

On a Quantitative Measure for Modularity Based on Information Theory

Daniel Polani¹, Peter Dauscher², and Thomas Uthmann²

¹ University of Hertfordshire, Hatfield, UK

d.polani@herts.ac.uk

Web: <http://homepages.feis.herts.ac.uk/~comqdp1/>

² Johannes Gutenberg Universität Mainz, Germany

{dauscher,uthmann}@informatik.uni-mainz.de

Web: <http://www.informatik.uni-mainz.de/~dauscher/>

Abstract. The concept of modularity appears to be crucial for many questions in the field of Artificial Life research. However, there have not been many quantitative measures for modularity that are both general and viable. In this paper we introduce a measure for modularity based on information theory. Due to the generality of the information theory formalism, this measure can be applied to various problems and models; some connections to other formalisms are presented.

1 Introduction

In the studies of complex systems and Artificial Life, a central question is how it is possible that, over time, systems can emerge with ever increasing complexity. This question is particularly prominent if one considers the Darwinian evolution which, from a naive point of view, appears to be mainly directed random search with a large test population. However, even the powerful parallelism that is available to evolution in form of huge numbers of individuals cannot alone explain how the vast search space of possible configurations of living organisms can be efficiently searched and exploited towards increasing complexity.

It seems that nature employs to some degree the same method as human programmers in large software systems (or vice versa). In the latter, with the advent of the software crisis in the 1970s [1], it became clear that large monolithic software systems in which each part depends on many others (a form of nonlocality) are unmanageable. Even if they should work reasonably reliably at a certain point in time, they cannot easily be adapted to new tasks. This is being solved by introducing *modules* which solve subproblems independently from the rest of the system and organizing larger systems by building them up from these smaller, manageable modules.

Adaptability is one of the central motifs of natural evolution. Therefore, the question arises whether evolution manages complexity in a similar way as human software engineers, via modularity. It turns out that there are several phenomena in nature that can be construed as exhibiting elements of modularity. The

existence of genes that encode certain traits of the phenotype, the crossover operator (which is construed by researchers of artificial evolution as to be preserving *building blocks* which encode for separable (i.e. modular) properties of the phenotype. Of course, the situation in systems evolving in nature is much more involved as there is no human designer, but even in artificial software systems pure modularity does not exist in general.

As different as the different instances of systems are that exhibit or do not exhibit modularity and as different the language is that is being used in conjunction with those, they seem to share common properties. It would be very useful to formalize these properties in a common language. It would enable us to understand better what modularity is, when it can be made use of, or even when we can expect it to emerge [2–4] thus helping us to obtain further clues how natural evolution manages to climb the ladder of complexity.

It seems, however, that only relatively recently systematic approaches have been made to decompose or tackle the decompositional structures of complex systems [5]. In the wake of the success of information-theoretic methods in the study of dynamical systems [6], recent approaches to address the question of modular decomposition of networks convert static networks into dynamic systems via a diffusion dynamics approach and analyse it applying spectral graph and information theory [7, 8]. Independence graphs derived from probabilistic relations [9] and related information-theoretical notions [10] provide a growing toolbox to address these questions.

The specific systems we will address here already have an a priori dynamical structure and do not require it to be artificially imposed, as done in above models. At the same time, they are of high relevance for Artificial Life studies. We begin with the illustrative and inspiring model from [11] and show how information theory can be used in order to adapt its classifications as to become both more intuitive and finer-grained. We will then relate this approach to modularity arising due to the variation operator in Evolutionary Algorithms. For a special case, we will establish a direct connection between the measure of coupling introduced here and the *modularity matrix elements* in [2, 4].

2 Towards a Quantitative Notion for Modularity

In a recent discussion of possible characterizations of modularity by Watson, several illustrative scenarios are presented that highlight different important aspects of the issue [11]. Since we felt it offers many relevant points and fruitful ideas, in the present paper we wish to build upon some of these discussions, offer alternative formal notions of modularity and study some of their conceptual and quantitative properties.

2.1 Probability Notation

We will apply the following notation: for a random variable X we will denote its domain by \mathcal{X} and a concrete sample value of X will be denoted by x . Let now

X, Y be jointly distributed random variables. Let $P(X = x)$ be the probability that X assumes the value $x \in \mathcal{X}$ for which, by abuse of notation, we write instead $p(x)$ wherever this is unambiguous. Similarly, let $P(X = x, Y = y) \equiv p(x, y)$ be the joint and $P(Y = y | X = x) \equiv p(y|x)$ the conditional distribution.

2.2 Watson's Regulatory Network scenario

One of the scenarios introduced by Watson is a (stochastic) dynamical system which can be seen as a very simple model for a regulatory network [11]. Consider thus a stochastic dynamical system of four random variables S_1, S_2, S_3, S_4 , each of them binary valued, i.e. $S_i \in \{0, 1\}$ for $i \in \{1, 2, 3, 4\}$. Write $S = (S_1, S_2, S_3, S_4)$ for the random variable denoting the complete system which therefore can assume 16 states. Write $S(t)$ for the whole system or $S_i(t)$ for a single variable at a time t . Consider the dynamics

$$P(S_i(t+1) = 1 | S(t) = s(t)) = \sum_j w_{ij} s_j(t) \quad (1)$$

$$P(S_i(t+1) = 0 | S(t) = s(t)) = 1 - P(S_i(t+1) = 1 | S(t) = s(t)) \quad (2)$$

where w_{ij} is the weight by which the variable S_j influences S_i . To guarantee well-defined probabilistic expressions, the weights in Eq. (1) are normalized such that $\sum_j w_{ij} = 1$. The discussion by Watson now strives to study the system as composed of two subsystems (“modules”). He achieves this by splitting up the system into two groups of variables (subsystems, “modules”), $M_1 = (S_1, S_2)$ and $M_2 = (S_3, S_4)$ and using different coupling inside and between the subsystems. Watson considers a system with $w_{ij} = 4/10$ if i and j belong to the same subsystem (including $i = j$) and $w_{ij} = 1/10$ if they belong to different subsystems. In this example, increasing coupling reinforces the probability that different S_i will assume the same state. Note that, although designed as two separate coupled subsystems, the question arises here whether M_1 and M_2 can be considered separate subsystems from a *dynamical* point of view.

In principle, $C = 4$ states are possible for each of the subsystems M_1 and M_2 . Consider now only systems that converge into a fixed point attractor; there might be several such attractors. Watson distinguishes three kinds of properties for such a system: *non-decomposability*, *separability* and *decomposability but no separability*. Watson concentrates on the subsystem M_1 and asks about the “most stable states”, i.e. the states of M_1 found if the entire system has converged to one of its attractors. Paraphrased from [11], non-decomposability would mean that “for every configuration of M_1 there is some configuration of M_2 (the remainder of the system) that would make that configuration of M_1 the most stable”.

In this case, there is a one-to-one relationship between the two subsystems after convergence. Watson characterizes this case by the number C' of possible states of M_1 after the system has converged. In this case $C' = C = 4$, i.e. every possible state of M_1 can be an attractor, depending on the rest. A possible set of attractors fulfilling this property could be $a_1 = (0000)$; $a_2 = (0101)$; $a_3 = (1010)$; $a_4 = (1111)$ where the first two bits correspond to M_1 and the last two bits to M_2 . In the other extreme, *separability*, “the module is fully independent

in the property of interest” [11]. Watson characterizes this case of separability by the fact that M_1 can only have one single state ($C' = 1$). Indeed, in this case (which means that all attractors of the total system are identical w.r.t. the bits of M_1), the attractor state of M_1 does not depend at all on M_2 . An example attractor set for this case would consist of $a_1 = (01\ 00)$ and $a_2 = (01\ 11)$, with M_1 attaining a single value 01. Finally, the intermediate case of *decomposability but non-separability* is now characterized by Watson via $1 < C' < C$. An attractor set showing this property would consist of $a_1 = (11\ 11)$ and $a_2 = (00\ 00)$ where one has $C' = 2$. As Watson points out, one can see “that there is something we know about the property of interest, the most stable configurations (00 or 11), that is independent of inter-module interactions.” [11]

2.3 Open Problems

Although Watson’s method of counting the configurations of interest, i.e. the attractors, has some attractive features, two problems remain: 1. The measure is not continuous; therefore continuous changes in coupling cannot be measured appropriately. 2. As we will show in the following, the number C' does not always correspond to an intuitive classification of *separability* or (*non-*) *decomposability*. Problem 2 can easily be seen by considering the following (not previously discussed) attractor structures:

Case 1: assume the two subsystems were not coupled at all and the corresponding attractors are $a_1 = (00\ 00)$; $a_2 = (00\ 11)$; $a_3 = (11\ 00)$; $a_4 = (11\ 11)$ and assumed with equal probability. As the attractors for the individual subsystems are independent of each other, this should be ideally considered as *separable*; in Watson’s framework, however, we have $C' = 2$, hence $1 < C' < C$, and thus the system is classified as *decomposable but not separable*. **Case 2:** Assume that each module would converge into all of its 4 configurations. This would be e.g. the case if each binary variable $S_i, i \in \{1, 2, 3, 4\}$ was only fed back to itself in a positive way: Then all 16 combinations could be attractors for the total system. For both modules we would observe $C' = 4$ classifying the system as *non-decomposable*. Intuitively, however, we would classify the system as *separable*, again as there is no coupling at all between the individual variables and thus between the subsystems.

In the following, we will suggest how to amend these misclassifications as well as how to provide a continuous quantification of Watson’s modularity classes of *separability*, *non-decomposability* and *decomposability but not separability* by using concepts from information theory.

3 Quantification of Dynamical Modularity

3.1 Information-Theory: Notation

We need some further notions. Define the *entropy* of a random variable X by $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$ It denotes the expected uncertainty about

a single outcome of X [12]. If the logarithm is chosen w.r.t. base 2, the entropy is quantified by bits; in the following, wherever this value vanishes, we will interchangeably write 0 or 0 bit. Another important quantity is the conditional entropy between two random variables X, Y which is given by $H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) = - \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x)$ and quantifies how much expected uncertainty in Y remains if X is known. The difference of two related entropies often has the character of entropy or uncertainty loss, or information gain. Specifically, the difference between $H(Y)$ and $H(Y|X)$ is known as the *mutual information* $I(X; Y) = H(Y) - H(Y|X)$ and quantifies how much the knowledge of X adds about the knowledge of Y . The mutual information is symmetric in X and Y and the following relation holds: $I(X; Y) = H(X) + H(Y) - H(X, Y)$. We will use this relation later on.

3.2 An Intuitive Quantification

We will show that mutual information can serve as a modularity measure which is both, intuitive and quantitative. Consider a system with fixed dynamics following Eq. (1) and consider the random variables M_1 and M_2 , denoting the first or second subsystem, and taking on values $\{00, 01, 10, 11\}$. Now we measure the coupling by considering the mutual information $I(M_1; M_2)$ which replaces the counting variable C' in Watson's original model. We can now generalize Watson's modularity classes in the following definition.

Definition 1 (Generalized Modularity Classes). *Consider a stochastic dynamical system and assume its attractor is split into an subsystem attractor M_1 and an attractor for the rest M_2 . Let $I(M_1; M_2)$ be the mutual information between the two attractor random variables. Then call M_1 separable, if $I(M_1; M_2) = 0$; non-decomposable, if $I(M_1; M_2) = H(M_1)$; decomposable but not separable, if $0 < I(M_1; M_2) < H(M_1)$*

Table 1 summarizes the cases we have considered so far. One observes that, while the classification by counting (C') sometimes deviates from what one would intuitively expect, modularity classification based on mutual information meets intuition in all of the cases. It should be mentioned that $I(M_1; M_2)$ does not measure modularity, but its opposite, coupling. A perfectly modular system would be characterized by separability, i.e. $I(M_1; M_2) = 0$.

4 Experiments

It is instructive to see the operation of the measure from Def. 1 in the concrete scenario. For this, consider the dynamical system Eq. (1). For didactical reasons, we use coupling strengths different from [11]: we set $w_{ij} = 1/N$ if i and j belong to the same subsystem and $w_{ij} = c/N$ if they belong to different subsystems (with $N = 2(1 + c)$ a normalization term); $c \in [0, 1]$ is a dynamical coupling strength between the subsystems — if $c = 0$, the subsystems are entirely uncoupled, if $c = 1$, there is no dynamical separation between the subsystems.

Scenario	Intuitive Classif.	C'	C' -Class.	Entropies	$I(M_1; M_2)$	I -Class.
Selecting attractor state M_2 , one can obtain any attractor state M_1	non-decomposable (n.-d.)	4	n.-d.	$H(M_1) = 2$ $H(M_2) = 2$ $H(M_1, M_2) = 2$	2	n.-d.
M_1 can only converge to one single state	separable (s.)	1	s.	$H(M_1) = 0$ $H(M_1, M_2) = H(M_2)$	0	s.
M_1 and M_2 can independently converge to one of two attractors (Case 1)	separable (s.)	2	d.b.n.s.	$H(M_1) = 1$ $H(M_2) = 1$ $H(M_1, M_2) = 2$	0	s.
M_1 and M_2 converge to one of two attractors, but not independently	decomposable but not separable (d.b.n.s.)	2	d.b.n.s.	$H(M_1) = 1$ $H(M_2) = 1$ $0 < H(M_1, M_2) < 2$	> 0 but < 2	d.b.n.s.
Each bit behaves independently (Case 2)	separable (s.)	4	n.-d.	$H(M_1) = 2$ $H(M_2) = 2$ $H(M_1, M_2) = 4$	0	s.

Table 1. Summary of the different cases considered with the values for C' and $I(M_1; M_2)$ and the corresponding classifications. In contrast to the classification based on C' (C' -Class), the mutual information $I(M_1; M_2)$ (I -Class) classifies all of the cases according intuition.

We run 100,000 independent simulations of the system for different values of the coupling c obtaining empirical distributions for the attractors of the system (as heuristics derived from experiments, we considered a state to be an attractor and ended the run if the system stays in this state for 10 time steps). Several mutual information quantities obtained for these states are shown in Fig. 1. The results show that for small c also the dynamical coupling $I(M_1; M_2)$ is small, as expected, and it grows continuously with growing c (thus fulfilling the continuity requirement from Sec. 2.3). It reaches a value close to 1 bit for $c = 1$ (the fully coupled system) reflecting the fact that in this case the system ends mostly in the equally distributed stochastic “attractors” $S = (0, 0, 0, 0)$ and $S = (1, 1, 1, 1)$ ³. In this case, we do not expect a natural split of the system in two subsystems.

Up to now we have always assumed that we do know the specific decomposition of the system into subsystems by knowing the form of w_{ij} and the particular structure of the network. However, our measure provides us with a

³ Because of finite-size effects in the dynamics of (1) and (2), other states are also found with some small probability, thus for $c = 1$ the values for $I(M_1; M_2)$ and $I(S_1, S_3; S_2, S_4)$ drop slightly below 1 bit.

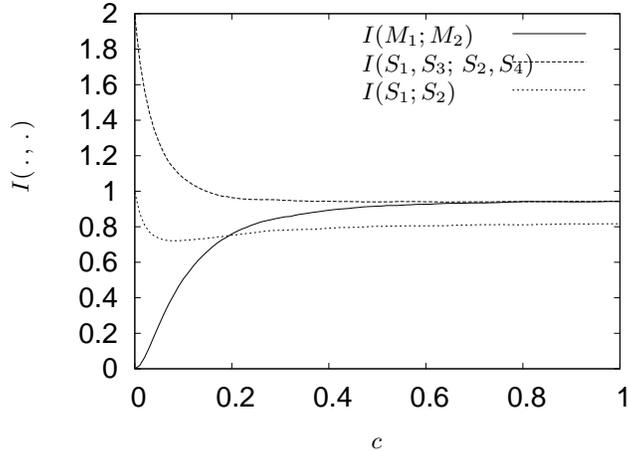


Fig. 1. The plots show the dynamic coupling $I(M_1; M_2)$ between the two subsystems, depending on the coupling constant c (the x -axis); in addition, the mutual information between the crosswise mixed system (S_1, S_3) and (S_2, S_4) as well as the mutual information between the variables (S_1, S_2) inside the system M_1 is shown. For details see text.

way to separate the subsystems without using this knowledge. Plotting the mutual information between alternative subsystems of S , Fig. 1 shows, e.g. the crosswise mutual information $I(S_1, S_3; S_2, S_4)$ arising from an alternative split of S into subsystems. For small c , the value is almost 2 bit, far from 0, indicating that this split does not provide natural modules⁴. On the other hand, once c becomes close to 1, this crosswise information converges towards the value of the coupling information, indicating that the split into S_1, S_2 and S_3, S_4 becomes indistinguishable from S_1, S_3 and S_2, S_4 ; this is expected, as for $c = 1$ both splits are dynamically equivalent.

Another interesting measure is the intrinsic information in $M_1 = (S_1, S_2)$, i.e. $I(S_1, S_2)$. For $c = 0$, this becomes 1 bit indicating that the variables are fully coupled in the subsystem. Increasing c first slightly reduces the coupling as the subsystem is being perturbed, but then the overall dynamics “stiffens” and again creates a correlation between the variables. However, when $I(M_1; M_2)$ grows, it indicates that modularity is increasingly lost. Once this value becomes larger than $I(S_1, S_2)$, M_1 can be basically seen as having lost its identity as a module, because the coupling between the module and its environment is stronger than the intrinsic coupling. Other combinations can also be compared. This illustrates how the information-theoretic measures provide a whole family of useful characterizations of modularity.

⁴ For the extreme case $c = 0$, S_1 is fully aligned with S_2 and S_3 with S_4 , thus $I(S_1, S_3; S_2, S_4) = 2$ bit.

5 Applicability to Evolutionary Operators

To round up the argument, we will show that our concepts apply directly to issues of modularity of individuals in Evolutionary Algorithms. To this purpose, we consider a measure $m_{jk}(\psi)$ from [2, 4] which quantifies the modularity of mating two individuals of given genotypes $j \in \mathcal{S}$ and $k \in \mathcal{S}$, where \mathcal{S} is the search space and ψ is an equivalence relation (more details below). This formalism combines the algebraic formalism from [13] with the dynamical system formalism from [14] and has been used to study under which circumstances self-organization of modularity occurs. In that formalism modularity is always considered w.r.t. a specific equivalence relation⁵ ψ defined over the search space \mathcal{S} . We will briefly sketch the formalism and show how it fits neatly into the framework developed in the earlier sections. Define then the *modularity matrix element* by

$$m_{jk}(\psi) = \sum_{i \in \mathcal{S}} p(i|j, k) r_{i \sim_{\psi} jk}(\psi) \quad . \quad (3)$$

Here $p(i|j, k)$ is the probability that crossover and mutation will generate an offspring individual of type i by mating of individuals of types j and k ⁶; $r_{i \sim_{\psi} jk}$ is a binary indicator variable describing whether i is equivalent to one of its parents j or k , i.e. $r_{i \sim_{\psi} jk}(\psi) = 1$ if $i \sim_{\psi} j$ or $i \sim_{\psi} k$ and 0 else⁷. Thus, for parents of types j and k , the quantity $m_{jk}(\psi)$ measures the expected degree to which their offspring is equivalent to one of the parents (with respect to ψ). The relation ψ can e.g. be used to model a number of relevant concepts, e.g. traits (like certain phenotypical properties, size, strength, etc.) or schemata [15, 16]. It is important to note that the formalism makes no assumption whatsoever about the underlying variation operator(s) or about the representation, e.g. bit-strings, GP-trees etc.

Consider now a simple example of a bitstring-based Genetic Algorithm and two of the bits of the individuals, $b^{(1)}$ and $b^{(2)}$; the genotype of an individual has therefore the following form: $(\dots b^{(1)} \dots b^{(2)} \dots)$. Let an individual of type $j = (\dots 1 \dots 1 \dots)$ and an individual of type $k = (\dots 0 \dots 0 \dots)$ be given. Let furthermore ψ be the equivalence relation which declares two types equivalent iff they are equal in both $b^{(1)}$ and $b^{(2)}$. Then $m_{jk}(\psi)$ becomes the probability that an offspring individual looks completely like j or completely like k with respect to these two bits: $m_{jk}(\psi) = p(\dots 0 \dots 0 \dots |j, k) + p(\dots 1 \dots 1 \dots |j, k)$.

Here, we consider a one-point crossover with $0 \leq p_{\text{crossover}} \leq 0.5$ (for larger crossover probabilities one experiences mirroring effects which we will ignore for the discussion). If we assume symmetry in the variation (i.e. neither $(\dots 0 \dots 0 \dots)$ nor $(\dots 1 \dots 1 \dots)$ is preferred by the variation), the definition of $m_{jk}(\psi)$ immediately leads to $p(\dots 0 \dots 0 \dots |j, k) = p(\dots 1 \dots 1 \dots |j, k) = \frac{1}{2}m_{jk}(\psi)$ and,

⁵ An equivalence relation is a relation \sim (“equivalent to”) having the properties of reflexivity ($\forall x : x \sim x$), symmetry ($\forall x, y : x \sim y \Leftrightarrow y \sim x$) and transitivity ($\forall x, y, z : x \sim y \wedge y \sim z \Rightarrow x \sim z$).

⁶ In [14] and [2, 4], $p(i|j, k)$ is called the *transmission function* and written as $T(i \leftarrow j, k)$.

⁷ In [2, 4], this is denoted as $r_{ijk}(\psi)$.

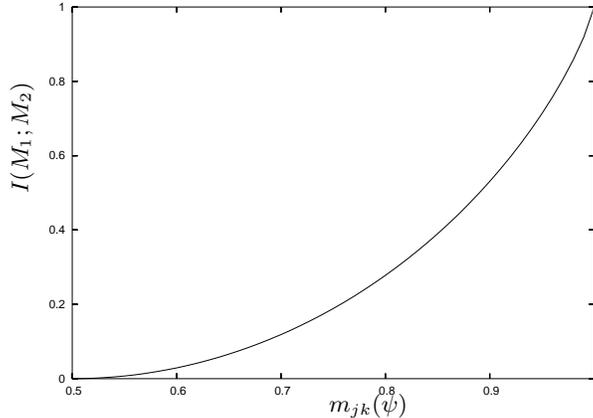


Fig. 2. Relation between $I(M_1; M_2)$ and $m_{jk}(\psi)$ in the example presented.

with similar assumptions, $p(\dots 1 \dots 0 \dots |j, k) = p(\dots 0 \dots 1 \dots |j, k) = \frac{1}{2}(1 - m_{jk}(\psi))$ holds. Most common crossover operators, as N -point-crossover or uniform crossover fulfill this requirement.

We now create the connection to the approach to measure modularity in the dynamical system. Let B_1 and B_2 be the random variables associated to the probabilities of values for the two bits $b^{(1)}$ and $b^{(2)}$. From the symmetry assumptions, it is clear that

$$P(b^{(i)} = 0) = P(b^{(i)} = 1) = 0.5 \quad i = 1, 2 \quad (4)$$

and therefore $H(B_1) = H(B_2) = 1$. As the above equations show the joint distribution of both bits depends on $m_{jk}(\psi)$, and thus the mutual information. Using above relations for $p(\dots 0 \dots 0 \dots |j, k)$ and $p(\dots 1 \dots 0 \dots |j, k)$, the mutual information can be computed. The functional dependency between $I(B_1; B_2)$ and $m_{jk}(\psi)$ is shown in Fig. 2. It corresponds to the intrinsic information $I(S_1; S_2)$ from Sec. 4 which quantifies how strongly the variables of a specific subsystem are coupled and shows that increasing modularity corresponds to stronger coupling of the intrinsic variables. An information-theoretic measure for inter-module coupling in that scenario can also be formulated, but it is more technically involved and will have to be discussed elsewhere for lack of space.

6 Conclusion and Future Work

We have presented two different directions of thrust towards a formalization of modularity. For this purpose, we have discussed Watson's example of a dynamical system and used information theory to obtain a generalized version of Watson's modularity classes. In particular, we were able to classify properly several cases that are unintuitive under Watson's C' counting approach. In addition, we were

able to reconstruct the modular structure of the regulatory network by looking at the mutual information between different parts of the system (Fig. 1).

In the other line of thrust, we were able to create a connection between the measure $m_{jk}(\psi)$ of modularity by which self-organized modularity in evolving systems has been studied in earlier work and the information-theoretic picture of dynamical systems modularity developed here. This indicates that the present approach can be extended to become a powerful tool to establish when and how modularity arises and, perhaps, to design Alife systems that are able to evolve their own modular decompositions in a targeted manner and thus are able to move more swiftly towards higher rungs in the ladder of complexity.

References

1. W. Zuser, S. Biff, T. Grechenig, and M. Köhle. *Software Engineering*. Pearson Studium, 2001.
2. P. Dauscher and T. Uthmann. On self-organizing modularization of individuals in evolutionary scenarios. In D. Polani, J. Kim, and T. Martinetz, editors, *Proc. Fifth German Workshop on Artificial Life, March 18-20, 2002, Lübeck, Germany*. Akademische Verlagsgesellschaft Aka GmbH, Berlin, 2002.
3. P. Dauscher. *Selbstorganisierte Modularisierung von Individuen in Evolutionären Algorithmen*. PhD thesis, Johannes Gutenberg-Universität Mainz, 2003.
4. P. Dauscher and T. Uthmann. Self-organized modularity in evolutionary algorithms. *Evolutionary Computation*, 2005. To appear.
5. Stefan Winter. Zerlegung von gekoppelten Dynamischen Systemen (Decomposition of Coupled Dynamical Systems). Diploma thesis, Johannes Gutenberg-Universität Mainz, 1996. (In German).
6. Gustavo Deco and Bernd Schuermann, editors. *Information Dynamics: Foundations and Applications*. Springer, 2001.
7. E. Ziv, M. Middendorf, and C. Wiggins. An information-theoretic approach to network modularity. *Physical Review E*, In press, 2005. arXiv:q-bio.QM/0411033v1.
8. K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen. Modularity and extreme edges of the internet. *Phys. Rev. Lett.*, 90:148701, 2003.
9. P. Magwene. New tools for studying integration and modularity. *Evolution*, 55(9):1734–1745, 2001.
10. Thomas Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, 2000.
11. R.A. Watson. Modular interdependency in complex dynamical systems. In Bilotta et al., editors, *Workshop Proceedings of the 8th International Conference on the Simulation and Synthesis of Living Systems, UNSW Australia, December 2002*, 2003.
12. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.
13. Nicholas J. Radcliffe. The Algebra of Genetic Algorithms. *Annals of Maths and Artificial Intelligence*, 10(4), 1994.
14. Lee Altenberg. The evolution of evolvability in genetic programming. In J. K. E. Kinnear, editor, *Advances in Genetic Programming*. MIT Press, 1994.
15. D.E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley, 1989.
16. M. Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, 1996.