

Network Measures of Ecosystem Complexity

Alan Dorin and Kevin B. Korb

Faculty of Information Technology,
Monash University, Clayton, Australia
[alan.dorin, kevin.korb] @ infotech.monash.edu.au

Abstract

We argue that the networks that can be constructed to represent ecosystems may inform us about the open-endedness of the evolutionary systems that underlie their dynamics. By adopting this approach we circumvent problems that arise from looking for open-endedness at the level of the organism, the more usual approach. We then examine various measures of ecosystem (niche web) complexity and propose a new information-theoretic approach, *Shannon Web Complexity*. We compare its behaviour to that of the more common measures in ecology, in the light of common intuitions about complexity over a set of test networks and real ecosystem trophic webs. We show that our measure better accommodates intuitions about the complexity of these networks.

Introduction

The search for open-ended evolutionary simulations is compelling and has driven a sub-community of Artificial Life researchers to join philosophers and theoretical biologists in pondering the manner in which biological evolution is open-ended. This has resulted in various simulation environments that attempt to replicate the behaviour of real ecosystems (e.g. see the review in (Dorin, Korb et al. 2008)) and open problems such as the call to, "Create a formal framework for synthesizing dynamical hierarchies at all scales" (Bedau, McCaskill et al. 2000).

To achieve the goal of open-ended evolutionary software, we must first unambiguously identify open-ended complexity increase when we see it – we require a measure. Typically, as we show below, the search has focused on the increasing complexity of *organisms*, their structure and behaviour. For reasons we outline, we believe this to be wrong-headed and the source of much confusion. Instead, we propose to measure the complexity of the ecosystems of which organisms are a part, and to show that these *do* increase in complexity over evolutionary time periods. We achieve this by looking at ecosystem networks.

Ecosystem networks

Biological evolution operates within ecosystems on changing populations that define for themselves new ways of accumulating and consuming energy and matter to be employed for reproduction. Through feedback loops, organisms construct their own niches, passively and actively

organising their environment, modifying the selection pressures acting on themselves, their progeny, and their cohabiters (Odling-Smee, Laland et al. 2003). The moulding of self-selection pressures by a population shifts the constraints within which future generations are introduced. Ecosystems can be described by a variety of networks linking these biotic and abiotic physical, chemical and behavioural relationships. We, like many ecologists, focus our attention on such networks as a way of understanding the global properties of the systems they represent (Watts and Strogatz 1998; Barrat and Weigt 2000; Dunne, Williams et al. 2002; Proulx, Promislow et al. 2005; Blüthgen, Fründ et al. 2008).

We examine several techniques employed in the ecological and other literature for measuring the properties of ecosystem food webs and networks, describing also the *Shannon web complexity* based on information theory (Boulton and Wallace 1969). We then assess how these measures stack up against one another and against our intuitions about the complexity of ecosystem networks in a set of examples.

Open-Ended Complexity Increase

A common opinion about evolution has been that it swims against the tide of entropy and in particular that evolution over time constructs more and more complex organisms (e.g., see (Bronowski 1970)). This idea of creative complexity increase equates at its most extreme, to the view that evolution is progressing from bacteria to invertebrates and thence to vertebrates and mammals and, finally, to the pinnacle of life forms, us.¹ Such a view of Progress, however, ignores some quite basic features of evolution. For example, that the bacteria being "progressed from" still exist today and, indeed, have exactly as long an evolutionary history as we do, since we all have common ancestry. So, progress can hardly be characterized by endurance. Instead, progress has been recast as complexity, and complexity itself has been cast in terms favorable to ourselves; for example, as owning complex neural organizations — an account that fails to address the vast majority of earth's life (Maynard-Smith and Szathmáry

¹ For a skeptical review of this consensus opinion, identifying culprits, see McShea, D. W. (1991). "Complexity and evolution: what everybody knows." *Biology and Philosophy* 6(3): 303-324.

1995). An infatuation with ourselves is also behind the “C-value paradox” — our chromosomes appear no more complex than those of other mammals and less complex than those of some plants. The two long-standing antagonists Dawkins and Gould have together, and quite rightly, castigated this view as human chauvinism in an exchange in *Evolution* (Dawkins, 1997; Gould, 1997). Gould preferred to see in every attempt to characterize complexity and attribute its increase to evolutionary processes this hidden agenda of congratulating ourselves on our own unique wonderfulness. Dawkins, on the other hand, considers the evolutionary increase in complexity to be not just compatible with evolution, but intrinsic to it. Evolution climbs “Mount Improbable” (Dawkins, 1996). Dawkins’ line of defence for ongoing complexity increase is to suggest that, whereas adaptive processes responding to the abiotic environment may just track meandering changes in the climate, coevolutionary processes acting between species work to develop coadaptations in trajectories that can be regarded as progressive in an engineering sense. Arms races lead to better weaponry and better defences, including better speed, flight, hearing and vision, for example.

In the Artificial Life literature, Bedau takes up the debate, offering his evolutionary activity statistics to assess whether or not an evolutionary system is evolving in an open-ended fashion (Bedau, Snyder et al. 1998) and an *Arrow of Complexity Hypothesis* that evolutionary systems show a systematic tendency to increase the complexity of organisms over time (Bedau 2006). Some Artificial Life researchers, notably (Ray 1990), have attempted to replicate this apparent evolutionary complexity increase in software, thus far, without any consensus of success, although some claim a limited success whilst improving Bedau et al’s measures of open-endedness (Channon 2006). The fundamental problem we have with the activity statistics, however, is that whatever they measure is not what we want to measure: they make no attempt to assess the complexity of organisms or ecosystems, but only the volume of new, adaptive “components” within an evolutionary system.

An attempt to dismiss complexity increases in species’ organisation and behaviour over evolutionary time periods invokes a metaphoric “passive diffusion” (McShea 1994) through species design space, rather than a directed drive towards greater complexity. While diffusion may well contribute to increases in species complexity, it is unlikely to explain it entirely (Korb and Dorin 2010). In any case, we prefer to sidestep the issue and focus on complexity at a higher level: in the organization of niches in the ecosystem. Niche web complexity is not subject to the diffusion effects cited by McShea and others; furthermore, it, and its correlate species biodiversity, relatively non-controversially have shown sustained increases over geological time. Indeed, it is arguable that niche web complexity exhibits an exponential trajectory over evolutionary time periods, which we call the *Arrow of Niche Complexity Hypothesis* (Korb and Dorin 2010): with complexity interpreted simply as the number of niches, this hypothesis states that any ecosystem acting beneath the ceiling of its capacity constraints whilst maintaining its stability will robustly tend to produce new niches, at an exponential growth rate — every species, without

exception, creates multiple new niches by its waste products, its impact as an ecosystem engineer (the existence of its body as habitat, for instance), its availability as food for other organisms, and its removal of resources from the environment changing their relative abundance and distribution. Elsewhere we offer a simulation that demonstrates the effect of an exponentially increasing number of niches (Korb and Dorin 2009). Furthermore, the network of dependence of species (in niches) to other species (in other niches) also increases in complexity. In order to argue that the latter increases are exponential and, in general, to assess changes in niche web complexity, we require a principled way of measuring such complexity.

Complexity Measures for Ecosystems

We require a measure for the complexity of (virtual or real) ecosystems in order to assess whether or not our *Arrow of Niche Complexity* hypothesis holds true under some circumstances. This measure must correspond to our (educated) intuitions about what constitutes the complexity of a network (such as a food web). A few useful intuitions are listed next. We then present some measures of network properties that have been employed in the literature and our own suggestion.

Intuitions about network complexity

Intuition 1 (simple): A network with a regular, repeating structure is simple (e.g. a lattice or a fully-connected network).

Intuition 2 (simple): Networks with few links are simple (e.g. a single long chain or a fully disconnected network).

Intuition 3 (simple): A random network is simple (with a high probability; but since random processes can produce any structure, such a net will sometimes accidentally be complex!).

Intuition 4 (simple): Small world networks — those with low “degrees of separation” — are simple.

Intuition 5 (complex): A complex network has organisation (e.g. clusters, loops) at multiple scales.

Intuition 6 (complex): A complex network has organisation (clusters, loops) of multiple sizes.

Intuition 7 (complex): A bigger network is more complex than a small one.

These intuitions, while widely commented upon in the ecological literature, are not universal; nor are they unambiguous. For one thing, they only make sense with *ceteris paribus* clauses — other things remaining equal. And there are potential interactions between some of them. For example, Intuition 7 may be undermined by increasing the size of the network while simultaneously deleting arcs and bringing in Intuition 2. They work perhaps as heuristic guides to assessing networks and their complexity measures.

Intuitions 5 and 6 are likely to capture some aspects of the major transitions in evolution, which can lead to tightly organized groups of niches.

Intuitions 1 and 2 have an interesting joint consequence which we will make use of later: networks of low density are simple, but so too are networks of high density (fully connected or worse). We infer that there may well be some kind of "Goldilocks effect", i.e., that there is a maximum of network complexity achieved at some middle level of density, which tapers off when there is either too many or too few arcs. The same effect applies to Intuition 4 as well: very small worlds mean very high interconnection, while very large worlds imply very low interconnection.

Measures of network properties

Networks have been widely studied in biology and several measures have been used to inform us about their properties in general (Watts and Strogatz 1998; Dunne, Williams et al. 2002; Proulx, Promislow et al. 2005; Neutel, Heesterbeek et al. 2007; Blüthgen, Fründ et al. 2008). Here we list some of relevance,² gauging the extent to which each informs us about network complexity. We conclude this list by introducing our own proposal.

Number of nodes: n

Ceteris paribus, smaller networks are simpler networks (cf. Intuition 7).

Number of edges: e

Having fewer edges is another way in which networks can be smaller and therefore simpler (Intuition 2).

Density: $D = e / n^2$

Given that there are n^2 potential directed arcs in a network (where a node may have an arc directed back to itself), this is the frequency of arcs (relevant to Intuitions 1, 2, 4, 5 and 6).

Density-Mass: $D \times n = e / n$

This combines Intuitions 2 and 7. Given that denser networks are more complex (other things being held equal *and* up to a point of diminishing returns) and larger networks are more complex, it's reasonable to suppose that a measure of complexity might be proportional to both simultaneously, so we multiply the two measures.

Characteristic path length (CPL):

$$CPL = \sum_{j=1}^n \sum_{i=1}^n \frac{s_{ij}}{p}$$

where s_{ij} is the shortest path between nodes i and j (0 in case the shortest path is infinite) and p is the number of finite shortest paths between two nodes in the network (i.e. $p < n^2$ just in case some shortest paths are infinite). Thus, *CPL* is the average shortest path length ("degree of separation") between nodes. Low values would normally indicate a highly connected network, i.e., high edge density, or perhaps strategically placed edges allowing for shortcuts, corresponding to Intuitions 1 and 4.

² The literature contains many measures and variations. We focus on a few popular unweighted measures. Measures such as the *maximum omnivorous loop weight* (Neutel, Heesterbeek, et al. 2007) are useful in some ecological applications but obviously not applicable to networks with unweighted edges.

Clustering Coefficient (CC):

$$CC = \frac{1}{n} \sum_i \frac{S_i}{N_i}$$

where N_i is the number of i 's neighbors and S_i is the number of shared neighbors, i.e., neighbors which are also neighbors of neighbors. This measures, on average, how "cliquey" the neighbors are across a network.

In a niche web a high clustering coefficient shows the presence of tightly coupled clusters of niche-dependencies. So, this is a partial indicator of the clusters and loops of Intuitions 5 and 6.

Shannon web complexity (SWC).

This is a new use of a prior information-theoretic complexity measure, measuring niche web complexity by the number of bits needed to efficiently encode a network with n nodes, where the web may be any directed graph between the nodes. The code should be Shannon efficient for specifying the network structure to a receiver. In this first version of SWC we make the simplifying assumption that the density of arcs in the network is uniform; i.e., the number of arcs in any two subgraphs of the same size is approximately the same. This assumption admittedly will be untrue for many networks, when the measure will no longer be Shannon efficient; however, SWC can be refined in the future to deal with such networks. As it stands, this measure will still be useful for a very large range of networks.

First, we need to identify (label, number) all the nodes. We can do this simply by specifying how many there are, i.e., coding the number n , assuming the labels will be 1, 2, ..., n .

$$\log_2 n$$

Now we need to specify all arcs. We can do this in two steps. First we encode an estimate p of the probability that an arc exists between any two nodes; call this code length $M(p)$. Given knowledge of p , specifying an existing arc takes $-\log_2 p$ bits and specifying the absence of an arc takes $-\log_2 (1 - p)$ bits. The number of possible arcs (going in either direction between nodes) is n^2 (since nodes may be parents of themselves), so

$$p = e / n^2$$

where e is the number of arcs in the graph (i.e., this is the density measure from above).

Hence, we can identify the arc structure in the following number of bits:

$$e (-\log_2 p) + [n^2 - e] (-\log_2 (1 - p))$$

The first summand is the bit cost of specifying e arcs; the second is the bit cost of specifying all other potential arcs are missing. So, our final measure is:

$$M(p) + \log_2 n + e (-\log_2 p) + [n^2 - e] (-\log_2 (1 - p))$$

This has the reasonable Goldilocks property above: a low density web is counted as simple; complexity increases as the number of arcs increase; but as the web becomes very dense – as, for example, an ecosystem turns into an indiscriminate

mush – it starts losing complexity. Maximal complexity is reached when $p \approx 0.5$. This measure is shown by (Boulton and Wallace 1969) to be effectively the same as the following adaptive code, which is simpler to compute (meaning, e.g., we don't actually have to measure $M(p)$):

$$\log_2 \frac{(n^2 + 1)!}{e!(n^2 - e)!}$$

This measure doesn't respond directly to the Intuitions that loopiness implies complexity (5 and 6), however as the arc density goes from low towards 0.5, loopiness is inevitable. Loopiness is improbable at low arc densities, while in some way meaningless at very high arc densities.

Testing Our Measures

Sample graphs.

Figure 1 shows four test graphs C1... 4 that we have designed with a constant number of nodes but increasing number of edges to highlight the behaviour of the network measures.

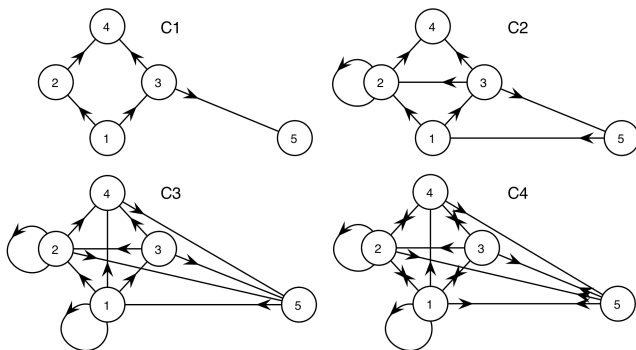


Figure 1. Graphs of five nodes with increasing number of edges.

Figure 2 is a set of networks showing successional stages of a subterranean food web redrawn from (Neutel, Heesterbeek et al. 2007). To the authors of that paper and this alike, these networks appear to be of increasing complexity³. In the following section we present the results of our measurement of the properties of these two sets of graphs.

³ Sch(iermonnikoog) and Hul(shorsterzand) 1 are both successional stage 1 food webs. Hul 2, 2-3 and 4 are subsequent stages of development at the latter site. Nodes represent trophic groups detailed in the original paper.

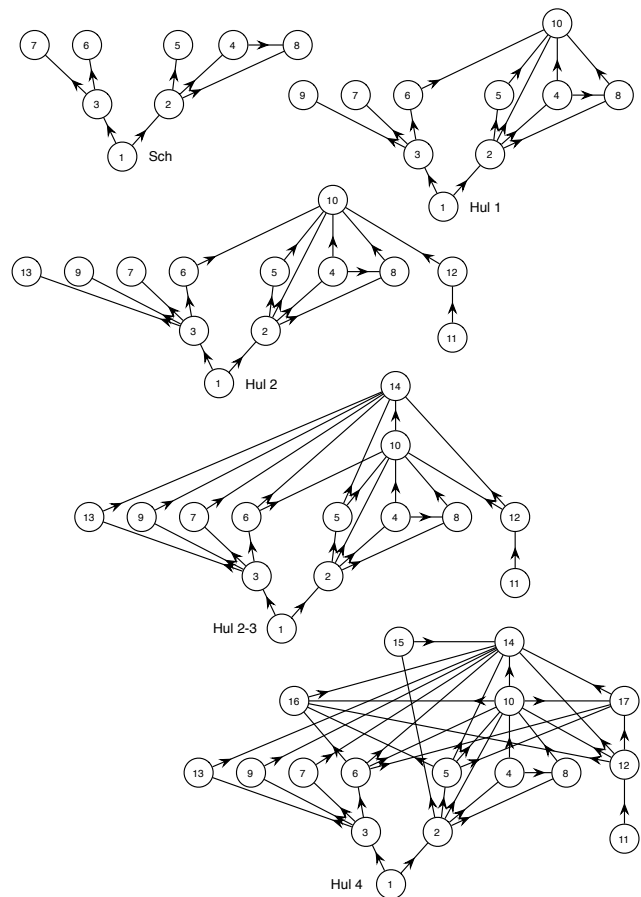


Figure 2. Graphs of subterranean food webs at progressing successional stages (Schiermonnikoog and Hulshorsterzand in the Netherlands).

Results

Figure 3 allows us to read the trends of the measures given above for graphs C1-4. Apart from SWC and CPL, all measures rise with the number of edges in the network. This certainly corresponds with naïve Intuition 2. But this suggests the measures are actually poor indicators of complexity as the sustained increase contradicts Intuitions 1 and 4 that as the network becomes more fully connected, it is becoming more homogeneous, less likely to have long loops and distinct clusters, and therefore *less* complex. In contrast, we see here that SWC and CPL both take the requisite dive after C3 (which has density ≈ 0.5) as the network connectedness climbs "too far".

Figures 4 and 5 show the CPL and SWC respectively, as applied to the webs of figure 2. The CPL drops in the middle stages, before rising once again. As the ratio of the number of arcs to number of nodes increases (i.e., the edge density increases), the chance of having differentiated sub-networks actually decreases – the network will become one large structure with many internally connecting arcs. Depending on how these edges are added, the characteristic path length may drop, as is the case here and as we saw above, in moving from our network C3 to C4. If more nodes are later added in such a

way as to add *lengthy* loops, then the CPL too will rise. The SWC demonstrates a continued increase in complexity across the webs as we, and the ecologists, would wish.

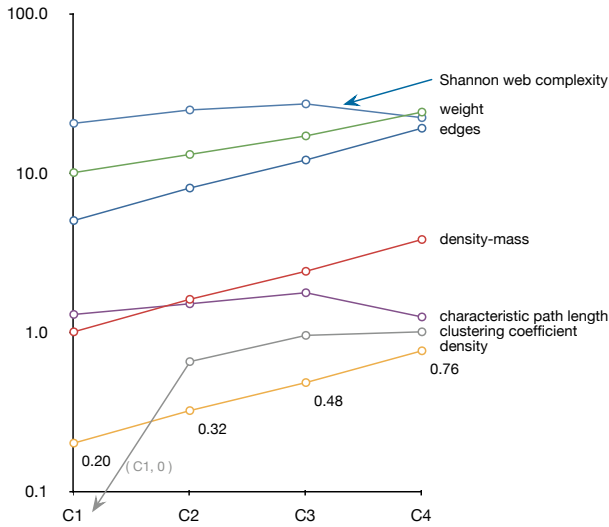


Figure 3. Trends of the various dimensionless measures across test graphs, C1-4 (vertical axis has a log scale).

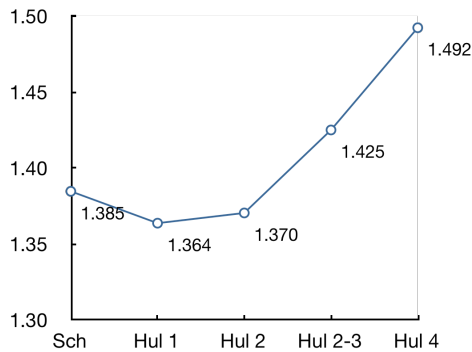


Figure 4. The drop in characteristic path length (CPL) of the food webs from Sch to Hul 1 and 2 is counter to our intuition about the webs' complexity.

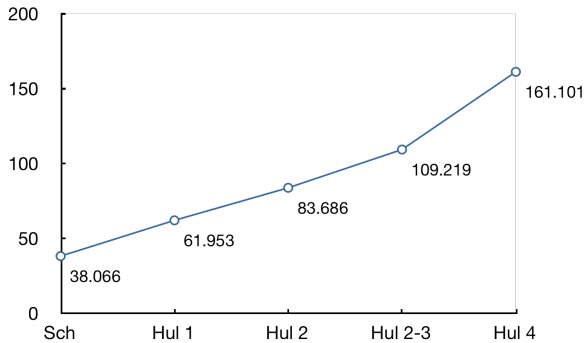


Figure 5. The increase in the Shannon web complexity (SWC) of the food webs matches our intuitions about their complexity.

Discussion and Future Work

Our proposed measure of niche web complexity is an improvement upon Bedau's evolutionary activity plots for identifying open-ended evolution – in particular, it measures the right thing, biological complexity, at at least one of the right levels of organisation, the niche web. Even this first SWC measure appears tricky to subvert; it is at least better than the measures actually employed in the ecological literature. In particular, we have shown that SWC corresponds to basic intuitions regarding complexity and, at least in our test cases, tells us more than its competitors in this regard. There are various options for improvement nevertheless. We can anticipate in the future looking at: the number of iterations required to reduce a non-planar graph to planarity by subtraction of maximal planar subgraphs; the standard deviations of shortest path lengths clustering coefficients across subgraphs; dropping the assumption of uniform arc densities in the SWC measure by compounding the SWCs of subgraphs.

Even before we extend our existing measures, we plan to apply them to the networks generated by various artificial-life ecosystems, especially our own (Korb and Dorin 2009) and those measured by others using their own statistics (e.g. (Ray 1990; Bedau, Snyder et al. 1998; Channon and Damper 2000)) to see what they may tell us about the simulations' open-endedness. Should they prove to support open-endedness, one significant hurdle must still be overcome — accommodating the “major transitions” of evolution (Maynard-Smith and Szathmáry 1995) that play a key role in the open-endedness of real evolution. Can these be replicated in simulation? Would our measures detect them if they did occur?

The major transitions such as the evolution of eukaryotes and the development of sexual reproduction, relate in part to changes in how information is passed between generations. Niche webs do not explicitly model such behaviour; however, another prominent feature of many of these transitions is the incorporation of one entity in the life cycle of another (e.g., bacteria in digestion or the development of mitochondria) or, again, the differentiation of subparts into specialising modules (e.g., new organs and tissues). These kinds of transitions have impacts on niche webs, either explicitly or implicitly, and will often show up in the ways in which subgraphs of niches are interrelated. So, while there are limitations to what examining niche webs can reveal about major transitions, there are also potential impacts of the transitions on niche webs that should not go unexamined.

Conclusions

Niche web complexity is a promising focus for understanding biological complexity growth and so for assessing also the complexity of Artificial Life simulations. While there is a long tradition in ecology of considering this kind of complexity, most of the literature uncritically adopts one or another measure on the basis of intuitive arguments. We have codified these intuitions, formalized a variety of measures corresponding to them, as well as an information-theoretic measure, and tested them using a range of networks. We think

the information-theoretic measure has considerable promise for assisting us in understanding biological complexity growth and, therefore, open-ended evolution.

Acknowledgements

The authors wish to thank Lloyd Allison for spotting bugs in an early version of our SWC and Mark Bedau and Tim Taylor for discussions on the issues addressed by this paper.

References

- Barrat, A. and M. Weigt (2000). "On the properties of small-world network models." *European Physical Journal B* 13: 547-560.
- Bedau, M. (2006). The evolution of complexity. *Symposium on the Making up of Organisms*, Ecole Normale Supérieure.
- Bedau, M. A., J. S. McCaskill, et al. (2000). "Open Problems in Artificial Life." *Artificial Life* 6(4): 363-376.
- Bedau, M. A., E. Snyder, et al. (1998). *A Classification of Long-Term Evolutionary Dynamics*. Artificial Life VI, MIT Press.
- Blüthgen, N., J. Fründ, et al. (2008). "What do interaction network metrics tell us about specialization and biological traits." *Ecology* 89(12): 3387-3399.
- Boulton, D. M. and C. S. Wallace (1969). "The information content of a multistate distribution." *Journal of Theoretical Biology* 23: 269-278.
- Bronowski, J. (1970). "New concepts in the evolution of complexity." *Synthese* 21: 228-246.
- Channon, A. (2006). "Unbounded evolutionary dynamics in a system of agents that actively process and transform their environment." *Genetic Programming and Evolvable Machines* 7(3): 253-281.
- Channon, A. D. and R. I. Dampier (2000). "Towards the evolutionary emergence of increasingly complex advantageous behaviours." *International Journal of Systems Science* 31(7): 843-860.
- Dorin, A., K. B. Korb, et al. (2008). Artificial-Life Ecosystems: What are they and what could they become? *Eleventh International Conference on Artificial Life*. S. Bullock, J. Noble, R. A. Watson and M. A. Bedau, MIT Press: 173-180.
- Dunne, J. A., R. J. Williams, et al. (2002). "Food-web structure and network theory: The role of connectance and size." *Proceedings of the National Academy of Sciences* 99(20): 12917-12922.
- Korb, K. B. and A. Dorin (2009). A simulation of niche construction, Bayesian Intelligence. 2009/2: 13.
- Korb, K. B. and A. Dorin (2010). Evolution unbound: releasing the arrow of complexity. *Bayesian Intelligence Technical Report*, Bayesian Intelligence. 2010/2: 21.
- Maynard-Smith, J. and E. Szathmáry (1995). *The Major Transitions in Evolution*. New York, W.H. Freeman.
- McShea, D. W. (1991). "Complexity and evolution: what everybody knows." *Biology and Philosophy* 6(3): 303-324.
- McShea, D. W. (1994). "Mechanisms of large-scale evolutionary trends." *Evolution* 48(6): 1747-1763.
- Neutel, A.-M., J. A. P. Heesterbeek, et al. (2007). "Reconciling complexity with stability in naturally assembling food webs." *Nature* 449: 599-602.
- Odling-Smee, F. J., K. N. Laland, et al. (2003). *Niche Construction, the neglected process in evolution*. Princeton, Princeton University Press.
- Proulx, S. R., D. E. L. Promislow, et al. (2005). "Network thinking in ecology and evolution." *Trends in Ecology and Evolution* 20(6): 345-353.
- Ray, T. S. (1990). *An approach to the synthesis of life*. Artificial Life II, Santa Fe, New Mexico, Addison Wesley.
- Watts, D. J. and S. H. Strogatz (1998). "Collective dynamics of small-world networks." *Nature* 393: 440-442.