Brigham Young University
BYU ScholarsArchive

International Congress on Environmental
Modelling and Software

2nd International Congress on Environmental
Modelling and Software - Osnabrück, Germany -
June 2004

Jul 1st, 12:00 AM

Ecosystems as Evolutionary Complex Systems: A Synthesis of Two System-Theoretic Approaches Based on Boolean Networks

Brian D. Fath

W. E. Grant

Follow this and additional works at: https://scholarsarchive.byu.edu/iemssconference

Ecosystems as Evolutionary Complex Systems: A Synthesis of Two System-Theoretic Approaches Based on Boolean Networks

Brian D. Fath[a] and W. E. Grant[b]

[a] *Biology Department, Towson University, Towson, Maryland 21252, USA*

[b] *Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, Texas, USA*

Abstract: Understanding and managing ecosystems as biocomplex wholes is the compelling scientific challenge of our times. Several different system-theoretic approaches have been proposed to study biocomplexity and two in particular, Kauffman's NK networks and Patten's ecological network analysis, have shown promising results. This research investigates the similarities between these two approaches, which to date have developed separately and independently. Kauffman (1993) has demonstrated that networks of non-equilibrium, open thermodynamic systems can exhibit profound order (subcritical complexity) or profound chaos (fundamental complexity). He uses Boolean NK networks to describe system behavior, where N is the number of nodes in the network and K the number of connections at each node. Ecological network analysis uses a different Boolean network approach in that the pair-wise node interactions in an ecosystem food web are scaled by the throughflow (or storage) to determine the probability of flow along each pathway in the web. These flow probabilities are used to determine system-wide properties of ecosystems such as cycling index (Finn 1976), indirect-to-direct effects ratio, and synergism. Here we use a modified version of the NK model to develop a fitness landscape of interacting species and calculate how the network analysis properties change as the model's species coevolve. We find that, of the parameters considered, network synergism increases modestly during the simulation whereas the other properties generally decrease. This research is largely a proof of concept test and will lay the foundation for future integration and model scenario analysis between two important network techniques.

Keywords: Boolean networks, Coevolution, Ecological Modeling, Fitness landscapes, Network Analysis.

1. INTRODUCTION

One goal of theoretical ecosystem ecology is to identify and quantify system-level concepts and find general patterns of ecosystem organization. One promising method has been to conceptualize ecosystems as networks connected by their transfer and exchange of energy and matter within and across system boundaries. Several different developments of this conceptualization have been realized. Independently, they have added significantly to our understanding of ecosystems yet there has been a lack of integration with these methods because of the different terminology, notation, history, disciplinary genesis, emphasis, and application. The main goal of this project is to find linkages between two commonly used Boolean representations of ecological networks. In particular, we link ecosystem theory based on network analysis to Kauffman's theory of self-organized systems in order to test the hypothesis that network properties of homogenization, amplification indirect effects, and synergism increase as an ecosystem co-evolves to higher fitness levels.

2. BACKGROUND

2.1 Ecological Network Analysis

Bernard Patten used mathematical systems theory as a foundation for studying ecosystems (Patten et al. 1976, Patten 1978, 1981). He stressed the utility of the inclusion-exclusion principle of set theory as a way to formalize the transactions that naturally occur in food webs. A binary interaction exists in ecological networks, simplified often as a question of "who eats whom", but more broadly as the transfer of conservative energy–matter between any two entities in the system. Much of the subsequent work in network ecology builds on this

basic premise of direct energy-matter transactions between coupled binary pairs. These transactions form the basis of both direct and indirect ecological relations, such as predation (direct), neutralism (direct), altruism (direct), mutualism (indirect) and competition (indirect) that are of importance to community ecology. Some of the primary findings of this research include the importance of indirect effects as they propagate through the myriad of network connections (Higashi and Patten 1989) and synergism, individual compartments in an ecosystem gaining positive value from being embedded in a larger network (Patten 1992, Fath and Patten 1998).

2.2 Ecological Network Properties

Several network properties have been developed with four in particular: amplification, indirect effects, homogenization, and synergism used most regularly to investigate ecosystem behavior. Since they have been described elsewhere, only a brief description is provided here (see Fath and Patten (1999) for the details). The four properties relate the distribution and contribution of conservative energy-matter flow through the network's many direct and indirect pathways. One measure of resource distribution is given in the direct flow intensity, or transfer efficiency, matrix $\mathbf{G}$, whose values, $g_{ij}=f_{ij}/T_j$, represent the likelihood of flow along a given path, where f_{ij} corresponds to the flow from compartment j to compartment i, and $T_j=\sum_{j(\neq i)=0,n} f_{ij}$ is the total sum of flow through compartment j including input and output boundary flows (f_{i0} and f_{0j}, respectively). $T_j^{in}=T_j^{out}$ at steady state. In the direct flow intensity matrix, $\mathbf{G}$, all elements have a non-negative value less than one ($0 \leq g_{ij} < 1$) and can be interpreted as a probability of flow along each pathway. Using standard input-output analysis techniques, an integral flow intensity matrix, $\mathbf{N}$, is computed from the convergent power series:
$$\mathbf{N} = \mathbf{G}^0 + \mathbf{G}^1 + \mathbf{G}^2 + \mathbf{G}^3 + ... = (\mathbf{I} - \mathbf{G})^{-1} \qquad (1)$$
where $\mathbf{I}$ is the multiplicative identity matrix.

Elements n_{ij} in the matrix $\mathbf{N}$ include the contribution of direct (m=1) and indirect (m>1) pathways, and therefore are always greater than or equal than the values of $\mathbf{G}$. The $\mathbf{G}$ and $\mathbf{N}$ matrices are used to define the amplification, homogenization, and indirect effects properties A specific quantitative test exists to determine each property (Figure 1).

Amplification occurs whenever an off-diagonal element of the integral flow matrix is greater than one ($n_{ij}>1$). The integral flow from j to i, can exceed one when cycling drives more than the equivalent of one unit of input flow over the

pathway. This property was observed in several of the small-scale models but is rare in large-scale models (Fath 2004).

Property	Test
Amplification	$n_{ij} > 1 \; for \; i \neq j$
Homogenization	$\dfrac{CV(G)}{CV(N)} > 1$
Synergism	$\dfrac{\sum + utility}{\left| \sum - utility \right|} > 1$
Ratio of direct to indirect effects	$\dfrac{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} (n_{ij} - i_{ij} - g_{ij})}{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} g_{ij}} > 1$

Figure 1. Four network properties

The homogenization property compares the resource distribution between the direct and integral flow intensity matrices. It was observed that, due to the contribution of indirect pathways, flow in the integral matrix was more evenly distributed or more homogenized than that in the direct matrix, meaning that flow is comprised of contributions from many parts of the network. Network homogenization occurs when the coefficient of variation of $\mathbf{N}$ is less than the coefficient of variation of $\mathbf{G}$ because this indicates that the network flow is more evenly distributed in the integral matrix.

Indirect effects are calculated as the integral contributions minus the direct and initial boundary input (Indirect = $\mathbf{N}$–$\mathbf{I}$–$\mathbf{G}$). The indirect to direct effects ratio is a measure of the relative strength of these two factors. When the ratio is greater than one, then indirect effects are greater than direct effects.

The fourth property, network synergism is based on a net flow intensity matrix, $\mathbf{D}$, where $d_{ij}=(f_{ij}-f_{ji})/T_i$. Unlike the other series in which the elements are non-negative, entries in $\mathbf{D}$ can be positive or negative ($-1 \leq d_{ij} < 1$). The elements of $\mathbf{D}$ represent the relative utility between that (i,j) pairing. An integral utility matrix $\mathbf{U}$, is obtained from the power series as:
$$\mathbf{U} = \mathbf{D}^0 + \mathbf{D}^1 + \mathbf{D}^2 + \mathbf{D}^3 + ... = (\mathbf{I} - \mathbf{D})^{-1} \qquad (2)$$

This methodology is used to determine qualitative relations between any two components in the

network such as predation, mutualism, competition, etc. Synergism arises when integral positive utility exceeds negative utility because of mutualistic relations in the system and is calculated as the ratio of the magnitude of the positive and negative utilities.

2.3   Kauffman's NK Model

Stuart Kauffman uses binary Boolean networks to find general laws of system self-organization (Kauffman 1993, 1996, 2000). His main thesis is that biological systems are composed of autonomous agents, or self-replicating systems that perform work, which are "co-constructing and propagating organization" (Kauffman 2000, p. 5). An emphasis is placed on co-construction and coevolution because of the cybernetic feedback that makes agents adapt to other agents at the same time they modify their own environment. There recently has been renewed interest in the impact species have on each other and on their environment (e.g., Jones et al. 1997, Odling-Smee et al. 2003). Coevolution and indirect effects are both manifestations of interacting networks.

In his NK model, Kauffman (2000) addresses species coevolution by coupling the influence from genes of one species to genes of another species. The basic module of the NK model represents an organism with N genes, each having two alleles, 0 and 1. The contribution of each gene to the fitness of the organism depends on the allele of that gene and the alleles of K other genes in its genome, called "epistatic" inputs. In this simple model there are 2^N combinations of alleles that influence fitness. Each allele combination is randomly assigned a fitness contribution value. The average fitness of the N genes is taken as the mean of the random values. The result is a fitness landscape, such that every allele combination has a specific fitness value (Table 1 shows an example for N=3).

When there is a flip in one allele from 0 to 1 or vice versa, the fitness contribution of the gene changes. If the result is higher fitness, then the allele shift is accepted, if not, then it is not accepted. Kauffman found that when the number of connections to other genes, K, is low the system quickly evolves to a global fitness maximum. As the number of connections increases there are more local peaks until the point when the system is completely interconnected (K=N–1) and the resulting fitness landscape is fully random. The more local peaks that occur, the more improbable it is to "climb" to the global peak, resulting, on average, in an overall lower fitness. However, Kauffman maintains that fitness landscapes are not random but instead are generated by the coevolutionary interactions of the various species.

Therefore, the next step is to link NK models of various species.

Table 1.   There are eight possible binary combinations of 3 genes. Each is assigned a random fitness value between 0 and 1, and the fitness for each allele combination is the mean of the three values. This procedure is used to construct a fitness landscape. For example, starting with each gene expressing a 0, the fitness is 0.37. If the allele on the first gene flips to "1" then fitness increases to 0.43. This simple model has only one fitness peak at (0,1,0).

1	2	3	fitness value w_1	fitness value w_1	fitness value w_3	Average fitness $\overline{w}$
0	0	0	0.2	0.5	0.4	0.37
0	0	1	0.7	0.1	0.2	0.33
0	1	0	0.5	0.9	0.8	0.73
0	1	1	0.3	0.3	0.1	0.23
1	0	0	0.5	0.4	0.4	0.43
1	0	1	0.1	0.5	0.3	0.30
1	1	0	0.9	0.2	0.8	0.63
1	1	1	0.6	0.8	0.4	0.60

In the multi-species version of Kauffman's NK model, the fitness value of each allele depends not only on the allele of that gene and on the alleles of K epistatic genes, but also on the alleles of C other genes in each of S other species. If there are two species coupled together, then each gene has K+C inputs, and a table of random fitness contributions is generated that has $2^{(K+C)}$ combinations. A model in which each species is connected with S other species has $2^{(K+CS)}$ possible states, so the number of possible states grows combinatorically. The fitness of the species is calculated as the mean of the fitness values of the alleles in its current genotype; each species is assumed to be isogenic. Now, when one species evolves (a flipping of an allele on a gene) this likely has ramifications for the other species by deforming the overall fitness landscape. Kauffman found that in general coevolving systems coupled in this manner behave either in an ordered or chaotic regime, separated by a phase transition depending on the number of couplings.

We have recreated Kauffman's multi-species NK model here to investigate the fitness of coevolving species with a particular interest in understanding how ecosystem properties may be affected by the resultant coevolutionary processes. A few modifications to the original model as presented above are noted. Each time step during the simulation, any one of four events, randomly chosen, may happen.   (1) A randomly-chosen species may evolve to a new genotype via

recombination, if the randomly-chosen new genotype has a higher fitness value than the current genotype. A randomly-chosen species may be replaced by a new species that (2) may have a different K than the current species, but has the same C and S, (3) may have a different C, but has the same K and S, or (4) may have a different S, but has the same K and C, if the new species has a higher fitness value than the current species. Thus, as species evolve or are replaced by invading species, they change their own fitness landscape (Kauffman 1996, 2000) as well as the fitness landscape of the other species. The above restrictions could be relaxed in future research to study more general cases, but for now the model was used to generate a time series of connectance matrices. We apply ecological network analysis to each matrix. Eventually, it would be useful to look at models that have more realistic ecosystem structures by using the methodologies developed in Fath (2004) or perhaps to see if over time species in the models naturally evolve into a configuration similar to a trophic structure. However, that is beyond the scope of this paper. Here we present the initial results from this research, which uses a five species model coevolving for 10 time steps under two different species coupling regimes.

3. INTEGRATED MODEL

In the first simulation, all species were initialized with S=1 (i.e., each species is connected with one other species), and in the second simulation all species were initialized with S=4 (connected to all other species). Every time interval during the simulation, we generate a connectance matrix based on the current fitness and S values of the set of species. Elements of the connectance matrix are equal to 0 if the fitness values of the genes of the "to" species are not affected by the genes of the "from" species, and diagonal elements are equal to 0, that is, species are not connected to themselves. Values of the other elements of the connectance matrix are calculated as the fitness value of the "to" species divided by its S value, that is, the sum of all elements "to" a given species is equal to its fitness value.

The elements of the connectance matrix represent the fitness contribution among connected species. In order to apply ecological network analysis to these matrices, we assume that elements of the connectance matrix represent relative rates of energy flow among the set of species. Obviously, fitness is not flow, but in a more general sense the fitness represents a measure of influence between species. The flow probability between two compartments is the proportion of flow to total throughflow ($g_{ij} = f_{ij}/T_j$) where T_j is the total throughflow into compartment j. This could also be interpreted as the probability of influence between two compartments (Patten et al. 1976). Here we assume that the fitness contribution (from 0 to 1) can be used as a measure of the weighted influence. This allows us to apply ecological network analysis to each matrix and calculate the cycling index (Finn 1976) as well as the 4 ecological network properties described above.

We then examined the temporal dynamics of these properties as the set of species co-evolve through different fitness landscapes to test the hypothesis that cycling index, homogenization, amplification, indirect effects, and synergism increase as the ecosystem co-evolves. Note, that ecological network analysis is a steady-state analysis, however we treat the model generated from each time step as a snapshot in time. As the system changes over time, we can determine the network properties of the system in that particular state. One other assumption is needed to run the analysis, which is that the model ecosystems, as open systems, receive external input. Energy enters the system largely through primary producer and lower trophic level species. Usually, for a model this size (5 compartments) external input into one compartment is enough, but in some of these simulations the first compartment is eliminated after which time there would be no further input available to higher trophic levels. Therefore, a unit of input is given to each of the first two compartments. The other compartments receive flow from the network of interactions, which subsequently affects their fitness.

4. NUMERICAL SIMULATION RESULTS

In the first simulation, each species is connected to one other species. The connectance values can change at each time step given the occurrence of a randomly chosen event, as described above (Table 2 shows two matrices generated by the model at time steps 2 and 3). For example, we see that in the third time step a new species 2 appears which is also dependent on species 4 and the overall connectance or fitness from species 2 to species 1 increases. Changes such as these continue through to the end of the simulation after 10 time steps. When the ecological network properties of these connectance matrices from each time step are calculated we find the following: amplification does not occur at any time step; the cycling index, homogenization, and ratio of indirect to direct effects all decrease over time; and the synergism parameter rises steadily until a certain point at which it starts to drop (Figure 2).

Table 2. Example of 2 connectance matrices generated by the first simulation model. Reading from columns to rows, at time 2, Sp 2 affects Sp 1 (0.52), Sp 3 affects Sp 2 (0.42), Sp 4 affects Sp 3 (0.42), Sp 5 affects Sp 4 (0.66), and Sp 1 affects Sp 5 (0.61). At T=3, Sp 2 is replaced by a new Sp 2 that is affected by Sp 3 and Sp 4 (overall fitness is higher (0.46 versus 0.42). The new Sp 2 also has caused a change in the fitness value of Sp 1.

T=2	Sp 1	Sp 2	Sp 3	Sp 4	Sp 5
Sp 1	0	0.52	0	0	0
Sp 2	0	0	0.42	0	0
Sp 3	0	0	0	0.42	0
Sp 4	0	0	0	0	0.66
Sp 5	0.61	0	0	0	0

T=3	Sp 1	Sp 2	Sp 3	Sp 4	Sp 5
Sp 1	0	0.66	0	0	0
Sp 2	0	0	0.23	0.23	0
Sp 3	0	0	0	0.42	0
Sp 4	0	0	0	0	0.66
Sp 5	0.61	0	0	0	0

In the second simulation, each species is initially linked to four other species. Several changes occur immediately, most notably, the connection between Sp 4 and Sp 5 is lost. During the 10-step simulation the system becomes more articulated, meaning there are fewer connections between species, but these changes would only be accepted if the overall fitness of the species increases. One simple measure to consider is the total number of connections in the system during each time step (Table 3). We see a similar pattern in the network parameters in the second simulation as well. Amplification does not occur at any time step. Cycling index and indirect effects ratio decrease, while in this simulation homogenization bounces around but is fairly flat. Synergism also oscillates reaching a peak in the middle of the simulation and dropping again near the end (Figure 3, note in the figure that synergism is plotted on the alternate y-axis).

Table 3. Connections in Simulation A (species initially connected to one species) and Simulation B (species initially connected to four species)

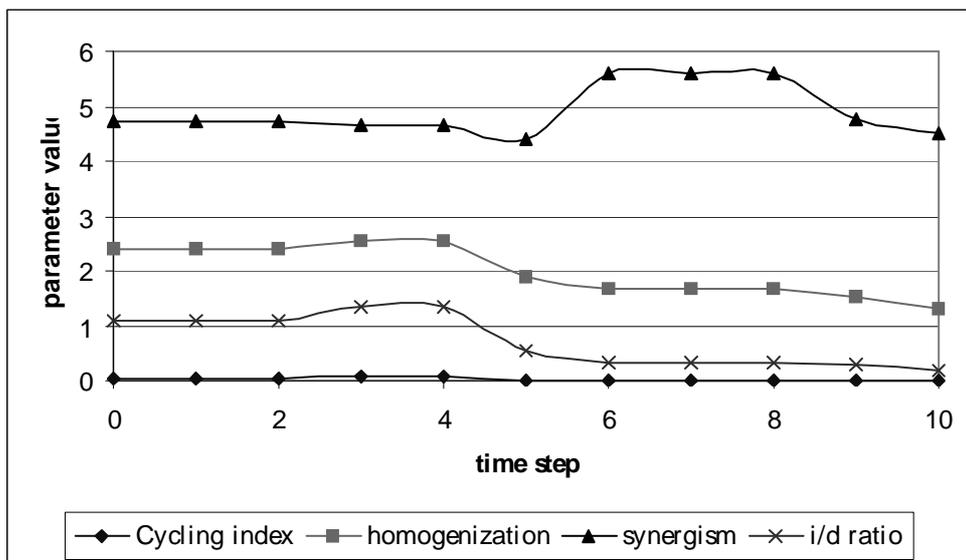| T	A: # links	B: # links
0	5	20
1	5	19
2	5	19
3	6	19
4	6	19
5	5	19
6	4	19
7	4	18
8	4	18
9	5	18
10	7	14

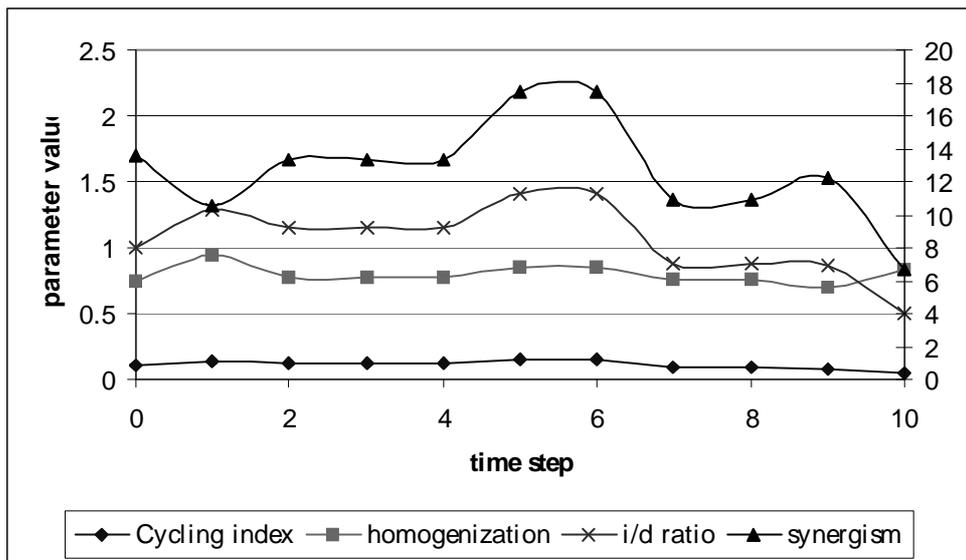Figure 2. Behavior of network properties over time for first simulation.

Figure 3. Behavior of network properties for second simulation. Synergism is plotted on the alternate y-axis.

5. CONCLUSIONS

In conclusion, we have recreated Kauffman's multi-species NK model and used it to investigate the coevolution of a simple model ecosystem. Furthermore, we have used the fitness values generated by the model as surrogates for the probability of influence between the compartments. This allows the application of network analysis techniques to determine the values of specific network properties. In particular, we found that network synergism appears to respond positively as fitness increases, and the other properties respond negatively. This paper represents the first attempt to integrate the two Boolean techniques; further research is needed to more deeply understand the interrelation between them. Future work along these lines is currently underway, in particular to see how various network-based ecological goal functions (Fath et al. 2001) respond to changes in fitness in these coevolutionary models.

6. REFERENCES

Fath, B.D., Network analysis applied to large-scale cyber-ecosystems, *Ecol. Modell.* 171, 329-337, 2004.

Fath, B.D., and B.C. Patten, Network synergism: emergence of positive relations in ecological systems. *Ecol. Modell.* 107, 127-143, 1998.

Fath B.D., and B.C. Patten, Review of the foundations of network environ analysis. Ecosystems, 2, 167-179, 1999.

Fath B.D., B.C. Patten, and J.S. Choi, Complementarity of ecological goal functions. *J. Theor. Biol.*, 208(4), 493-506, 2001.

Finn, J.T., Measures of ecosystem structure and function derived from flow analysis. *J. Theor. Biol.*, 56, 363-380, 1976.

Higashi, M., and B.C. Patten, Dominance of indirect causality in ecosystems. *Amer. Nat.*, 133, 288-302, 1989.

Jones, C.G., J.H. Lawton, and M. Shachak, Positive and negative effects of organisms as physical ecosystem engineers. *Ecology* 78, 1946-1957, 1997.

Kauffman, S.A., *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford University Press, New York, 1993

Kauffman, S.A., *At Home in the Universe: The Search for Laws of Self-Organization and Complexity.* Oxford University Press, New York, 1996.

Kauffman, S.A. *Investigations.* Oxford University Press, New York, 2000.

Odling-Smee, F.J, K.N, Laland, and M.W. Feldman. *Niche Construction: The neglected process in evoloution.* Princeton University Press, 472 pp. Princeton, 2003.

Patten, B.C., Systems approach to the concept of environment. *Ohio J. Sci.*, 78, 206-222. 1978.

Patten, B.C., Environs: the superniches of ecosystems. *Amer. Zoo.*, 21, 845-852, 1981.

Patten, B.C., Energy, emergy and environs. *Ecol. Modell.* 62, 29-69, 1992.

Patten, B.C., R.W. Bosserman, J.T. Finn, and W.G. Cale, Propagation of cause in ecosystems. In Patten, B.C. (Ed.), *Systems Analysis and Simulation in Ecology*, Vol. IV. Academic Press, 593 pp. New York, 457-579, 1976.