

Complexity: Quantity or Quality

Russell K. Standish, *Department of Mathematics and Statistics, University of New South Wales (UNSW), Kensington, NSW, Australia*

ABSTRACT

The term complexity is used informally both as a quality and as a quantity. As a quality, complexity has something to do with our ability to understand a system or object—people understand simple systems, but not complex ones. On another level, complexity is used as a quantity, when people talk about something being more complicated than another. In this article, the author explores the formalisation of both meanings of complexity, which happened during the latter half of the twentieth century.

Keywords: Complexity, Complex Systems, Emergence, Graph Theory, Information Theory, Occam's Razor

INTRODUCTION: IS COMPLEXITY A QUALITY OR A QUANTITY?

The term *Complexity* has two distinct usages, which may be categorised simply as either a quality or a quantity. We often speak of complex systems as being a particular class of systems that are difficult to study using traditional analytic techniques. We have in mind that biological organisms and ecosystems are *complex*, yet systems like a pendulum, or a lever are simple. Complexity as a *quality* is therefore what makes the systems complex.

However, we may also speak of complexity as a quantity—with statements like a human being being more complex than a nematode worm, for example. Under such usage, complex and

simple systems form a continuum, characterised by the chosen complexity measure.

Bruce Edmonds (1999) performed a comprehensive survey of complexity measures as part of his PhD thesis, but it has not been updated to include measures proposed since that time. However, it remains one of the most comprehensive resources of complexity measures available to date.

The aim of this article is not to provide a catalogue of complexity measures, but rather to select key measures and show how they interrelate with each other within an overarching information theoretic framework. This article extends a chapter written by the author some time ago (Standish, 2008), to include more recent developments in this field, most notably in the area of graph complexity.

DOI: 10.4018/ijsss.2014010103

COMPLEXITY AS A QUANTITY

We have an intuitive notion of complexity as a quantity; we often speak of something being more or less complex than something else. However, capturing what we mean by complexity in a formal way has proved far more difficult, than other more familiar quantities we use, such as length, area and mass.

In these more conventional cases, the quantities in question prove to be decomposable in a linear way, ie a 5cm length can be broken into 5 equal parts 1 cm long; and they can also be directly compared a mass can be compared with a standard mass by comparing the weights of the two objects on a balance.

However, complexity is not like that. Cutting an object in half does not leave you with two objects having half the complexity overall. Nor can you easily compare the complexity of two objects, say an apple and an orange, in the same way you can compare their masses.

The fact that complexity includes a component due to the interactions between subsystems rapidly leads to a combinatorial explosion in the computational difficulty of using complexity measures that take this into account. Therefore, the earliest attempts at deriving a measure simply added up the complexities of the subsystems, ignoring the component due to interactions between the subsystems.

The simplest such measure is the *number of parts* definition. A car is more complex than a bicycle, because it contains more parts. However, a pile of sand contains an enormous number of parts (each grain of sand), yet it is not so complex since each grain of sand is conceptually the same, and the order of the grains in the pile is not important. Another definition used is the *number of distinct parts*, which partially circumvents this problem. The problem with this idea is that a shopping list and a Shakespearian play will end up having the same complexity, since it is constructed from the same set of parts (the 26 letters of the alphabet—assuming the shopping list includes items like zucchini, wax and quince, of course). An even bigger problem is to define precisely

what one means by “part”. This is an example of the *context dependence* of complexity, which we’ll explore further later.

Bonner and McShea have used these (organism size, number of cell types) and other *proxy* complexity measures to analyse complexity trends in evolution (Bonner, 1988; McShea, 1996), and in a later paper, *hierarchical maximum* (number of hierarchy levels, similar to the *graph radius* measure introduced in the next section) (Marcot & McShea, 2007). They argue that all these measures trend in the same way when figures are available for the same organism, hence are indicative of an underlying organism complexity value. This approach is of most value when analysing trends within a single phylogenetic line, such as the diversification of trilobites.

GRAPH THEORETIC MEASURES OF COMPLEXITY

Since the pile of sand case indicates complexity is not simply the number of components making up a system, the relationships between components clearly contribute to the overall complexity. One can start by caricaturing the system as a *graph*—replacing the components by abstract *vertices* or *nodes* and relationships between nodes by abstract *edges* or *arcs*.

Graph theory (Diestel, 2005) was founded by Euler in the 18th century to solve the famous Königsberg bridge problem. However, until the 1950s, only simple graphs that could be analysed in toto were considered. Erdős and Renyi (1959) introduced the concept of a *random graph*, which allowed one to treat large complex graphs statistically. Graphs of various sorts were readily recognised in nature, from food webs, personal or business contacts, sexual relations and the Internet amongst others. However, it soon became apparent that natural networks often had different statistical properties than general random graphs. Watts and Strogatz (1998) introduced the *small world* model, which has sparked a flurry of activity in recent years to measure networks such as the

Internet, networks of collaborations between scientific authors and food webs in ecosystems (Albert & Barabási, 2002).

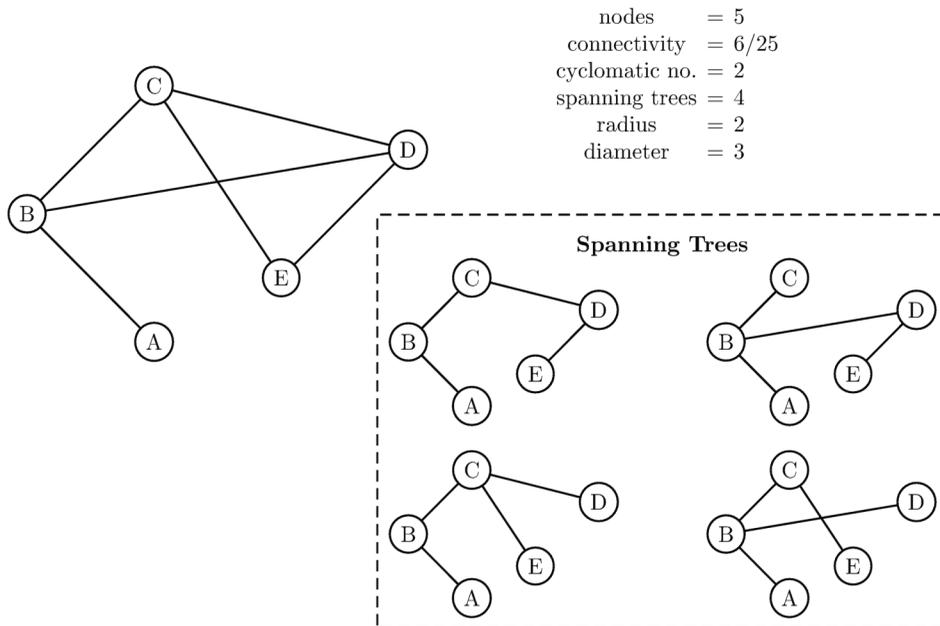
Graph theory provides a number of measures that can stand in for complexity (see Figure 1). The simplest of these is the *connectivity* of a graph, namely the number of edges connecting vertices of the graph. A fully connected graph, however, is no more complex than one that is completely unconnected. As connectivity increases from zero, a *percolation threshold* is reached where the graph changes from being mostly discontinuous to mostly continuous. The most complex systems tend to lie close to the percolation threshold. Another graph measure used is *cyclomatic number* of a graph, basically the number of independent loops it contains. The justification for using cyclomatic number as a measure of complexity is that feedback loops introduce nonlinearities in the system's behaviour, that produce complex behaviour.

Related to the concept of cyclomatic number is the *number of spanning trees* of the graph.

A spanning tree is a subset of the graph that visits all nodes but has no loops (ie is a tree). A graph made up from several disconnected parts has no spanning tree. A tree has exactly one spanning tree. The number of spanning trees increases rapidly with the cyclomatic number.

Another useful measure related to complexity is the height of the flattest spanning tree, known as the *graph radius*. Equivalently the maximum number of hops separating two nodes on the graph, known as the *graph diameter* can be used. This idea was popularised in the movie *six degrees of separation* -- which refers to the maximum number of acquaintances connecting any two people in the World. Networks having small diameter relative to their node count (so called *small world networks*) tend to support more complex dynamics than networks having a large diameters. The reason is that any local disturbance is propagated a long way through a small world network before dying out, giving rise to chaotic dynamics, whereas in the other

Figure 1. Various graph theoretic measures for a simple graph. The spanning trees are shown in the dashed box.



networks, disturbances remain local, leading to simpler linear dynamics.

Diameter and radius are not simply related for graphs (unlike with circles), about all that can be said is that $r \leq d \leq 2r$, but then only for graphs where the hop count forms a proper metric, which needn't be the case for directed graphs (Gliviak & Kyš, 1995).

Offdiagonal Complexity

Claussen (2007) introduced a complexity measure called *offdiagonal complexity* that is low for regular and randomly connected graphs, but takes on extremal values for *scale-free* graphs, such as typically seen in naturally occurring networks like metabolic and foodweb networks, the internet, the world wide web and citation networks. The apparent ubiquity of the scale-free property amongst networks we intuitively associate as complex (Newman, 2003) is the justification for using offdiagonal complexity, the other advantage being its computational practicality.

To compute offdiagonal complexity, start with the *adjacency matrix*:

$$g_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Let $\ell(i)$ be the node degree of i , and let $c_{mn}, m \leq n$, be the number of edges between all nodes i and j with node degrees $m = \ell(i)$, $n = \ell(j)$:

$$c_{mn} = \sum_{ij} g_i \delta_{m, \ell(i)} \delta_{n, \ell(j)} \quad (2)$$

Then the normalised n -th diagonal sum is:

$$a_n = \frac{\sum c_{i, i+n}}{\sum c_{ij}} \quad (3)$$

The *offdiagonal complexity* is defined by a Boltzmann-Gibbs entropy-like formula over the normalised diagonal sums:

$$C_{\text{offdiag}} = -\sum_n a_n \ln a_n \quad (4)$$

For regular lattices, each node has the same link degree, so c_{mn} is diagonal, $a_n = \delta_{n0}$ and $C_{\text{offdiag}} = 0$.

For random graphs, most edges will connect nodes with similar link degree (the characteristic link degree scale), so c_{mn} will have a mostly banded structure, and $a_n \rightarrow 0$ as n increases. This leads to small non-zero values of the offdiagonal complexity.

Scale free networks have a power law distribution of link degree, which leads to the c_{mn} matrix having a wide spread of entries. In the case of all a_n being equal, offdiagonal complexity takes its maximum value as equal to the number of nodes.

INFORMATION AS COMPLEXITY

The single simplest unifying concept that covers all of the preceding considerations is *information*. The more information required to specify a system, the more complex it is. A sandpile is simple, because the only information required is that it is made of sand grains (each considered to be identical, even if they aren't in reality), and the total number of grains in the pile. However, a typical motorcar requires a whole book of blueprints in its specification.

Prokopenko et al. (2009) undertook a recent review of information theoretic concepts as applied to the topics of complexity, self-organisation and emergence, which is a useful resource for exploring this topic.

Information theory began in the work of Shannon (1949), who was concerned with the

practical problem of ensuring reliable transmission of messages. Every possible message has a certain probability of occurring. The less likely a message is, the more information it imparts to the listener of that message. The precise relationship is given by a logarithm:

$$I = -\log_2 p \quad (5)$$

where p is the probability of the message, and I is the information it contains for the listener. The base of the logarithm determines what units information is measured in—base 2 means the information is expressed in *bits*. Base 256 could be used to express the result in *bytes*, and is of course equivalent to dividing Equation 5 by 8.

Shannon, of course, was not so interested in the semantic content of the message (i.e. its meaning), rather in the task of information transmission so instead considered a message composed of symbols x_i drawn from an alphabet A . Each symbol had a certain probability $p(x_i)$ of appearing in a message—consider how the letter “e” is far more probable in English text than the letter “q”, thus carrying $-\log_2 p(x_i)$ bits of information. These probabilities can be easily measured by examining extant texts. The entropy, or average information carried by the string of symbols $\{x_i\}$ (ignoring letter order) is given by:

$$H(x_1 x_2 \dots x_n) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (6)$$

This equation can be refined by considering an alphabet based on possible pairs of letters, then possible triplets, in the limit converging on the minimum amount of information required to be transmitted in order for the message to be reconstructed in its original form (given by Equation 5). That this value may be considerably less than just sending the original message in its

entirety is the basis of compression algorithms, such as those employed by the well-known *gzip* or *PKzip* (aka WinZip) programs.

The issue of semantic content discouraged a lot of people from applying this formalism to complexity measures. The problem is that a message written in English will mean something to a native English speaker, but be total gibberish to someone brought up in the Amazon jungle with no contact with the English speaking world. The information content of the message depends on exactly who the listener is! Whilst this context dependence appears to make the whole enterprise hopeless, it is in fact a feature of all of the measures discussed so far. When counting the number of parts in a system, one must make a decision as to what exactly constitutes a part, which is invariably somewhat subjective, and needs to be decided by consensus or convention by the parties involved in the discussion. Think of the problems in trying decide whether a group of animals is one species of two, or which genus they belong to. The same issue arises with the characterisation of the system by a network. When is a relationship considered a graph edge, when often every component is connected to every other part in varying degrees.

However, in many situations, there appears to be an obvious way of partitioning the system, or categorising it. In such a case, where two observers agree on the same way of interpreting a system, then they can agree on the complexity that system has. If there is no agreement on how to perform this categorisation, then complexity is meaningless.

To formalise complexity then, assume as given a classifier system that can categorise descriptions into equivalence classes. Clearly, humans are very good at this—they’re able to recognise patterns even in almost completely random data. Rorschach plots are random ink plots that are interpreted by viewers as a variety of meaningful images. However, a human classifier system is not the only possibility. Another is the classification of programs executed by a computer by what output they produce. Tech-

nically, in these discussions, researchers use a *Universal Turing Machine* (UTM), an abstract model of a computer.

Consider then the set of possible binary strings, which can be fed into a UTM U as a program. Some of these programs cause U to produce some output then halt. Others will continue executing forever. In principle, it is impossible to determine generally if a program will halt or continue on indefinitely. This is the so called *halting problem*. Now consider a program p that causes the UTM to output a specific string s and then halt. Since the UTM halts after a certain number of instructions executed (denoted $\ell(p)$) the same result is produced by feeding in any string starting with the same $\ell(p)$ bits. If the strings have equal chance of being chosen (*uniform measure*), then the proportion of strings starting with the same initial $\ell(p)$ bits is $2^{-\ell(p)}$. This leads to the *universal prior* distribution over descriptions s , also known as the Solomonoff-Levin distribution:

$$P(s) = \sum_{\{p:U(p)=s\}} 2^{-\ell(p)} \quad (7)$$

The complexity (or information content) of the description is given by Equation 5, or simply the logarithm of (7). In the case of an arbitrary classifier system, the complexity is given by the negative logarithm of the equivalence class size:

$$C(x) = \lim_{s \rightarrow \infty} s \log_2 N - \log_2 \omega(s, x) \quad (8)$$

where N is the size of the alphabet used to encode the description and $\omega(s, x)$ is the number of equivalent descriptions having meaning x of size s or less (Standish, 2001).

It turns out that the probability $P(s)$ in Equation 7 is dominated by the shortest program (Li & Vitányi, 1997; Thm 4.3.3), namely:

$$K(s) + \log_2 P(s) \leq C \quad (9)$$

($\log_2 P(s) < 0$ naturally)

where C is a constant independent of the description s . $K(s)$ is the length of the shortest program p that causes U to output s , and is called the *Kolmogorov complexity* or *algorithmic complexity*.

An interesting difference between algorithmic complexity, and the general complexity based on human observers can be seen by considering the case of random strings. *Random*, as used in algorithmic information theory, means that no shorter algorithm can be found to produce a string than simply saying “print ...”, where the ... is a literal representation of the string. The algorithmic complexity of a random string is high, at least as high as the length of the string itself. However, a human observer simply sees a random string as a jumble of letters, much the same as any other random string. In this latter case, the equivalence class of random strings is very large, close to N^5 , so the perceived complexity is small. Thus the human classifier defines an example of what Gell-Mann calls *effective complexity* (Gell-Mann, 1994), namely a complexity that has a high value for descriptions that are partially compressible by complex schema, but low for random or obviously regular systems.

A good introduction to information theoretical concepts for complex systems studies can be found in Adami (1998).

Information Theoretic Graph Complexity

In an attempt to bridge the information theoretic approach to complexity with graph theoretical approaches, Standish recently introduced a coding scheme and equivalence class for which practical (though still NP-hard) algorithms exist (Standish, 2012). The intention is to use this method with networks that have a meaning or function attached, such as metabolic networks, or food webs. A randomly constructed food web will collapse fairly quickly under ecosystem dynamics to a much smaller stable food web, so random or regular networks will tend to have a lower complexity value.

Graphs (and digraphs) are encoded with the numbers of nodes and links in binary, with a leading string of 1 bits followed by a 0 stop bit to indicate the width of the node binary field. Finally, a bit vector describing the presence or absence of an edge is appended. Graphs that are automorphisms of each other (nodes and edges can be placed in 1:1 correspondence with each other) are considered to be members of the same equivalence class. Determining the size of the equivalence class (ω) turns out to be equivalent to determining the size of the automorphism group, for which several practical algorithms are available. This measure is minimal for empty and full networks, and peaks for intermediate values of connectivity.

It turns out that this specific complexity measure is very closely related to the concept of *graph entropy*, introduced by Rashevsky (1955), and Trucco (1956) and described in a sequence of articles by Mowshowitz (1968a), Mowshowitz (1968b), Mowshowitz (1968c), and Mowshowitz (1968d). A recent, detailed review can be found in Dehmer and Mowshowitz (2011).

Given a graph $G = V \times E$ of nodes V and links E , and *graph invariant* function α defined on the nodes:

$$\alpha : V \rightarrow A$$

we can form the graph entropy measure:

$$S(G, \alpha) = |V| \log |V| - \sum_{a \in A} |\alpha^{-1}(a)| \log |\alpha^{-1}(a)| \quad (10)$$

where $|\cdot|$ is the usual notation for set cardinality. The sum in Equation 10 is over sets of nodes that are equivalent under the map α . This plays the analogous role to the previously mentioned classifier system. In Standish's (2012) measure, the automorphism relation between graphs as the observer function is used—the corresponding α function maps nodes to their *orbits*. Other graph invariants have also been used in

the literature, such as node degree, or level in a tree structure.

A very similar measure to (10) is obtained by averaging the information contained in each orbit:

$$I(G, \alpha) = - \sum_{a \in A} P_i \log P_i = - \sum_{a \in A} \frac{|\alpha^{-1}(a)|}{|V|} \log \frac{|\alpha^{-1}(a)|}{|V|} \quad (11)$$

One can likewise form similar measures by considering graph invariants over links, rather than nodes.

MOTIF ENTROPY

Adami et al. (2011) introduced the concept of *motif entropy*. By breaking the network into motifs (eg pairs of nodes connected by a link, triangles, quads, 3-pointed stars, etc), and forming the Shannon entropy $H = - \sum p_i \log_2 p_i$, where p_i are the probabilities of the various motifs occurring, one gets a measure which they call motif entropy. This should converge to (8) as more motifs are included, in just the same way as (6) converges to (5) as longer sequences are included. Adami et al. restrict themselves to motifs of two and three nodes only in examining the neural network of *C. elegans*, and also in examining the epistatic interaction networks in the Avida digital organism system. They show that this suffices to capture meaningful adaptive information about the systems, but not whether it captures all the pertinent information. Further work linking motif entropy with network complexity is called for.

EXTENSION TO PROCESSES ON NETWORKS

What has been discussed up until now is the static, or *structural* complexity of a network. Often, a dynamic process occurs on a network,

such as neural network dynamics, or the ecological dynamics of a foodweb. One is interested in the amount of complexity contributed to the process by the network structure. Since two distinct networks with the same attached dynamics may well be considered identical in some context (perhaps by having the same attractor basins, for instance), then in general the *dynamic complexity* is less than the *structural complexity*.

With continuous processes, there is a practical difficulty of establishing whether two networks generate the same process, particularly if there is an element of stochasticity involved. When comparing continuous-valued time series, one would need to choose a metric over (in general) a multidimensional space, and an error threshold within which two time series are considered the same. Furthermore, a maximum time period for comparison needs to be chosen, as dynamical chaos effects are likely to render two arbitrarily close trajectories significantly different after a finite period of time.

If the initial transients of the processes aren't important, one could compare basins of attraction instead, which only eliminates the choice of time period in the comparison.

With discrete (or symbolic) processes, the problem is conceptually simpler in that one can determine if two networks generate identical processes according to an observer function $O(x)$.

Nevertheless, the computational complexity of this approach rules it out for all but the simplest of networks.

An alternative approach is given by considering the amount of information flowing between nodes, a notion known as *transfer entropy*.

TRANSFER ENTROPY

Given a time series X_t , let $X_t^- = \{X_t, X_{t-1}, X_{t-2}, \dots\}$ be the *history* of X up to time t .

The mutual information $I(X_{t+1} : Y_t^-)$ gives a measure of the extent to which the his-

tory of Y disambiguates the future of X . However, Y may itself depend on the past of X , giving rise to spurious directional effects between Y and X (Kaiser & Schreiber, 2002). So we should also condition on the past of X , giving rise to the notion of *transfer entropy*:

$$\tau_{Y \rightarrow X} = \left(X_{t+1} : Y_t^- \mid X_t^- \right) \tag{12}$$

Transfer entropy has been applied to random boolean networks (Lizier et al., 2011), but is more usually used to infer network structure from time series data such as neural networks (Wibral et al., 2011) or genetic regulatory networks (Tung et al., 2007).

We may also condition the transfer entropy on the state of the rest of system U (not including X or Y):

$$\tau_{Y \rightarrow X|U} = I\left(X_{t+1} : Y_t^- \mid X_t^-, U_t^-\right) \tag{13}$$

Lizier et al (2011) call (12) the *apparent* transfer entropy, and (13) the *complete* transfer entropy. They find that apparent transfer entropy is maximised around a critical point corresponding to a connectivity of around 2 links per node, whereas complete transfer entropy rises near the critical point, and continues to rise as connectivity increases and the system moves into the chaotic regime, up to a connectivity of 5 links per node.

A related concept to transfer entropy is *Granger causality*. Granger causality between two nodes X and Y is found by considering a linear multivariate model of the lags:

$$X_{t+1} = \sum_k A_k X_{t-k} + \sum_k B_k Y_{t-k} + \epsilon$$

and a restricted linear model with the Y terms removed:

$$X_{t+1} = \sum_k C_k X_{t-k} + \epsilon'$$

If the former full model gives a statistically significant better fit to the data than the latter restricted model, we say that *Y* Granger causes *X*. To quantify the statistical significance, we use the F-statistic:

$$F_{Y \rightarrow X} = \ln \frac{\langle \epsilon^{12} \rangle}{\langle \epsilon^2 \rangle} \quad (14)$$

as the variance of residuals in the restricted model ($\langle \epsilon^{12} \rangle$) will be more than that of the full model.

Granger causality has the advantages of being computationally simpler, as well as having an interpretation in terms of statistical significance. The downside is that it captures linear relationships only, whereas transfer entropy is model-free, capturing all that is relevant between entities. The two concepts are very closely related, and for the special case of Gaussian processes, are identical up to a factor of 2 (Barnett et al., 2009).

The models used in Granger causality may also include the remainder of the system **U**, and this is used for computing the *causal density* of the system, which is the proportion of pairs of nodes where one node Granger causes the other (Seth et al., 2011). The measure has a minimum for weakly interacting nodes, and likewise for strongly interacting nodes (as everything influences everything else, so is conditioned out). It has a maximum in between, expressing a balance between integration and segregation in a system. It is very similar to an earlier measure proposed by Tononi, Sporns and Edelman (TSE complexity) (Tononi et al., 1994) which is based on mutual information across bipartitions of the network rather than transfer entropy.

Like Standish's network complexity, both TSE complexity and causal density are minimal for sparse and dense networks, rising to a maximum value in between. However the maximum value of causal density occurs around the order-chaos transition (approx 2

links per node), which is a distinctly different peak to that of structural complexity, which is at a maximum at $n/2$ links per node.

MEDIUM ARTICULATION

Wilhelm (Wilhelm & Hollunder, 2007; Kim & Wilhelm, 2008) introduced a new complexity like measure that addresses the intuition that complexity should be minimal for the empty and full networks, and peak for intermediate values. It is obtained by multiplying the mutual information between all pairs of nodes by the conditional entropy across all links (which they call the *redundancy*). The resulting measure also has a quality of measuring the segregation/integration balance reminiscent of causal density.

Precisely, medium articulation is given by:

$$MA = -\sum_{ij} w_{ij} \log \frac{w_{ij}}{\sum_k w_{ik} \sum_k w_{kj}} \times \sum_{ij} w_{ij} \log \frac{w_{ij}^2}{\sum_k w_{ik} \sum_k w_{kj}} \quad (15)$$

where w_{ij} is the normalised weight ($\sum w_{ij} = 1$) of the link from node *i* to node *j*. This can be compared with the product of the two terms *A* and Φ representing the degree of constraint and the extent of freedom of the system in Ulanowicz (2012).

There is no clear relationship between medium articulation and complexity for the average network (Standish, 2012). Medium articulation does not appear to discriminate between complex networks. However if we restrict our attention to simple networks (Figure 3) medium articulation is strongly correlated with complexity, and so could be used as a proxy for complexity for these cases.

This lends some credence to the notion that causal density, TSE complexity and network complexity are all related.

COMPUTATIONAL COMPLEXITY AND LOGICAL DEPTH

Algorithmic complexity takes no account of the time required to execute the shortest algorithm. An almanac of tide charts and Newton's equations of motion plus law of gravity for the Earth-Moon-Sun system contain the same information, yet the almanac contains the information in a more useful form for the maritime sailor, as it requires less work to determine when the tides occur. *Computational complexity* and *logical depth* are two concepts designed to address the issue of a description's value.

Computational Complexity is defined as the execution time of the algorithm. Since this is highly dependent on what operations are available to the processor, usually only the scaling class of the algorithm is considered as the input size is increased. Algorithms may scale *polynomially*, which means the execution time increases as some power of the problem size ($t \propto n^s$), or may scale faster than this (eg exponentially: $t \propto s^n$), in which case they have *nonpolynomial complexity*. The class of polynomial algorithms is called *P*, and the class of nonpolynomial algorithms for which the solution can be checked in polynomial time is called *NP*. It is known that all *NP* algorithms can be transformed into any other by means of a polynomially preprocessor, but it is unknown whether or not $P=NP$, ie whether it is possible to transform any nonpolynomial algorithm into a polynomial one. This issue is of great importance, as certain public key encryption schemes depend on nonpolynomial algorithms to ensure the encryption scheme cannot be cracked within a practical amount of time.

Bennett's *logical depth* (Bennett, 1988) is the execution time of the most highly compressed representation of an object, relative to some reference machine. It is meant to be a measure of the value of the object—the almanac of tide tables has a high value of logical depth compared with the equations of motion that generate it.

In terms of the observer-based complexity notions introduced in §4, assume that the observer has a limited amount of computing resources. Perhaps *e* is only prepared to spend 5 minutes computing the information, and prior to the widespread availability of electronic computers this meant that the almanac was not equivalent to the equations of motion, since it requires more than 5 minutes to compute the tidal information from the equations of motion via manual paper and pencil techniques. Since the almanac is inequivalent to the equations of motion *in this context*, it is clear that the almanac has greater complexity *in this context*.

This reciprocal relationship between logical depth and complexity led to Baez and Stay's (2012) interesting proposal for an *algorithmic thermodynamics*, which draws an analogy between:

- *E*, expected value of the logarithm of a program's runtime (like logical depth);
- *V*, expected value of a program's length (KCS complexity);
- *N*, expected value of a program's output and derive the algorithmic analogue of the basic thermodynamic relation:

$$dE = TdS - PdV + \mu dN \quad (16)$$

where *T, S, P* and μ are defined from the algorithmic analogue of the standard statistical mechanics partition function:

$$p = \frac{1}{Z} e^{-\beta E - \gamma V - \delta N} \quad (17)$$

OCCAM'S RAZOR

The practice of preferring a simpler theory over a more complex one when both fit the observed evidence is called *Occam's Razor*, after William de Occam: Entities should not be multiplied unnecessarily.

What is not widely appreciated, is that this strategy is remarkably successful at picking better theories. Often, when tested against further empirical evidence, the simpler theory prevails over the more complex. A classical example of this sort of thing is Einstein's General Theory of Relativity. The key field equations of General Relativity are really quite simple:

$$G = 8\pi kT \quad (18)$$

Of course unraveling what these equations means for a specific instance involves a nontrivial amount of 4 dimensional tensor calculus, so the General Relativity computations have high logical depth. Einstein proposed the equations in this form because they seemed the most "beautiful"—there were a large number of alternative formulations that fitted the data at the time. One by one, these alternative formulations were tested empirically as technology developed through the 20th century, and found wanting.

However, by what criteria is a particular theory more simple than another. Goodman (1972) developed a theory of *simplicity* to put the practice of Occam's Razor on a more rigorous footing. His idea was to formalise the theories into formal logic predicates, and then count the number of primitive clauses required to encode the theory.

Solomonoff (1964) developed the concept of *algorithmic information complexity* in the 1960s as a way of explaining why Occam's razor works. He considered the set of all possible descriptions and computed the probability distribution that a particular description would be generated by a program picked at random by the reference machine. His work had some technical problems that were solved by Levin (1974), which led to the *universal prior* distribution (7). Basically, simple descriptions have a much higher probability than more complex ones, thus Occam's razor. The same behaviour is true of the arbitrary classifier system at Equation 8 (Standish, 2004).

While the world is expected to be remarkably simple by the above arguments, it is also logically very deep (10^{10} years of runtime so far!). This appears to be the result of another poorly understood principle called the *Anthropic Principle*. The Anthropic Principle states that the world must be consistent with our existence as intelligent, reasoning beings (Barrow & Tipler, 1986). So while Occam's razor says we should live in the simplest of universes, the Anthropic Principle says it shouldn't be too simple, as a certain level of complexity is required for intelligent life. The simplest means of generating this level of complexity is by accruing random mutations, and selecting for functional competence, i.e. Darwinian evolution.

COMPLEXITY AS A QUALITY: EMERGENCE

It is often thought that *complex systems* are a separate category of systems to *simple systems*. So what is it that distinguishes a complex system, such as a living organism, or an economy, from a simple system, such as a pair of pliers? This question is related to the notorious question of *What is Life?*, however may have a simpler answer, since not all complex systems are living, or even associated with living systems.

Consider the concept of *emergence* (Holland, 1997; Fromm, 2004). We intuitively recognise emergence as patterns arising out of the interactions of the components in a system, but not implicit in the components themselves. Examples include the formation of hurricanes from pressure gradients in the atmosphere, crashes in stock markets, flocking behaviour of many types of animals and of course, life itself.

Let us consider a couple of simple illustrative examples, that are well known and understood. The first is the *ideal gas*, a model gas made up of large numbers of non-interacting point particles obeying Newton's laws of motion. A *thermodynamic* description of the gas is obtained by averaging:

- **Temperature (T):** Is the average kinetic energy of the particles;
- **Pressure (P):** Is the average force applied to a unit area of the boundary by the particles colliding with it;
- **Density (ρ):** Is the average mass of particles in a unit volume.

The *ideal gas law* is simply a reflection of the underlying laws of motion, averaged over all the particles:

$$P_{\rho} \propto T \quad (19)$$

The thermodynamic state is characterised by the two parameters T and ρ . The so-called *first law of thermodynamics* is simply a statement of conservation of energy and matter, in average form.

An entirely different quantity enters the picture in the form of *entropy*. Consider discretising the underlying phase-space into cubes of size h^N , (N being the number of particles) and then counting the number of such cubes having temperature T and density ρ , $\omega(T, \rho, N)$. The entropy of the system is given by:

$$S(T, \rho, N) = k_B \ln \omega(T, \rho, N) \quad (20)$$

where k_B is a conversion constant that expresses entropy in units of Joules per Kelvin. One can immediately see the connection between complexity (Equation 8) and entropy. Readers familiar with quantum mechanics will recognise h as being an analogue of Planck's constant. However, the ideal gas is not a quantum system, and as $h \rightarrow 0$, entropy diverges! However, it turns out that in the thermodynamic limit ($N \rightarrow \infty$), the average entropy S/N is independent of the size of h .

The *second law of thermodynamics* is a recognition of the fact that the system is more likely to move a state occupying a larger region of phase space, than a smaller region of phase space, namely that $\omega(T, \rho, N)$ must increase in

time. Correspondingly entropy must also increase (or remain constant) over time. This is a probabilistic statement that only becomes exact in the thermodynamic limit. At the syntactic, or specification level of description (ie Newton's laws of motion), the system is perfectly *reversible* (we can recover the system's initial state by merely reversing the velocities of all the particles), yet at the semantic (thermodynamic) level, the system is *irreversible* (entropy can only increase, never decrease).

The property of irreversibility is an *emergent* property of the ideal gas, as it is not *entailed* by the underlying specification. It comes about because of the additional identification of thermodynamic states, namely the set of all micro-states possessing the same temperature and density. This is extra information, which in turn entails the second law.

The second example I'd like to raise (but not analyse in such great depth) is the well known *Game of Life*, introduced by John Conway (1982). This is a *cellular automaton* (Wolfram, 1984), in this case, a 2D grid of cells where each cell can be one of two states. Dynamics on the system is imposed by the rule that the state of a cell depends on the values of its immediate neighbours at the previous time step.

Upon running the Game of Life, one immediately recognises a huge bestiary of emergent objects, such as blocks, blinkers and gliders. Take gliders for example. This is a pattern that moves diagonally through the grid. The human observer recognises this pattern, and can use it to *predict* the behaviour of the system with less effort than simulating the full cellular automaton. It is a *model* of the system. However, the concept of a glider is not *entailed* by the cellular automaton specification, which contains only states and transition rules. It requires the additional identification of a pattern by the observer.

This leads to a general formulation of *emergence* (Standish, 2001). Consider a system specified in a language L_s , which can be called the specification, or *syntactic* layer (see Figure 2). If one accepts the principle of *reduction*, all systems can ultimately be specified the common

Figure 2. Diagram showing the syntactic and semantic spaces. Two different messages, having meanings A and B , can each be coded in many equivalent ways in syntactic space, represented by the sets \mathcal{A} and \mathcal{B} . The information or complexity of the messages is related to the size it occupies in syntactic space by Formula (5).

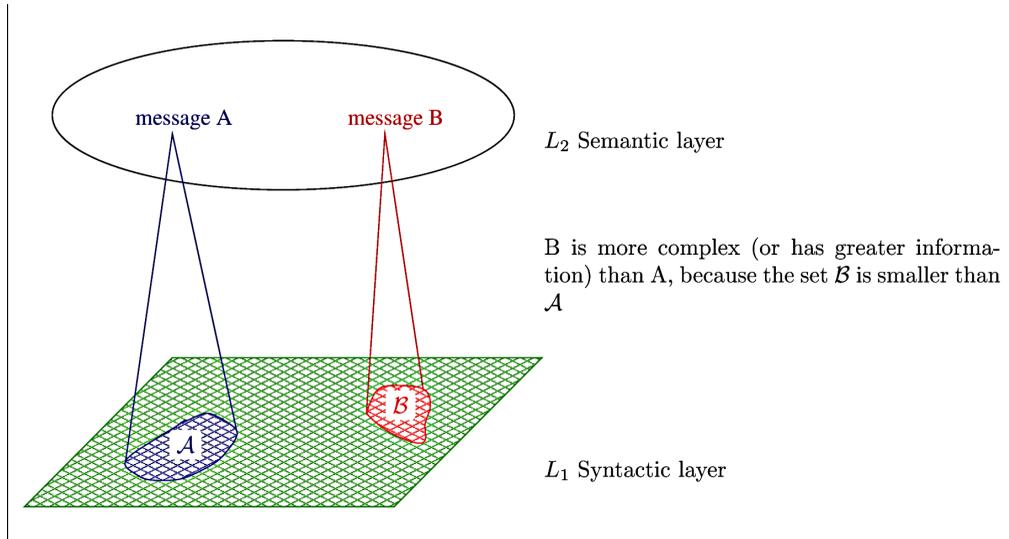
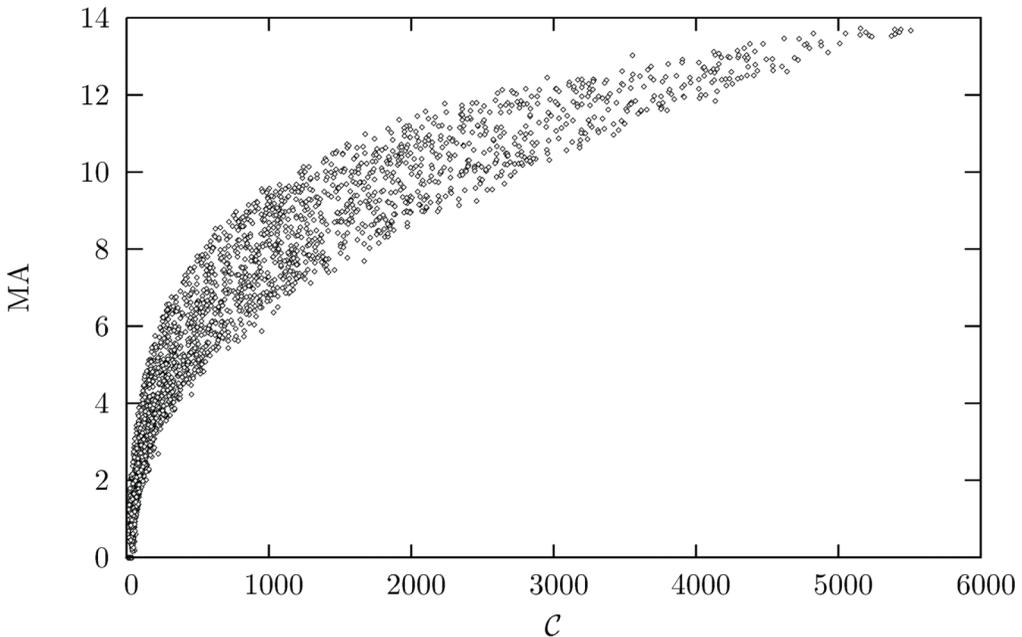


Figure 3. Medium Articulation plotted against complexity for 1000 randomly sampled Erdős-Rényi graphs up to order 500 with no more than $2n$ links



language of the theoretical physics of elementary particles. However, an often believed corollary of reduction is that this specification encodes all there is to know about the system. The above two examples shows this corollary to be manifestly false. Many systems exhibit one or more *good models*, in another language L_2 , which can be called the *semantic layer*. The system's specification does not entail completely the behaviour of the semantic model, since the latter also depends on specific identifications made by the observer. In such a case, we say that properties of the semantic model is emergent with respect to the syntactic specification.

It should be noted that this notion of emergence corresponds to what Bedau (2002) calls *nominal emergence*, which is the weakest and most generic of the emergence concepts he considers. He introduces an important notion of *weak emergence*, which relates to emergent behaviour that cannot be predicted from knowledge of the underlying system, except by full simulation of that underlying system. Finally, he also introduces the concept of *strong emergence*, the philosophically troubling notion that the emergent phenomenon has real causative power over the underlying syntactic layer.

The concept of "good" model deserves further discussion. In our previous two examples, neither the thermodynamic model, nor the glider model can be said to perfectly capture the behaviour of the system. For example, the second law of thermodynamics only holds in the thermodynamic limit—entropy may occasionally decrease in finite sized systems. A model based on gliders cannot predict what happens when two gliders collide. However, in both of these cases, the semantic model is cheap to evaluate, relative to simulating the full system specification. This makes the model "good" or "useful" to the observer. We don't prescribe here exactly how to generate good models here, but simply note that in all cases of recognised emergence, the observer has defined a least one semantic and one syntactic model of the system, and that these models are fundamentally incommensurate. Systems

exhibiting emergence in this precise sense can be called *complex*.

A school of thought founded by Robert Rosen holds that complex systems cannot be described by a single best model as reductionists would assume, but rather has a whole collection of models that in the limit collectively describe the system (Rosen, 1991). That such systems exist, at least formally, is assured by Gödel's incompleteness theorem, (Hofstadter, 1979) which shows that number theory is just such a system that cannot be captured by a finite specification. He further argues mechanical systems (those that have a finite specification such as the examples I have given above) can never be complex, since the specification contains all there is to know about the system. However, he implicitly assumes that all models must be perfect (ie in perfect correspondence with the underlying system), rather than merely good as I do here. This constitutes a *straw man* argument, and leads him to the *false* conclusion that mechanical systems (eg computer simulations) can never exhibit emergence. The two examples presented above, which are perfectly good mechanical systems, are counter-examples to this claim. Furthermore, the definition of complex systems presented here is known to be non-empty, a fact not known of Rosen's definition since no physical counterpart to Gödel's incompleteness theorem is known.

Algorithmic Detection of Emergence

Because emergence involves the recognition of patterns or concepts that by definition are not part of the systems syntactic description, at first glance it seems a hopeless task to automate the process of detecting when emergence exists in a system. Yet it would be a very desirable prize in the search for syntactic rules that are optimal at generating emergent phenomenon, such as living systems.

In recent years, there have been a number of attempts to devise computationally tractable measures to indicate when emergence is occurring.

Gershenson and Fernandez (2012) define an emergence measure as being the ratio $E = I_{out} / I_{in}$ of the information content of the system's outputs (state space traversed by the system) to its inputs (initial conditions and state rules). This must, of course, be relative to some observer classification function, as it is clear that if the observer understands completely the algorithm producing the system, I_{out} is always strictly less than I_{in} , i.e. no emergence is possible with the omniscient observer.

Gershenson and Fernandez also define the concept $S = I_{in} - I_{out}$ to be *self-organisation*, and the product ES to be complexity of the system.

Another noteworthy effort is the use of Granger causality to detect when emergence is occurring in a system (Seth, 2010). Granger causality is based on detection of linear correlations in the system, which might be a deficiency to some, as neural networks (including naturally occurring observers) are nonlinear in general. It is unclear at this point in time whether it is necessary to use the more computationally demanding transfer entropy concept for this purpose, as used by Lizier et al. (2007).

CONCLUSION

When connoting a quality, complexity of a system refers to the presence of emergence in the system, or the exhibition of behaviour not specified in the systems specification. When connoting a quantity, complexity refers to the amount of information needed to specify the system. Both notions are inherently observer or context dependent, which has led to a disparate collection of formalisations for the term, and has led to some despairing of the concept being adequately formalised. This would be a mistake, as within a given application, the meaning can be well defined.

An additional difficulty is the combinatorial size of the underlying syntactic space, which can lead to the intractability of computing complex-

ity. Furthermore, the details of the syntactic layer may be inaccessible, for example the absence of historical genetic data in the study of evolution from the fossil record. So being able to establish easy to measure proxies for complexity is often important, and many proposals for complexity are of this nature.

REFERENCES

- Adami, C. (1998). *Introduction to artificial life*. Springer. doi:10.1007/978-1-4612-1650-6
- Adami, C., Qian, J., Rupp, M., & Hintze, A. (2011). Information content of colored motifs in complex networks. *Artificial Life*, 17, 375–390. doi:10.1162/artl_a_00045 PMID:21762026
- Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 47. doi:10.1103/RevModPhys.74.47
- Baez, J. C. (2012). Algorithmic thermodynamics. *Computability of the Physical. Mathematical Structures in Computer Science*, 22, 771–787. doi:10.1017/S0960129511000521
- Barnett, L., Barrett, A., & Seth, A. (2009). Granger causality and transfer entropy are equivalent for Gaussian variables. *Physical Review Letters*, 103(23), 238701. doi:10.1103/PhysRevLett.103.238701 PMID:20366183
- Barrow, J. D., & Tipler, F. J. (1986). *The anthropic cosmological principle*. Oxford: Clarendon.
- Bedau, M. A. (2002). Downward causation and the autonomy of weak emergence. *Principia*, 6, 5–50.
- Bennett, C. H. (1988). Logical depth and physical complexity. In R. Harkin (Ed.), *The universal turing machine: A halfcentury survey* (Vol. 1, pp. 227–258). Oxford UP.
- Bonner, J. T. (1988). *The evolution of complexity*. Princeton UP.
- Claussen, J. C. (2007). Offdiagonal complexity: A computationally quick complexity measure for graphs and networks. *Physica A*, 375, 365–373. arXiv:q-bio.MN/0410024.
- Conway, J. H. (1982). What is life? In E. Berlekamp, J. H. Conway, & R. Guy (Eds.), *Winning ways for your mathematical plays* (Vol. 2). Academic.

- Dehmer, M., & Mowshowitz, A. (2011). A history of graph entropy measures. *Information Sciences, 181*(1), 57–78. doi:10.1016/j.ins.2010.08.041
- Diestel, R. (2005). *Graph theory* (3rd ed.). Berlin, Germany: Springer.
- Edmonds, B. (1999). *Syntactic measures of complexity*. PhD thesis, University of Manchester. Retrieved from <http://bruce.edmonds.name/thesis/>
- Erdős, P., & Rényi, A. (1959). On random graphs. *Publ. Math. Debrecen, 6*, 290–291.
- Fromm, J. (2004). *The emergence of complexity*. Kassel UP.
- Gell-Mann, M. (1994). The quark and the jaguar: Adventures in the simple and the complex. *The Freeman*.
- Gershenson, C., & Fernandez, N. (2012). Complexity and information: Measuring emergence, self-organization, and homeostasis at multiple scales. *Complexity, 18*, 29–44. arXiv1205.2026 doi:10.1002/cplx.21424
- Gliviak, F., & Kyš, P. (1995). Note on the relation between radius and diameter. *Mathematica Bohemica, 120*, 203–207.
- Goodman, N. (1972). *Problems and projects*. Indianapolis, IN: Bobbs-Merrill.
- Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An eternal golden braid*. New York, NY: Basic Books.
- Holland, J. (1997). *Emergence: From chaos to order*. Addison Wesley.
- Kaiser, A., & Schreiber, T. (2002). Information transfer in continuous processes. *Physica D. Non-linear Phenomena, 166*, 43–62. doi:10.1016/S0167-2789(02)00432-3
- Kim, J., & Wilhelm, T. (2008). What is a complex graph. *Physica A, 387*, 2637–2652. doi:10.1016/j.physa.2008.01.015
- Levin, L. A. (1974). Laws of information conservation (non-growth) and aspects of the foundation of probability theory. *Problems of Information Transmission, 10*, 206–210.
- Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications*. Springer. doi:10.1007/978-1-4757-2606-0
- Lizier, J., Pritam, S., & Prokopenko, M. (2011). Information dynamics in small-world Boolean networks. *Artificial Life, 17*, 293–314. doi:10.1162/artl_a_00040 PMID:21762020
- Lizier, J. T., Prokopenko, M., & Zomaya, A. (2007). Detecting non-trivial computation in complex dynamics. In Costa et al. (Eds.), *Advances in artificial life: Proceedings of ECAL2007* (Vol. 4648 of LNCS, pp. 895–904). Springer.
- Marcot, J. D., & McShea, D. W. (2007). Increasing hierarchical complexity throughout the history of life: Phylogenetic tests of trend mechanisms. *Paleobiology, 33*, 182–200. doi:10.1666/06028.1
- McShea, D. W. (1996). Metazoan complexity and evolution: Is there a trend? *Evolution; International Journal of Organic Evolution, 50*, 477–492. doi:10.2307/2410824
- Mowshowitz, A. (1968a). Entropy and the complexity of graphs: I. an index of the relative complexity of a graph. *Bulletin of Mathematical Biology, 30*(1), 175–204. PMID:5666816
- Mowshowitz, A. (1968b). Entropy and the complexity of graphs: II. The information content of digraphs and infinite graphs. *Bulletin of Mathematical Biology, 30*(2), 225–240. PMID:5674895
- Mowshowitz, A. (1968c). Entropy and the complexity of graphs: III. Graphs with prescribed information content. *Bulletin of Mathematical Biology, 30*(3), 387–414.
- Mowshowitz, A. (1968d). Entropy and the complexity of graphs: IV. Entropy measures and graphical structure. *Bulletin of Mathematical Biology, 30*(4), 533–546.
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review, 45*(2), 167–256. arXivcond-mat/0303516 doi:10.1137/S003614450342480
- Prokopenko, M., Boschetti, F., & Ryan, A. J. (2009). An information-theoretic primer on complexity, self-organization, and emergence. *Complexity, 15*, 11–28. doi:10.1002/cplx.20249
- Rashevsky, N. (1955). Life, information theory, and topology. *Bulletin of Mathematical Biology, 17*(3), 229–235.
- Rosen, R. (1991). *Life itself*. Columbia UP.

- Seth, A. K. (2010). Measuring autonomy and emergence via Granger causality. *Artificial Life*, 16, 179–196. doi:10.1162/artl.2010.16.2.16204 PMID:20067405
- Seth, A. K., Barrett, A. B., & Barnett, L. (2011). Causal density and integrated information as measures of conscious level. *Phil. Trans. R. Soc. A*, 369, 3748–3767. doi:10.1098/rsta.2011.0079 PMID:21893526
- Shannon, C. E. (1949). *The mathematical theory of communication*. Urbana-Champaign, IL: U. Illinois Press.
- Solomonoff, R. J. (1964). A formal theory of inductive inference, part 1 and 2. *Inform. Contr.*, 1–22, 224–254.
- Standish, R. K. (2001). On complexity and emergence. *Complexity International*, 9. arXiv:nlin.AO/0101006.
- Standish, R. K. (2004). Why Occam's razor? *Foundations of Physics Letters*, 17, 255–266. arXivphysics/0001020 doi:10.1023/B:FOPL.0000032475.18334.0e
- Standish, R. K. (2008). Concept and definition of complexity. In A. Yang, & Y. Shan (Eds.), *Intelligent complex adaptive systems* (pp. 105–124). IGI Global. doi:10.4018/978-1-59904-717-1.ch004
- Standish, R. K. (2012). Complexity of networks (reprise). *Complexity*, 17, 50–61. arXiv0911.3482 doi:10.1002/cplx.20393
- Tononi, G., Sporns, O., & Edelman, G. M. (1994). A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 5033–5037. doi:10.1073/pnas.91.11.5033 PMID:8197179
- Trucco, E. (1956). A note on the information content of graphs. *Bulletin of Mathematical Biology*, 18(2), 129–135.
- Tung, T. Q., Ryu, T., Lee, K., & Lee, D. (2007). Inferring gene regulatory networks from microarray time series data using transfer entropy. In *Proceedings of the Twentieth IEEE International Symposium on Computer-Based Medical Systems* (pp. 383–388).
- Ulanowicz, R. (2013). Circumscribed complexity in ecological networks. In A. Mowshowitz, M. Dehmer, & F. Emmert-Streib (Eds.), *Advances in network complexity* (pp. 249–258). John Wiley & Sons. doi:10.1002/9783527670468.ch11
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393(6684), 409–410. doi:10.1038/30918 PMID:9623993
- Wibral, M., Rahm, B., Rieder, M., Lindner, M., Vicente, R., & Kaiser, J. (2011). Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks. *Progress in Biophysics and Molecular Biology*, 105, 80–97. doi:10.1016/j.pbiomolbio.2010.11.006 PMID:21115029
- Wilhelm, T., & Hollunder, J. (2007). Information theoretic description of networks. *Physica A*, 385, 385–396. doi:10.1016/j.physa.2007.06.029
- Wolfram, S. (1984). Cellular automata as models of complexity. *Nature*, 311, 419–424. doi:10.1038/311419a0

APPENDIX

Glossary

Algorithmic Complexity: Length of shortest program capable of generating the description (also known as Kolmogorov complexity).

Anthropic Principle: The statement that the properties of the universe must be such as to permit our existence as human observers.

Cellular Automaton: A grid of cells, each of which can be in a finite number of states. The states of each cell depend only on the states of a neighbourhood of cells at the previous timestep, known as a transition rule.

Church-Turing Thesis: The proposition that all computable functions can be computed on a Turing machine.

Classifier system: A system that classifies a set of objects into a discrete set of classes. Formally equivalent to a map into the set of whole numbers.

Complexity (quality): The quality of possessing emergent properties.

Complexity (quantity): The amount of information a particular system represents to the observer.

Computational Complexity: The computational cost of executing an algorithm.

Emergence: The phenomenon of emergent properties. Also known as *nominal emergence*.

Emergent Properties: Properties of a system at the semantic level that are not entailed at the syntactic level.

Entail: To logically imply something.

Entropy: Logarithm of number of syntactic states equivalent to a given semantic state. It is related to information (I) by $S + I = S_{max}$ where S_{max} is the maximum possible entropy of the system.

Equivalence Class: A set of objects that are equivalent under some mapping i.e.

$$\{x : e(x) = e(x')\} \exists x'$$

Game of Life: A well known cellular automaton, with two states per cell, and a particular transition rule.

Granger Causality: A statistical measure of how much some process influences another process.

Graph Theory: Mathematical theory of objects consisting of atomic nodes linked by connections called edges.

Gödel Incompleteness Theorem: No finite set of axioms can prove all possible true theorems of number theory.

Information: The amount of meaning in any description; formally given as the logarithm of the proportion of syntactic space occupied by the description.

Link degree: The number of edges a node has in a graph.

Logical Depth: Execution time of the most compressed representation of an object.

Mutual Information: The amount of information held in common between two systems.

Newton's Laws of motion: Laws of ideal point particles: the acceleration a particle experiences is proportional to the force acting on it, which is a function of the positions and velocities of the particle and the environment.

Nominal Emergence: See *emergence*.

Occam's Razor: A statement of the practice of preferring a simpler theory over a more complex one.

Scale free: A scale free distribution has infinite mean. A power law distribution is a common scale free distribution. A scale free process is a stochastic process obeying a scale free distribution.

Second Law of Thermodynamics: Entropy can only increase, or remain constant in a closed system; it can never decrease.

Semantic Level (or space): The space of meanings for any description.

Strong Emergence: Emergent phenomena that have causal power over the syntactic level.

Syntactic Level (or space): The language in which a description is specified: letters of the alphabet, genetic code, laws of theoretical physics, as appropriate.

Transfer Entropy: A measure of the amount of influence one process has over another. Different from Granger Causality, but provably equivalent in some limits.

Turing Machine: A formal model of a computation.

Universal Turing Machine: A formal model of a computer. Is capable of simulating any Turing Machine with appropriate input.

Weak Emergence: Emergent phenomena that cannot be predicted with full simulation of the system.