netze*. Addison-Wesley.
- [21] Shalizi, C. R., [1996]. Is the Primordial Soup Done Yet?. <http://www.santafe.edu/~shalizi/Self-organization/soup-done/>, Jan 23, 2001
- [22] Spitzner, A., and Polani, D., [1998]. Order Parameters for Self-Organizing Maps. In Niklasson, L., Bodén, M., and Ziemke, T., editors, *Proc. of the 8th Int. Conf. on Artificial Neural Networks (ICANN 98)*, Skövde, Sweden, vol. 2, 517–522. Springer.
- [23] Tononi, G., Sporns, O., and Edelman, G. M., [1994]. A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. USA*, 91:5033–5037.
- [24] Villmann, T., Der, R., Herrmann, M., and Martinetz, T., [1997]. Topology Preservation in Self-Organizing Feature Maps: Exact Definition and Measurement. *IEEE Trans. Neural Networks*, 8(2):256–266.
- [25] Walter, W. G., [1951]. A Machine that Learns. *Scientific American*, 60–63.
- [26] Yovits, M. C., and Cameron, S., editors, [1960]. *Self-Organizing Systems – Proceedings of an Interdisciplinary Conference, 5-6 May 1959*, Computer Science and Technology and their Application. Pergamon Press.