16 Graph theory and the evolution of autocatalytic networks

Sanjay Jain and Sandeep Krishna

Abstract

We give a self-contained introduction to the theory of directed graphs, leading up to the relationship between the Perron-Frobenius eigenvectors of a graph and its autocatalytic sets. Then we discuss a particular dynamical system on a fixed but arbitrary graph, that describes the population dynamics of species whose interactions are determined by the graph. The attractors of this dynamical system are described as a function of graph topology. Finally we consider a dynamical system in which the graph of interactions of the species coevolves with the populations of the species. We show that this system exhibits complex dynamics including self-organization of the network by autocatalytic sets, growth of complexity and structure, and collapse of the network followed by recoveries. We argue that a graph theoretic classification of perturbations of the network is helpful in predicting the future impact of a perturbation over short and medium time scales.

16.1 Introduction

Studies of networks are useful at several different levels (for recent reviews see [1–4]). At one level one is interested in describing the structure of natural and man-made networks such as food webs in ecosystems, biochemical and neural networks in organisms, networks of social interaction among agents in societies, and technological networks like the internet, etc. A useful representation of a network is a graph (and its generalizations) where the components of the network (which could be species, neurons, agents, etc.) are represented by nodes, and their mutual interactions by the links of the graph. Graph theory provides important tools to capture various aspects of the network structure.

At a second level one wants to know how the network structure of the system influences what happens in the system. E.g., the food-web structure of an ecosystem affects the dynamics of populations of the species, the network of human contacts influences the spread of a contagious disease, etc. At this level of discussion the network is typically taken to be static on the time scales of interest; the prime concern is the dynamics of other variables on a network with some particular type of (fixed) structure. Here dynamical systems theory is a major tool, and network variables (like the adjacency matrix elements of the underlying graph) appear as fixed parameters in the dynamics of other system variables like population, etc.

At a third level one is interested in how networks themselves change with time. Biochemical, neural, ecological, social and technological networks are not static, but are products of evolution. Moreover this evolution is quite complex in real systems. Networks sometimes self-organize and grow in size and complexity, and sometimes disintegrate. Their evolution is usually intertwined with other system variables, e.g., a food-web influences populations of species, and if a species goes extinct, the food-web changes. Understanding the processes and mechanisms involved in the evolution of complex networks is a major intellectual challenge.

A problem that illustrates all these levels is the problem of the origin of life on earth. The simplest living structure that we know — a bacterial cell — is a complex collection of several thousand types of molecules interacting with each other in a complex network of chemical interactions. The network may be described by a graph in which the nodes represent the molecular types or molecular species, and links connecting nodes represent chemical interactions between the molecular species. By participating in specific chemical reactions each molecular species or node plays a rather definite functional role in the organization of the cell: it permits or creates certain specific processes or spatial structures. Note that the complex chemical network of a cell is needed to produce the processes and structures that exist in it, and conversely, the same processes and structures are essential for maintaining the network and allowing it to evolve. If we assume that life originated on earth about 3.5 to 3.8 billion years ago as suggested by the microfossil evidence, then about 4 billion years back there was neither such a complex network of interactions nor such processes and structures existing anywhere on the earth. One of the puzzles of the origin of life on earth is: how did the network and the processes and spatial structures bootstrap themselves into existence when none was present — how did a chemical 'organization' emerge with individual molecular species playing definite roles in it?

A second puzzle concerns the highly 'structured' nature of the organization. The molecules appearing in cells are very special (a small subset in a very large space of possible molecules) and so is the graph that describes their interactions (a special kind of graph in the very large space of graphs). The probability of such structures arising by pure chance is astronomically small. If we assume that it was not an unlikely chance event that created life, we are led to the question: what then are the mechanisms that can create highly structured or 'ordered' organizations? A similar question is relevant for economic and social networks.

In order to address such questions in a mathematical model, one is naturally led to dynamical systems in which the graph describing the network is also a dynamical variable, whose dynamics is coupled to that of other variables such as the population of the molecular species. Here we present a model with such a structure, which has been inspired by the work in refs. [5–10]. The analysis of such dynamical systems is facilitated by the development of some new tools in graph theory. Another purpose of this article is to discuss some of these new tools. Together, the model and these tools address the above two questions about the origin of life, and provide partial answers. The model exhibits a mechanism by which a chemical organization can emerge where none existed through the formation of small *autocatalytic sets* of molecular species. In the model we also observe a *self-organizing process* which results in the growth of the initial autocatalytic set into a complex and highly structured chemical organization in a short time.

In addition, the model also captures, in an analytically tractable form, several phenomena that one associates with the evolution of other biological and social systems. These include

emergence of cooperation and interdependence in the system; crashes and recoveries of the system as a whole; 'core-shifts'; appearance of 'keystone species'; etc. We also argue that the juxtaposition of graph theory and dynamical systems provides the possibility of formulating more precisely notions that are important and useful in everyday language but otherwise difficult to pin down. In particular we attempt to formulate the notion of 'innovation' in this dynamical system, and classify innovations into categories according to their graph theoretic structure. It turns out that different categories of innovation have different short and longer term impact on the dynamics of the system.

This article is organized roughly according to the three kinds of network studies indicated above. In section 2 we discuss aspects of graph theory in a self-contained manner, reviewing older results as well as recent work. Among other things we describe a relation between topological properties of a graph (namely its autocatalytic sets) and its algebraic properties (the structure of the eigenvectors of its adjacency matrix). In section 3 we discuss a simple dynamical system describing molecular population dynamics on a fixed interaction graph. Here we show how structure of the graph influences the dynamics of the system; in particular relating the nature of its attractors to graph topology. Section 4 describes a model of graph evolution, motivated by the origin of life problem. In section 5 we show that the dynamics of this model exhibits self-organization and growth of cooperation and structure in the network, with analytical estimates of the time scales involved. Section 6 discusses the phenomena of crashes and recoveries exhibited by the model. In this section we also formulate a definition of innovation that seems appropriate for this model, and discuss a hierarchy of different categories of innovation and the roles they play in the ups and downs of the system. Finally, section 7 contains a discussion of some limitations of the model, speculations regarding the origin of life problem and possible future directions.

16.2   Graph theory and autocatalytic sets

16.2.1   Directed graphs and their adjacency matrices

A *directed graph* $G = G(S, L)$, often referred to in the sequel as simply a *graph*, is defined by a set S of 'nodes' and a set L of 'links' (or 'arcs'), where each link is an ordered pair of nodes [11, 12]. It is convenient to label the set of nodes by integers, $S = \{1, 2, \ldots, s\}$ for a graph of s nodes. An example of a graph is given in Figure 16.1a where each node is represented by a small labeled circle, and a link (j, i) is represented by an arrow pointing from node j to node i. A graph with s nodes is completely specified by an $s \times s$ matrix, $C = (c_{ij})$, called the *adjacency matrix* of the graph, and vice versa. The matrix element in the ith row and j^{th} column of C, c_{ij}, equals unity if L contains a directed link (j, i) (arrow pointing from node j to node i), and zero otherwise. (This convention differs from the usual one where

$c_{ij} = 1$ if and only if there is a link from node i to node j; our adjacency matrix is the transpose of the usual one. We have chosen this convention because it is more natural in the context of the dynamical system to be discussed in subsequent sections.) Figure 16.1b shows the adjacency matrix corresponding to the graph in Figure 16.1a. We will use the terms 'graph' and 'adjacency matrix' interchangeably: the phrase 'a graph with adjacency matrix C' will often be abbreviated to 'a graph C'. Undirected graphs are special cases of directed
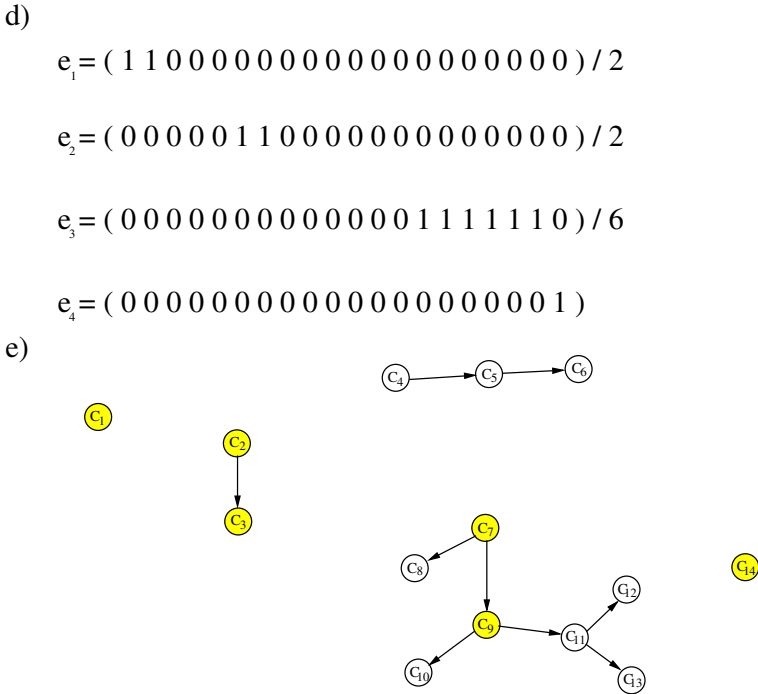
a)

b)

$$
C = \begin{pmatrix}
0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0 \\
0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1
\end{pmatrix}
$$

c)

d)

$$e_1 = (\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ )\ /\ 2$$

$$e_2 = (\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ )\ /\ 2$$

$$e_3 = (\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 0\ )\ /\ 6$$

$$e_4 = (\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ )$$

e)

Figure 16.1: a. A directed graph with 20 nodes. **b.** The adjacency matrix of the graph in Figure 16.1a. **c.** A subgraph of the graph in Figure 16.1a. The adjacency matrix of the subgraph is the shaded portion of the matrix in Figure 16.1b. **d.** Four Perron-Frobenius eigenvectors (PFEs) of the graph in Figure 16.1a. The first three vectors have been divided by factors of 2, 2 and 6 respectively to normalize them. **e.** The irreducible decomposition of the graph in Figure 16.1a into subgraphs C_α, with $\alpha = 1, 2, \ldots, 14$. Each of the 14 nodes of the graph in Figure 16.1e represents either an irreducible subgraph of the graph in Figure 16.1a, or a single node that is not part of any irreducible graph. The basic subgraphs of the graph in Figure 16.1a are represented by yellow nodes. The dotted lines in Figure 16.1b demarcate the adjacency matrices corresponding to the subgraphs C_α. Colours identify the attractor of the dynamics discussed in section 3, except in Figure 16.1e. In all graphs in the article (except Figure 16.1e), white nodes have zero relative population in the attractor, $X_i = 0$, while blue and red nodes have $X_i > 0$. In graphs that have an autocatalytic set, red nodes belong to the core of the dominant autocatalytic set of the graph, blue nodes to its periphery, and white nodes are outside the dominant autocatalytic set.

graphs whose adjacency matrices are symmetric. A single (undirected) link of an undirected graph between, say, nodes j and i, can be viewed as two directed links of a directed graph, one from j to i and the other from i to j.

A graph $G' = G'(S', L')$ is called a *subgraph* of $G(S, L)$ if $S' \subset S$ and $L' \subset L$. We will use the term 'subgraph' if G' satisfies a stronger property: every link in L with both endpoints in S' also belongs to L'. That is, for us, a subgraph will be a subset of nodes together with *all* their mutual links. (This is often called an 'induced subgraph' in the literature [12].) The graph in Figure 16.1c (comprising nodes 14, 15, 16, 17, 18 and 19 and all their mutual links) is thus a subgraph of the graph in Figure 16.1a. For a subgraph we will often find it more convenient to label the nodes not by integers starting from 1, but by the same labels the corresponding nodes had in the parent graph. The adjacency matrix of a subgraph can be obtained by deleting all the rows and columns from the full adjacency matrix that correspond to the nodes outside the subgraph. The highlighted portion of the matrix in Figure 16.1b is the adjacency matrix of the subgraph in Figure 16.1c.

A *walk* of length n (from node i_1 to node i_{n+1}) is an alternating sequence of nodes and links $i_1 l_1 i_2 l_2 \ldots i_n l_n i_{n+1}$ such that link l_1 points from node i_1 to node i_2 (or $l_1 = (i_1, i_2)$), l_2 points from i_2 to i_3 and so on. A walk with all nodes distinct (except possibly the first and last nodes) will be called a *path*. If the first and last nodes i_1 and i_{n+1} of a walk or path are the same, it will be referred to as a *closed* walk or path. The existence of even one closed walk in the graph implies the existence of an infinite number of distinct walks in the graph. In the graph of Figure 16.1a, there is an infinite number of walks from node 11 to node 17 (e.g., $11 \to 12 \to 14 \to 17$, $11 \to 12 \to 11 \to 12 \to 14 \to 17$, ...) but no walks from node 11 to node 10. An undirected graph trivially has closed walks if it has any undirected links at all.

In the graph theory literature, what we have defined above to be a 'closed path' is usually referred to as a 'cycle'. However, for later convenience, we define a cycle somewhat differently. We define an *n-cycle* to be a subgraph with $n \geq 1$ nodes which contains exactly n links and also contains a closed path that covers all n nodes. E.g., the subgraph formed by node 20 and its self link is a 1-cycle, that formed by nodes 1 and 2 is a 2-cycle and by nodes 3,4 and 5 a 3-cycle. The subgraph formed by nodes 1,2,3,4 and 5 is not a 5-cycle because it does not have a closed path covering all the five nodes. The word 'cycle' will be used generically for an *n*-cycle of unspecified length. Given a directed graph C, its *associated undirected graph* (or 'symmetrized version') $C^{(s)}$ can be obtained by adding additional links as follows: for every link (j, i) in L, add another link (i, j) if the latter is not already in L. Two nodes of a directed graph C will be said to be *connected* if there exists a path between them in the associated undirected graph $C^{(s)}$, and disconnected otherwise. Thus any directed graph can be decomposed into 'connected components' which are maximal sets of connected nodes (e.g., the graph of Figure 16.1a has five connected components that are disconnected from each other). In a directed graph C, we refer to a node i as being 'downstream' from a node j if there is a path in C leading from j to i, and no path from i to j. Similarly i is 'upstream' from j if there is a path in C leading from i to j, and no path from j to i. Thus in Figure 16.1a, node 17 is downstream from node 11, or equivalently node 11 is upstream from node 17. Node 10 is neither upstream nor downstream from node 11 since they are not connected, and node 12 is neither upstream nor downstream from 11 because each can be reached from the other along some directed path.

If C is the adjacency matrix of a graph then it is easy to see that $(C^n)_{ij}$ equals the number of distinct walks of length n from node j to node i. E.g., $C_{ij}^2 = \sum_{k=1}^s C_{ik}C_{kj}$; each term in the sum is unity if and only if there exists a link from j to k and from k to i; hence the sum counts the number of walks from j to i of length 2.

Perron-Frobenius eigenvalues and eigenvectors (PFEs)

A vector $\mathbf{x} = (x_1, x_2, \ldots, x_s)$ is said to be an eigenvector of an $s \times s$ matrix C with an eigenvalue λ if for each i, $\sum_{j=1}^s c_{ij}x_j = \lambda x_i$. The eigenvalues of a matrix C are roots of the *characteristic equation* of the matrix: $|C - \lambda I| = 0$ where I is the identity matrix of the same dimensionality as C and $|A|$ is the determinant of the matrix A. In general a matrix will have complex eigenvalues and eigenvectors, but an adjacency matrix of a graph has special properties, because it is a 'non-negative' matrix, i.e., it has no negative entries.

For any non-negative matrix, the Perron-Frobenius theorem [13, 14] guarantees that there exists an eigenvalue which is real and larger than or equal to all other eigenvalues in magnitude. This largest eigenvalue is often called the Perron-Frobenius eigenvalue of the matrix, which we will denote by $\lambda_1(C)$ for a graph C. Further the theorem also states that there exists an eigenvector of C corresponding to $\lambda_1(C)$ (which we will refer to as a Perron-Frobenius Eigenvector, PFE) all of whose components are real and non-negative. The Perron-Frobenius eigenvalue of the graph in Figure 16.1a is 1. Four PFEs of the graph in Figure 16.1a are displayed in Figure 16.1d.

The presence or absence of closed paths in a graph can be determined from the Perron-Frobenius eigenvalue of its adjacency matrix (see ref. [16] for a simple proof):

Proposition 1. *If a graph, C,*
(i) *has no closed walk then* $\lambda_1(C) = 0$,
(ii) *has a closed walk then* $\lambda_1(C) \geq 1$,
(iii) *has a closed walk and all closed walks only occur in subgraphs that are cycles then* $\lambda_1(C) = 1$.

Note that λ_1 cannot take values between zero and one because of the discreteness of the entries of C which are either zero or one. (Thus, for an undirected graph, if it has even one undirected link, $\lambda_1(C) \geq 1$.) Several results pertaining to the relationship of the graph structure to the structure of its PFEs can be found in ref. [15].

Irreducible graphs and matrices

A subgraph of a directed graph is termed *irreducible* if there is a path within the subgraph from each node in the subgraph to every other node in the subgraph. The simplest irreducible subgraph is a 1-cycle. In Figure 16.1a the subgraph comprising nodes 3,4 and 5 is irreducible, as is the subgraph of nodes 6 and 7, but the subgraph of nodes 3,4,5,6 and 7 is not irreducible since there is, for example, no path from node 6 to node 5.

If a graph or subgraph is irreducible then the corresponding adjacency matrix is also termed *irreducible*. Thus a matrix C is *irreducible* if for every ordered pair of nodes i and j there exists a positive integer k such that $(C^k)_{ij} > 0$. Refs. [13, 14] describes further properties of irreducible matrices.

The nodes of any graph can be grouped into a unique set of irreducible subgraphs as follows:

(1) Pick any node, say i. Find all the nodes which have paths leading to them starting at i. Denote this set by S_1; it may include i itself. Similarly find all the nodes which have paths leading to i. Denote this set by S_2. Denote the subgraph formed by the set of nodes $\{i\} \cup (S_1 \cap S_2)$ and all their mutual links as C_1. If $S_1 \cap S_2 \neq \phi$, then C_1 is an irreducible graph because every node of C_1 has a path within C_1 to every other node in it. If $S_1 \cap S_2 = \phi$, then i does not belong to any irreducible subgraph and C_1 consists of just the node i and no links.

(2) Pick another node which is not in C_1 and repeat the procedure with that node to get another subgraph, C_2. The sets of nodes comprising the two subgraphs will be disjoint.

(3) Repeat this process until all nodes have been placed in some C_α, $\alpha = 1, 2, \ldots, M$. Each C_α is either an irreducible subgraph or consists of a single node with no links.

Irrespective of which nodes are picked and in which order, this procedure will produce for any graph a unique set of disjoint subgraphs (upto labelling of the C_α) encompassing all the nodes of the graph. The graph in Figure 16.1a will decompose into 14 such subgraphs (see Figure 16.1e).

We say there is a path from an irreducible subgraph C_1 to another irreducible subgraph C_2 if there is a path in C from any node of C_1 to any node of C_2. The terms 'downstream' and 'upstream' can thus be used unambiguously for the C_α.

Structure of a general graph

A general adjacency matrix can be rewritten in a useful form by renumbering the nodes by the following procedure [13, 14]:

Determine all the subgraphs $C_1, C_2, \ldots, C_M$ of the graph as described above. Construct a new graph of M nodes, one node for each C_α, $\alpha = 1, \ldots, M$. The new graph has a directed link from C_β to C_α if, in the original graph, any node of C_β has a link to any node of C_α. Figure 16.1e illustrates what this new graph looks like for the graph of Figure 16.1a.

Clearly the resulting graph cannot have any closed paths. For if it were to have a closed path then the C_α subgraphs comprising the closed path would together have formed a larger irreducible subgraph in the first place. Therefore we can renumber the C_α such that if $\alpha > \beta$, C_β is never downstream from C_α. Now we can renumber the nodes of the original graph such that nodes belonging to a given C_α occupy contiguous node numbers, and whenever a pair of nodes i and j belong to different subgraphs C_α and C_β respectively, then $\alpha > \beta$ implies $i > j$. Such a renumbering is in general not unique, but with any such renumbering the adjacency matrix takes the following canonical form:

$$C = \begin{pmatrix} C_1 & & & & 0 \\ & C_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ R & & & & C_M \end{pmatrix}$$

where 0 indicates that the upper block triangular part of the matrix contains only zeroes while the lower block triangular part, R, is not equal to zero in general. It can be seen that the graph
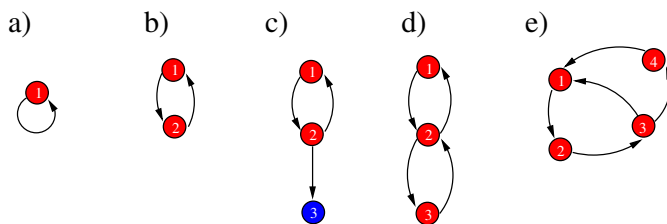
Figure 16.2: Various autocatalytic sets (ACSs). **a.** A 1-cycle, the simplest ACS. **b.** A 2-cycle. **c.** An ACS which is not an irreducible graph. **d,e** Examples of ACSs which are irreducible graphs but not cycles.

in Figure 16.1a is already in this canonical form. In Figure 16.1b, the dotted lines demarcate the block diagonal portions which correspond to the C_α.

From the above form of C it follows that

$$|C - \lambda I| = |C_1 - \lambda I| \times |C_2 - \lambda I| \times \ldots \times |C_M - \lambda I|$$

Therefore the set of eigenvalues of C is the union of the sets of eigenvalues of $C_1, \ldots, C_M$. $\lambda_1(C) = \max_\alpha \{\lambda_1(C_\alpha)\}$.

Therefore if a given graph C has a Perron-Frobenius eigenvalue $\lambda_1 > 0$ then it contains at least one irreducible subgraph with Perron-Frobenius eigenvalue λ_1. When $\lambda_1 > 0$, all irreducible subgraphs of C with Perron-Frobenius eigenvalue equal to λ_1 are referred to as *basic* subgraphs. The yellow nodes in Figure 16.1e correspond to the basic subgraphs of Figure 16.1a.

16.2.2   Autocatalytic sets

The concept of an autocatalytic set (ACS) was first introduced in the context of a set of catalytically interacting molecules. There it was defined to be a set of molecular species which contains a catalyst for each of its member species [17–19]. Such a set of molecular species can collectively self-replicate under certain circumstances even if none of its component molecular species can individually self-replicate. This property is considered important in understanding the origin of life. If we imagine a node in a directed graph to represent a molecular species and a link from j to i as signifying that j is a catalyst for i, this motivates the following graph-theoretic definition of an ACS in any directed graph: An *autocatalytic set* (ACS) is a subgraph, each of whose nodes has at least one incoming link from a node belonging to the same subgraph.

Figure 16.2 shows various ACSs. The simplest ACS is a 1-cycle; Figure 16.2a. There is the following hierarchical relationship between cycles, irreducible subgraphs and ACSs: all cycles are irreducible subgraphs and all irreducible subgraphs are ACSs, but not all ACSs are irreducible subgraphs and not all irreducible subgraphs are cycles. Figures 16.2a and 16.2b are graphs that are irreducible as well as cycles, 16.2c is an ACS that is not an irreducible

subgraph and hence not a cycle, while 16.2d and 16.2e are examples of irreducible graphs that are not cycles. It is not difficult to see the following [16]:

Proposition 2.
(i) *An ACS must contain a closed path*. Consequently,
(ii) *If a graph C has no ACS then $\lambda_1(C) = 0$.*
(iii) *If a graph C has an ACS then $\lambda_1(C) \geq 1$.*

Relationship between autocatalytic sets and Perron-Frobenius eigenvectors

The ACS is a useful graph-theoretic construct in part because of its connection with the PFE. Let $\mathbf{x}$ be a PFE of a graph. Consider the set of all nodes i for which x_i is non-zero. We will call the subgraph of all these nodes and their mutual links the 'subgraph of the PFE $\mathbf{x}$'. If all the components of the PFE are non-zero then the subgraph of the PFE is the entire graph. For example the subgraph of the PFE $\mathbf{e}_3$ mentioned in Figure 16.1d is is the graph shown in Figure 16.1c. One can show that [16]

Proposition 3
If $\lambda_1(C) > 0$, then the subgraph of any PFE of C is an ACS.

For the PFEs of Figure 16.1d this is immediately verified by inspection. Note that this result relates an algebraic property of a graph, its PFE, to a topological structure, an ACS. Further, this result is not true if we considered irreducible graphs intead of ACSs. E.g., the subgraph of $\mathbf{e}_3$, shown in Figure 16.1c, is not an irreducible graph.

Note also that the converse of the above statement is not true, i.e., there need not exist a PFE for every ACS in a given graph. Thus in Figure 16.1a, nodes 3,4,5,6 and 7 form an ACS but there is no eigenvector with eigenvalue λ_1 for which all these and only these components are non-zero.

Let $\mathbf{x}$ be a PFE of a graph C, and let C' denote the adjacency matrix of the subgraph of $\mathbf{x}$. Let $\lambda_1(C')$ denote the Perron-Frobenius eigenvalue of C'. It is not difficult to see that $\lambda_1(C') = \lambda_1(C)$. Figure 16.3 illustrates this point.

For the graph in Figure 16.3a $\lambda_1 = 1$. Figure 16.3b shows a PFE of the graph and how it satisfies the eigenvalue equations. For this PFE, nodes 1, 5 and 6 have $x_i = 0$. Removing these nodes produces the PFE subgraph shown in Figure 16.3c. Its adjacency matrix, C', is obtained by removing rows 1, 5, 6 and columns 1, 5, 6 from the original matrix. Figure 16.3d illustrates that the vector constructed by removing the zero components of the PFE is an eigenvector of C' with eigenvalue 1. The logic of this example is easily extended to a general proof that $\lambda_1(C') = \lambda_1(C)$.

We can now perform a graph decomposition of C' into irreducible subgraphs as before; since $\lambda_1(C') = \lambda_1(C)$, it follows that C' must contain at least one of the basic subgraphs of C. If C' contains only one of the basic subgraphs of C we will refer to $\mathbf{x}$ as a *simple* PFE, and to C' as a *simple ACS*. The graph in Figure 16.1a has only four simple PFEs which are displayed in Figure 16.1d. All PFEs of C are linear combinations of its simple PFEs.
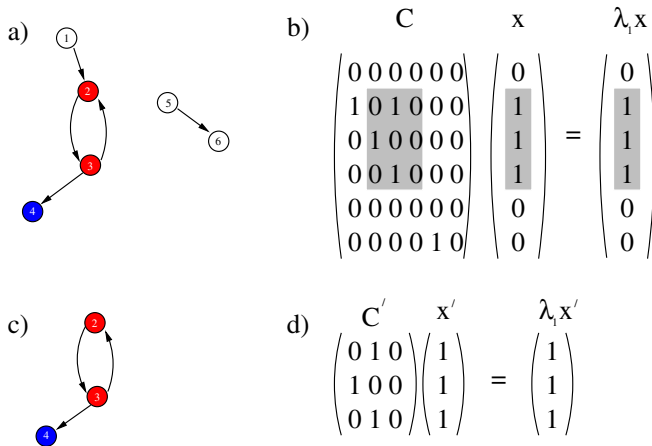
Figure 16.3: Example showing that the λ_1 of a PFE subgraph equals the λ_1 of the whole graph. **a.** A directed graph with 6 nodes. **b.** **x** is an eigenvector of its adjacency matrix C with eigenvalue $\lambda_1 = 1$, which is the Perron-Frobenius eigenvalue of the graph. The non zero components of **x** and the corresponding rows and columns of C are highlighted. **c.** The subgraph of the PFE **x**. **d.** The vector **x'** constructed by removing the zero components of **x** is an eigenvector of the adjacency matrix, C', of the PFE subgraph. Its corresponding eigenvalue is unity, which is also the Perron-Frobenius eigenvalue of the PFE subgraph.

Core and periphery of a simple PFE

If C' is the subgraph of a simple PFE, the basic subgraph of C contained in C' will be called the *core* of C' (or equivalently, the 'core of the simple PFE'), and denoted Q'. The set of the remaining nodes and links of C' that are not in its core will together be said to constitute the *periphery* of C'. For example, for the PFE in Figure 16.1c the core is the 2-cycle comprising nodes 14 and 15. Note that the periphery is not a subgraph in the sense we are using the word 'subgraph', since it contains links not just between periphery nodes but also from nodes outside the periphery (like the link from node 15 to 16 in Figure 16.1c).

The core and periphery can be shown to have the following topological property (which justifies the nomenclature):

Proposition 4. *From every node in the core of (the subgraph of) a simple PFE there exists a path leading to every other node of the PFE subgraph. From no periphery node is there any path leading to any core node.*
Thus all periphery nodes are downstream from all core nodes. Starting from the core one can reach the periphery but not vice versa.

It follows from the Perron-Frobenius theorem for irreducible graphs [13] that $\lambda_1(Q')$ will necessarily increase if any link is added to the core. Similarly removing any link will decrease $\lambda_1(Q')$. Thus λ_1 measures the multiplicity of internal pathways in the core. Figure 16.4 illustrates this point.

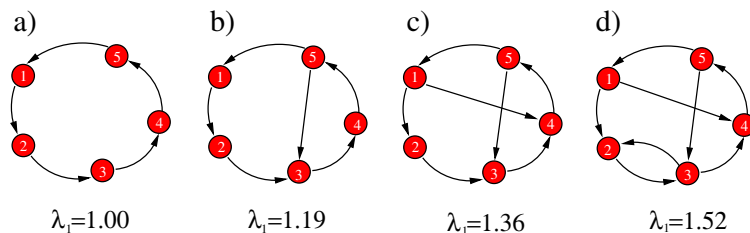**Figure 16.4:** λ_1 is a measure of the multiplicity of internal pathways in the core of simple PFE. Four irreducible graphs are shown. An irreducible graph always has a unique PFE that is simple and whose core is the entire graph. The Perron-Frobenius theorem ensures that adding a link to the core of a simple PFE necessarily increases its Perron-Frobenius eigenvalue λ_1. The figure also illustrates the concept of keystone nodes (see section 3).

Core and periphery of a non-simple PFE

Since any PFE of a graph can be written as a linear combination of a set of simple PFEs (this set is unique for any graph), the definitions of core and periphery can be readily extended to any PFE as follows:

The *core of a PFE*, denoted Q, is the union of the cores of those simple PFEs whose linear combination forms the given PFE. The rest of the nodes and links of the PFE subgraph constitute its periphery. It follows from the above discussion that $\lambda_1(Q) = \lambda_1(C)$. When the core is a union of disjoint cycles then $\lambda_1(Q) = 1$, and vice versa.

The structure of PFEs when there is no ACS

The above discussion about the structure of PFEs was for graphs C with $\lambda_1(C) > 0$. If $\lambda_1(C) = 0$, the graph has no ACS. Then the structure of PFEs is as follows: there exists a PFE for every connected component of the graph. Since there are no closed walks in the graph, all walks have finite lengths. Consider the longest paths in a given connected component. Identify the nodes that are the endpoints of these longest paths. The PFE corresponding to the given connected component will have $x_i > 0$ for each of the latter nodes and $x_i = 0$ for all other nodes in the graph. Again a general PFE is a linear combination of all such PFEs, one for each connected component of the graph. In this case since there is no closed path there is no core (or periphery) for any PFE of the graph. The core of all PFEs of such a graph may be defined to be the null set, $Q = \phi$.

16.3   A dynamical system on a fixed graph

In the previous section we have discussed the properties of graphs and their associated adjacency matrices, eigenvalues and eigenvectors. In this section we discuss the dynamical significance of the same constructs. In particular, we present an example of a dynamical system on

a fixed graph described by a set of coupled ordinary differential equations, whose attractors are precisely the PFEs discussed above. This dynamical system arises as an idealization of population dynamics of a set of chemicals.

Consider the simplex of normalized non-negative vectors in s dimensions: $J = \{\mathbf{x} \equiv (x_1, x_2, \ldots, x_s) \in \mathbf{R}^s | 0 \leq x_i \leq 1, \sum_{i=1}^{s} x_i = 1\}$. For a fixed graph $C = (c_{ij})$ with s nodes, consider the set of coupled ordinary differential equations [20]

$$\dot{x}_i = \sum_{i=1}^{s} c_{ij} x_j - x_i \sum_{j,k=1}^{s} c_{kj} x_j. \tag{16.1}$$

This will be the dynamical system of interest to us in this section.

Note that the dynamics preserves the normalization of $\mathbf{x}$, $\sum_{i=1}^{s} \dot{x}_i = 0$. For non-negative C it leaves the simplex J invariant. (For negative c_{ij}, additional conditions have to be added (see [21]) but we do not discuss that case here.)

The links of the graph represent the interactions between the variables x_i that live on the nodes. x_i could represent, for example, the relative population of the ith species in a population of s species, or the probability of the ith strategy among a group of s strategies in an evolutionary game, or the market share of the ith company among a set of competing companies, etc. It is useful to see how equation (16.1) arises in a population dynamic context.

Let $i \in \{1, \ldots, s\}$ denote a chemical (or molecular) species in a chemical reactor. Molecules can react with each other in various ways; we focus on only one aspect of their interactions: catalysis. The catalytic interactions can be described by a directed graph with s nodes. The nodes represent the s species and the existence of a link from node j to node i means that species j is a catalyst for the production of species i. In terms of the adjacency matrix, $C = \{c_{ij}\}$ of this graph, c_{ij} is set to unity if j is a catalyst of i and is set to zero otherwise. The operational meaning of catalysis is as follows:

Each species i will have an associated non-negative population y_i in the pond which changes with time. In a certain approximation (discussed below) the population dynamics for a fixed set of chemical species whose interactions are given by C, will be given by

$$\dot{y}_i = \sum_{j=1}^{s} c_{ij} y_j - \phi y_i, \tag{16.2}$$

where $\phi(t)$ is some function of time. To see how such an equation might arise, assume that species j catalyses the ligation of reactants A and B to form the species i, $A + B \xrightarrow{j} i$. Then the rate of growth of the population y_i of species i in a well stirred reactor will be given by $\dot{y}_i = k(1 + \nu y_j) n_A n_B - \phi y_i$, where n_A, n_B are reactant concentrations, k is the rate constant for the spontaneous reaction, ν is the catalytic efficiency, and ϕ represents a common death rate or dilution flux in the reactor. Assuming the catalysed reaction is much faster than the spontaneous reaction, and that the concentrations of the reactants are large and fixed, the rate equation becomes $\dot{y}_i = K y_j - \phi y_i$, where K is a constant. In general since species i can have multiple catalysts, we get $\dot{y}_i = \sum_{j=1}^{s} K_{ij} y_j - \phi y_i$, with $K_{ij} \sim c_{ij}$. We make the further idealization $K_{ij} = c_{ij}$ giving equation (16.2).

The relative population of species i is by definition $x_i \equiv y_i / \sum_{j=1}^{s} y_j$. Therefore $\mathbf{x} \equiv (x_1, \ldots, x_s) \in J$, since $0 \leq x_i \leq 1, \sum_{i=1}^{s} x_i = 1$. Taking the time derivative of x_i and

using (16.2) it is easy to see that $\dot{x}_i$ is given by (16.1). Note that the ϕ term, present in (16.2), cancels out and is absent in (16.1).

We remark that the quasispecies equation [17] has the same form as equation (16.2), albeit with a different interpretation and a special structure of the C matrix that arises from that interpretation.

16.3.1   Attractors of equation (16.1)

The rest of this section consists of examples and arguments to justify the

Proposition 5. *For any graph C,*

(i) *Every eigenvector of C that belongs to J is a fixed point of (16.1), and vice versa.*

(ii) *Starting from any initial condition in the simplex J, the trajectory converges to some fixed point (generically denoted $\mathbf{X}$) in J.*

(iii) *For generic initial conditions in J, $\mathbf{X}$ is a Perron-Frobenius eigenvector (PFE) of C.* (For special initial conditions, forming a space of measure zero in J, $\mathbf{X}$ could be some other eigenvector of C. Henceforth we ignore such special initial conditions.)

(iv) *If C has a unique PFE, $\mathbf{X}$ is the unique stable attractor of (16.1).*

(v) *If C has more than one linearly independent PFE, then $\mathbf{X}$ can depend upon the initial conditions. The set of allowed $\mathbf{X}$ is a convex linear combination of a subset of the PFEs.* The interior of this convex set in J may then be said to be the 'attractor' of (16.1), in the sense that for generic initial conditions all trajectories converge to a point in this set.

(vi) *For every $\mathbf{X}$ belonging to the attractor set, the set of nodes i for which $X_i > 0$ is the same and is uniquely determined by C. The subgraph formed by this set of nodes will be called the 'subgraph of the attractor' of (16.1) for the graph C.* Physically, this set consists of nodes that always end up with a nonzero relative population when the dynamics (16.1) is allowed to run its course, starting from generic initial conditions.

(vii) *If $\lambda_1(C) > 0$, the subgraph of the attractor of (16.1) is an ACS.* This ACS will be called the *dominant ACS* of the graph. The dominant ACS is independent of (generic) initial conditions and depends only on C.

For example for the graph of Figure 16.1a, $\mathbf{X}$ is a convex linear combination of $\mathbf{e}_2$ and $\mathbf{e}_3$, $\mathbf{X} = a\mathbf{e}_2 + (1-a)\mathbf{e}_3$, with $0 \le a \le 1$. a depends upon initial conditions; generically $0 < a < 1$. The subgraph of the attractor contains eight nodes, 6,7,14-19. Starting with generic initial conditions where all the x_i are nonzero, the trajectory will converge to a point $\mathbf{X}$ where these eight nodes have nonzero X_i and each of the other twelve nodes have $X_i = 0$. The eight populated nodes form an ACS, the dominant ACS of the graph.

To see (i), let $\mathbf{x}^\lambda \in J$ be an eigenvector of C, $\sum_j c_{ij} x_j = \lambda x_i$. Substituting this on the r.h.s. of (16.1), one gets zero. Conversely, if the r.h.s. of (16.1) is zero, one finds $\mathbf{x} = \mathbf{x}^\lambda$, with $\lambda = \sum_{k,j} c_{kj} x_j$.

To motivate (ii) and (iii) it is most convenient to consider the underlying dynamics (16.2) from which (16.1) is derived: Since (16.1) is independent of ϕ, we can set $\phi = 0$ in (16.2) without any loss of generality. With $\phi = 0$ the general solution of (16.2), which is a linear system, can be schematically written as:

$$\mathbf{y}(t) = e^{Ct}\mathbf{y}(0),$$

where $\mathbf{y}(0)$ and $\mathbf{y}(t)$ are viewed as column vectors. Suppose $\mathbf{y}(0)$ is a right eigenvector of C

with eigenvalue λ, denoted $\mathbf{y}^\lambda$. Then

$$\mathbf{y}(t) = e^{\lambda t}\mathbf{y}^\lambda.$$

Since this time dependence is merely a rescaling of the eigenvector, this is an alternative way of seeing that $\mathbf{x}^\lambda = \mathbf{y}^\lambda / \sum_{j=1}^s y_j^\lambda$ is a fixed point of (16.1). If the eigenvectors of C form a basis in R^s, $\mathbf{y}(0)$ is a linear combination: $\mathbf{y}(0) = \sum_\lambda a_\lambda \mathbf{y}^\lambda$. In that case, for large t it is clear that the term with the largest value of λ will win out, hence

$$\mathbf{y}(t) \overset{t\to\infty}{\sim} e^{\lambda_1 t}\mathbf{y}^{\lambda_1}$$

where λ_1 is the eigenvalue of C with the largest real part (which we know is the same as its Perron-Frobenius eigenvalue) and $\mathbf{y}^{\lambda_1}$ an associated eigenvector. Therefore, for generic initial conditions the trajectory of (16.1) will converge to $\mathbf{X} = \mathbf{x}^{\lambda_1}$, a PFE of C. If the eigenvectors of C do not form a basis in R^s, the above result is still true (as we will see in examples).

Note that λ_1 can be interpreted as the 'population growth rate' at large t, since $\dot{\mathbf{y}}(t) \overset{t\to\infty}{\sim} \lambda_1 \mathbf{y}$. In the previous section we had mentioned that λ_1 measures a topological property of the graph, namely, the multiplicity of internal pathways in the core of the graph. Thus in the present model, λ_1 has both a topological and dynamical significance, which relates two distinct properties of the system, one structural (multiplicity of pathways in the core of the graph), and the other dynamical (population growth rate). The higher the multiplicity of pathways in the core, the greater is the population growth rate of the dominant ACS.

Part (iv) follows from the above. We will give examples as illustrations of (v) and (vi). Further, from Proposition 3, previous section, we know that the subgraph of a PFE has to be an ACS, whenever $\lambda_1 > 0$. That explains (vii). It is instructive to consider examples of graphs and see how the trajectory converges to a PFE.

Example 1. A simple chain, Figure 16.5a:
The adjacency matrix of this graph has all eigenvalues (including λ_1) zero. There is only one (normalized) eigenvector corresponding to this eigenvalue, namely $\mathbf{e} = (0,0,1)$ and this is the unique PFE of the graph. (This is an example where the eigenvectors of C do not form a basis in R^s.)   Since node 1 has no catalyst, its rate equation is (henceforth taking $\phi = 0$) $\dot{y}_1 = 0$. Therefore $y_1(t) = y_1(0)$, a constant. The rate equation for node 2 is $\dot{y}_2 = y_1 = y_1(0)$. Thus $y_2(t) = y_2(0) + y_1(0)t$. Similarly $\dot{y}_3 = y_2$ implies that $y_3(t) = (1/2)y_1(0)t^2 + y_2(0)t + y_3(0)$. At large t, $y_1 = $ constant, $y_2 \sim t$, $y_3 \sim t^2$; hence y_3 dominates. Therefore, $X_i = \lim_{t\to\infty} x_i(t)$ is given by $X_1 = 0, X_2 = 0, X_3 = 1$. Thus we find that $\mathbf{X}$ equals the unique PFE $\mathbf{e}$, independent of initial conditions.

Example 2. A 1-cycle, Figure 16.5b:
This graph has two eigenvalues, $\lambda_1 = 1, \lambda_2 = 0$. The unique PFE is $\mathbf{e} = (1,0)$. The rate equations are $\dot{y}_1 = y_1, \dot{y}_2 = 0$, with the solutions $y_1(t) = y_1(0)e^t, y_2(t) = y_2(0)$. At large t node 1 dominates, hence $\mathbf{X} = (1,0) = \mathbf{e}$. The exponentially growing population of 1 is a consequence of the fact that 1 is a self-replicator, as embodied in the equation $\dot{y}_1 = y_1$.

Example 3. A 2-cycle, Figure 16.5c:
The corresponding adjacency matrix has eigenvalues $\lambda_1 = 1, \lambda_2 = -1$. The unique normal-
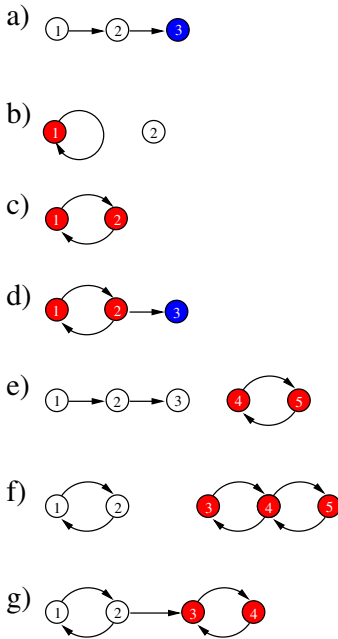
a)

b)

c)

d)

e)

f)

g)

Figure 16.5: Examples of graphs with a unique PFE. The subgraph of the PFE coincides with the nodes that are populated in the attractor.

ized PFE is $\mathbf{e} = (1/2, 1/2)$. The population dynamics equations are $\dot{y}_1 = y_2, \dot{y}_2 = y_1$. The general solution to these is (note $\ddot{y}_1 = y_1$)

$$y_1(t) = Ae^t + Be^{-t}, \qquad y_2(t) = Ae^t - Be^{-t}.$$

Therefore at large t, $y_1 \to Ae^t, y_2 \to Ae^t$, hence $\mathbf{X} = (1,1)/2 = \mathbf{e}$. Neither 1 nor 2 is individually a self-replicating species, but collectively they function as a self-replicating entity. This is true of all ACSs.

Example 4. A 2-cycle with a periphery, Figure 16.5d:
This graph has $\lambda_1 = 1$ and a unique normalized PFE $\mathbf{e} = (1,1,1)/3$. The population equations for y_1 and y_2 and consequently their general solutions are the same as Example 3, but now in addition $\dot{y}_3 = y_2$, yielding $y_3(t) = Ae^t + Be^{-t} + \text{constant}$. Again for large t, y_1, y_2, y_3 grow as $\sim Ae^t$, hence $\mathbf{X} = (1,1,1)/3 = \mathbf{e}$. The dominant ACS includes all the three nodes.

   This example shows how a parasitic periphery (which does not feed back into the core) is supported by an autocatalytic core. This is also an example of the following general result: when a subgraph C', with largest eigenvalue λ_1', is *downstream* from another subgraph C'' with largest eigenvalue $\lambda_1'' > \lambda_1'$, then the population of the former also increases at the rate λ_1''. Therefore if C'' is populated in the attractor, so is C'. In this example C' is the single node 3 with $\lambda_1' = 0$ and C'' is the 2-cycle of nodes 1 and 2 with $\lambda_1'' = 1$.

Example 5. A 2-cycle and a chain, Figure 16.5e:
The graph in Figure 16.5e combines the graphs of Figures 16.5a and c. Following the analysis of those two examples it is evident that for large t, $y_1 \sim t^0, y_2 \sim t^1, y_3 \sim t^2, y_4 \sim e^t, y_5 \sim e^t$. Because the populations of the 2-cycle are growing exponentially they will eventually completely overshadow the populations of the chain which are growing only as powers of t. Therefore the attractor will be $\mathbf{X} = (0, 0, 0, 1, 1)/2$ which, it can be checked, is a PFE of the graph (it is an eigenvector with eigenvalue 1).

In general when a graph consists of one or more ACSs and other nodes that are not part of any ACS, the populations of the ACS nodes grow exponentially while the populations of the latter nodes grow at best as powers of t. Hence ACSs always outperform non-ACS structures in the population dynamics (see also Example 2). This is a consequence of the infinite walks provided by the positive feedback inherent in the ACS structure, while non-ACS structures have no feedbacks and only finite walks.

Example 6. A 2-cycle and another irreducible graph disconnected from it, Figure 16.5f:
One can ask, when there is more than one ACS in the graph, which is the dominant ACS? Figure 16.5f shows a graph containing two ACSs. The 2-cycle subgraph has a Perron-Frobenius eigenvalue 1, while the other irreducible subgraph has a Perron-Frobenius eigenvalue $\sqrt{2}$. The unique PFE of the entire graph is $\mathbf{e} = (0, 0, 1, \sqrt{2}, 1)/(2 + \sqrt{2})$ with eigenvalue $\sqrt{2}$. The population dynamics equations are $\dot{y}_1 = y_2, \dot{y}_2 = y_1, \dot{y}_3 = y_4, \dot{y}_4 = y_3 + y_5, \dot{y}_5 = y_4$. The first two equations are completely decoupled from the last three and the solutions for y_1 and y_2 are the same as for Example 3. For the other irreducible graph the solution is (since $\ddot{y}_4 = \dot{y}_3 + \dot{y}_5 = 2y_4$)

$$y_4(t) = Ae^{\sqrt{2}t} + Be^{-\sqrt{2}t}, \qquad y_3(t) = \frac{1}{\sqrt{2}}(Ae^{\sqrt{2}t} + Be^{-\sqrt{2}t}) + C,$$

$$y_5(t) = \frac{1}{\sqrt{2}}(Ae^{\sqrt{2}t} + Be^{-\sqrt{2}t}) - C.$$

Thus, the populations of nodes 3,4 and 5 also grow exponentially but at a faster rate, reflecting the higher Perron-Frobenius eigenvalue of the subgraph comprising those nodes. Therefore this structure eventually overshadows the 2-cycle, and the attractor is $\mathbf{X} = \mathbf{e}$. The dominant ACS in this case is the irreducible subgraph formed by nodes 3,4 and 5.

More generally, when a graph consists of several disconnected ACSs with different individual λ_1, only the ACSs whose λ_1 is the largest (and equal to $\lambda_1(C)$) end up with non-zero relative populations in the attractor.

Example 7. A 2-cycle downstream from another 2-cycle, Figure 16.5g:
What happens when the graph contains two ACSs whose individual λ_1 equals $\lambda_1(C)$, and one of those ACSs is downstream of another? In Figure 16.5g nodes 3 and 4 form a 2-cycle which is downstream from another 2-cycle comprising nodes 1 and 2. The unique PFE of this graph, with $\lambda_1 = 1$, is $\mathbf{e} = (0, 0, 1, 1)/2$. The population dynamics equations are $\dot{y}_1 = y_2, \dot{y}_2 = y_1, \dot{y}_3 = y_4 + y_2, \dot{y}_4 = y_3$. Their general solution is:

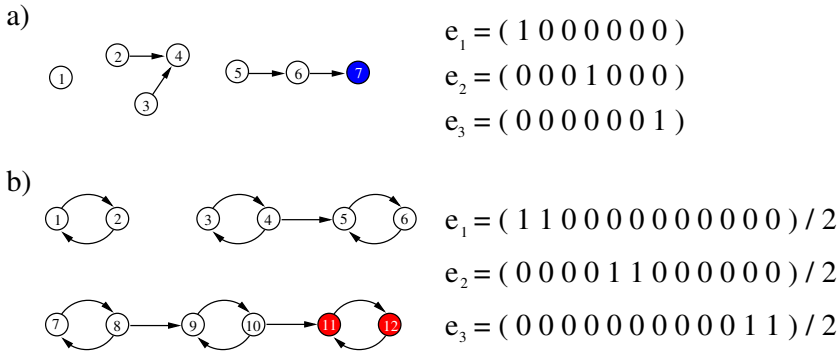$$y_1(t) = Ae^t + Be^{-t}, \qquad y_2(t) = Ae^t - Be^{-t},$$

a)

$$e_1 = (\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ )$$

$$e_2 = (\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ )$$

$$e_3 = (\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ )$$

b)

$$e_1 = (\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ )\ /\ 2$$

$$e_2 = (\ 0\ 0\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ )\ /\ 2$$

$$e_3 = (\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ )\ /\ 2$$

Figure 16.6: Examples of graphs with multiple PFEs. (a) e_1, e_2, e_3 are all eigenvectors with eigenvalue $\lambda_1 = 0$. Only e_3 is the attractor. Thus for generic initial conditions, only node 7, which sits at the end point of the longest chain of nodes is populated in the attractor. (b) e_1, e_2, e_3 are all eigenvectors with eigenvalue $\lambda_1 = 1$, but only e_3 is the attractor. Only the 2-cycle of nodes 11 and 12, which sits at the end of the longest chain of cycles, is populated in the attractor.

$$y_3(t) = \frac{t}{2}(Ae^t - Be^{-t}) + Ce^t + De^{-t},$$

$$y_4(t) = \frac{t}{2}(Ae^t + Be^{-t}) + (C - \frac{A}{2})e^t + (\frac{B}{2} - D)e^{-t}.$$

It is clear that for large t, $y_1 \sim e^t, y_2 \sim e^t, y_3 \sim te^t, y_4 \sim te^t$. While all four grow exponentially with the same rate λ_1, as $t \to \infty$ y_3 and y_4 will overshadow y_1 and y_2. The attractor will be therefore be $\mathbf{X} = (0, 0, 1, 1)/2 = \mathbf{e}$. Here the dominant ACS is the 2-cycle of nodes 3 and 4. This result generalizes to other kinds of ACSs: if one irreducible subgraph is downstream of another with the same Perron-Frobenius eigenvalue, the latter will have zero relative population in the attractor.

The above examples displayed graphs with a unique PFE, and illustrated Proposition 5 (iv). The stability of the global attractor follows from the fact that the constants A, B, C, D, etc., in the above examples, which can be traded for the initial conditions of the populations, appear nowhere in the attractor configuration $\mathbf{X}$. Now we consider examples where the PFE is not unique.

Example 8. Graph with $\lambda_1 = 0$ and three disconnected components, Figure 16.6a:
As mentioned in section 2 this graph has three independent PFEs, displayed in Figure 16.6a. The attractor is $\mathbf{X} = e_3$. This is an immediate generalization of Example 1 above. Using the same argument as for Example 1, we can see that $y_i \sim t^k$ if the longest path ending at node i is of length k. Therefore the attractor will have nonzero components only for nodes at the ends of the longest paths. Thus the populations of nodes 1,2,3 and 5 are constant, those of 4 and 6 increase $\sim t$ for large t, and of 7 as $\sim t^2$, explaining the result.

Example 9. Several connected components containing 2-cycles, Figure 16.6b:
Here again there are three PFEs, one for each connected component. The population of nodes in 2-cycles which are not downstream of other 2-cycles (nodes 1,2,3,4,7 and 8) will grow as e^t. As in Example 7, Figure 16.5g, the nodes of 2-cycles which are downstream of one 2-cycle (nodes 5,6,9 and 10) will grow as te^t. It can be verified that the populations of nodes in 2-cycles downstream from two other 2-cycles (nodes 11 and 12) will grow as t^2e^t. The pattern is clear: in the attractor only the 2-cycles at the ends of the longest chains of 2-cycles will have non-zero relative populations, explaining the result.

Example 10. Figure 16.1a:
From previous examples it is evident how the populations will change with time for Figure 16.1a. Here we list the result:

$$y_8 \sim t^0, \qquad y_9 \sim t^1, \qquad y_{10} \sim t^2,$$

$$y_1, y_2, y_3, y_4, y_5, y_{11}, y_{12}, y_{13}, y_{20} \sim e^t,$$

$$y_6, y_7, y_{14}, y_{15}, y_{16}, y_{17}, y_{18}, y_{19} \sim te^t.$$

Thus, starting from a generic initial population, only the eight nodes, 6,7,14-19, will be populated in the attractor. This explains the comments just after the statment of Proposition 5.

Note the structure of the dominant ACS in the above examples when $\lambda_1 > 0$. If there is a unique PFE in the graph, the dominant ACS is the subgraph of the PFE. If there are several PFEs only a subset of those may be counted as illustrated in Examples 9 and 10, Figures 16.6b and 16.1a, respectively. A general construction of the dominant ACS for an arbitrary graph will be described elsewhere.

How long does it take to reach the attractor?

The timescale over which the system reaches its attractor depends on the structure of the graph C. For instance in Example 2, the attractor is approached as the population of node 1, y_1, overwhelms the population y_2. Since y_1 grows exponentially as e^t, the attractor is reached on a timescale $\lambda_1^{-1} = 1$. In contrast, in Example 1, the attractor is approached as y_3 overwhelms y_1 and y_2. Because in this case all the populations are growing as powers of t, the timescale for reaching the attractor is infinite. In general, this timescale depends on the difference in growth rate between the fastest growing population and the next fastest growing population.

For graphs which have no basic subgraphs, i.e., graphs with $\lambda_1 = 0$ like those in Example 1 and 8, all populations grow as powers of t, hence the timescale for reaching the attractor is infinite.

For graphs which have one or more basic subgraphs (i.e., $\lambda_1 \geq 1$) but all the basic subgraphs are in different connected components, such as Examples 2-6, the timescale for reaching the attractor is given by $(\lambda_1 - \mathrm{Re}\lambda_2)^{-1}$, where λ_2 is the eigenvalue of C with the next largest real part, compared to λ_1.

For graphs having one or more basic subgraphs with at least one basic subgraph downstream from another basic subgraph, the ratio of the fastest growing population to the next fastest growing will always be a power of t (as in Examples 7, 9 and 10) therefore the timescale for reaching the attractor is again infinite.

Core and periphery of a graph

Since the dominant ACS is given by a PFE, we will define the core of the dominant ACS to be the core of the corresponding PFE. If the PFE is simple, the core of the dominant ACS consists of just one basic subgraph. If the PFE is non-simple the core of the dominant ACS will be a union of some basic subgraphs. Further, the dominant ACS is uniquely determined by the graph. This motivates the definition of the core and periphery of a graph: The *core* of a graph C, denoted $Q(C)$, is the core of the dominant ACS of C. The *periphery* of C is the periphery of the dominant ACS of C. This definition applies when $\lambda_1(C) > 0$. When $\lambda_1(C) = 0$, the graph has no ACS and by definition $Q(C) = \phi$. In all cases $\lambda_1(Q(C)) = \lambda_1(C)$. For all the graphs depicted in this paper, except the one in Figure 16.1e, the red nodes constitute the core of the graph, the blue nodes its periphery, and the white nodes are neither core nor periphery – they are nodes that are not in any of the PFE subgraphs. [1]

Core overlap of two graphs

Given any two graphs C and C' whose nodes are labeled, the *core overlap* between them, denoted $Ov(C, C')$, is the number of common links in the cores of C and C', i.e., the number of ordered pairs (j, i) for which Q_{ij} and Q'_{ij} are both non-zero [22]. If either of C or C' does not have a core, $Ov(C, C')$ is identically zero.

Keystone nodes

In ecology certain species are referred to as keystone species – those whose extinction or removal would seriously disturb the balance of the ecosystem [24–27]. One might similarly ask for the notion of a keystone node in a directed graph that captures some important organizational role played by a node. Consider the impact of the hypothetical removal of any node i from a graph C. One can, for example, ask for the core of the graph $C - i$ that would result if node i (along with all its links) were removed from C. We will refer to a node i as a *keystone node* if C has a non-vanishing core and $Ov(C, C - i) = 0$ [23]. Thus a keystone node is one whose removal modifies the organizational structure of the graph (as represented by its core) drastically. In each of Figures 16.4a–d, for example, the core is the entire graph. In Figure 16.4a, all the nodes are keystone, since the removal of any one of them would leave the graph without an ACS (and hence without a core). In general when the core of a graph is a single n-cycle, for any n, all the core nodes are keystone. In Figure 16.4b, nodes 3, 4 and 5 are keystone but the other nodes are not, and in Figure 16.4c only nodes 4 and 5 are keystone. In Figure 16.4d, there are no keystone nodes. These examples show that the more internal pathways a core has (generally, this implies a higher value of λ_1), the less likely it is to have keystone species, and hence the more robust its structure is to removal of nodes.

Figure 16.7 illustrates another type of graph structure which has a keystone node. The graph in Figure 16.7a consists of a 2-cycle (nodes 4 and 5) downstream from an irreducible subgraph consisting of nodes 1,2 and 3. The core of this graph is the latter irreducible subgraph. Figure 16.7b shows the graph that results if node 3 is removed with all its links. This

---

[1] The definition of the core of a graph given in refs. [22, 23] is a special case of this definition, holding only for graphs where each connected component of the dominant ACS has no more than one basic subgraph.
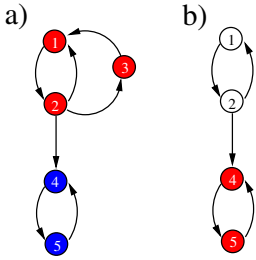
Figure 16.7: Example illustrating the notion of keystone species and the phenomenon of a core-shift. Node number 3 is keystone node of the graph in part **a** because its removal produces the graph in **b** which has a zero core overlap with the graph in **a**. The core nodes of both graph are coloured red. An event in which the core before the event and after the event have zero overlap is called a 'core-shift'.

consists of one 2-cycle downstream from another. Though both 2-cycles are basic subgraphs of the graph, as discussed in Example 7, Figure 16.5g, this graph has a unique (upto constant multiples) PFE, whose subgraph consists of the downstream cycle (nodes 4 and 5) only. Thus the 2-cycle 4-5 is the core of the graph in Figure 16.7b. Clearly $Ov(C, C-3) = 0$ therefore node 3 in Figure 16.7a is a keystone node.

We remark that the above purely graph theoretic definition of a keystone node turns out to be useful in the dynamical system discussed in this and the following sections. For other dynamical systems, other definitions of keystone might be more useful.

16.4 Graph dynamics

So far we have discussed the algebraic properties of a fixed graph, and the attractors of a particular dynamical system on arbitrary, but fixed graphs. However one of the most interesting properties of complex systems is that the graph of interactions among their components evolves with time, resulting in many interesting adaptive phenomena. We now turn to such an example, where the graph itself is a dynamical variable, and display how phenomena such as self-organization, catastrophes, innovation, etc, can arise. We shall see that the above discussion of (static) graph theory will be crucial in understanding these phenomena.

We consider a process which alters a graph in discrete steps. The series of graphs produced by such a process can be denoted $C_n, n = 1, 2, \ldots$. A graph update event will be one step of the process, taking a graph from C_{n-1} to C_n. In fact the process we consider is a specific example of a Markov process on the space of graphs. At time $n-1$, the graph C_{n-1} determines the transition probability to all other graphs. The stochastic process picks the new graph C_n using this probability distribution and the trajectory moves forward in graph space. In the example we consider, the transition probability is not specified explicitly. It arises implicitly as a consequence of the dynamics (16.1) that takes place on a fast time scale for the fixed graph C_{n-1}.

The graph dynamics is implemented as follows [20]:
Initially the graph is random: for every ordered pair (i, j) with $i \neq j$, c_{ij} is independently chosen to be unity with a probability p and zero with a probability $1 - p$. c_{ii} is set to zero for all i. Each x_i is chosen randomly in $[0, 1]$ and all x_i are rescaled so that $\sum_{i=1}^{s} x_i = 1$.

Step 1. With C fixed, $\mathbf{x}$ is evolved according to (16.1) until it converges to a fixed point, denoted $\mathbf{X}$. The set $\mathcal{L}$ of nodes with the least X_i is determined, i.e, $\mathcal{L} = \{i \in S | X_i = \min_{j \in S} X_j\}$.

Step 2. A node, say node k, is picked randomly from $\mathcal{L}$ and is removed from the graph along with all its links.

Step 3. A new node (also denoted k) is added to the graph. Links to and from k to other nodes are assigned randomly according to the same rule, i.e, for every $i \neq k$ c_{ik} and c_{ki} are independently reassigned to unity with probability p and zero with probability $1 - p$, irrespective of their earlier values, and c_{kk} is set to zero. All other matrix elements of C remain unchanged. x_k is set to a small constant x_0, all other x_i are perturbed by a small amount from their existing value X_i, and all x_i are rescaled so that $\sum_{i=1}^{s} x_i = 1$.

This process, from step 1 onwards, is iterated many times.

Notice that the population dynamics and the graph dynamics are coupled: the evolution of the x_i depends on the graph C in step 1, and the evolution of C in turn depends on the x_i through the choice of which node to remove in step 2. There are two timescales in the dynamics, a short timescale over which the graph is fixed while the x_i evolve, and a longer timescale over which the graph is changed.

This dynamics is motivated by the origin of life problem, in particular the puzzle of how a complex chemical organization might have emerged from an initial 'random soup' of chemicals, as discussed in section 1. Let us consider a pond on the prebiotic earth containing s molecular species which interact catalytically as discussed in the previous section, and let us allow the chemical organization to evolve with time due to various natural process which remove species from the pond and bring new species into the pond. Thus over short timescales we let the populations of the species evolve according to (16.1). Over longer timescales we imagine the prebiotic pond to be subject to periodic perturbations from storms, tides or floods. These perturbations remove existing species from the pond and introduce new species into it. The species most likely to be completely removed from the pond are those that have the least number of molecules. The new species could have entirely different catalytic properties from those removed or those existing in the pond. The above rules make the idealization that the perturbation eliminates exactly one existing species (that has the least relative population) and brings in one new species. The behaviour of the system does not depend crucially on this assumption [23].

While in previous sections we have considered graphs with 1-cycles, the requirement $c_{ii} = 0$ in the present section forbids 1-cycles in the graph. The motivation is the following: 1-cycles represent self-replicating species (see previous section, Example 2). Such species, e.g., RNA molecules, are difficult to produce and maintain in a prebiotic scenario and it is generally believed that it requires a complex self supporting molecular organization to be in place *before*

an RNA world, for example, can take off [28, 29]. Thus, we wish to address the question: can we get complex molecular organizations without putting in self-replicating species by hand in the model? As we shall see below, this does indeed happen, since even though self-replicating individual species are disallowed, collectively self-replicating autocatalytic sets can still arise by chance on a certain time scale, and when they do, they trigger a wave of self-organization in the system.

The rules for changing the graph implement *selection* and *novelty*, two important features of natural evolution. Selection is implemented by removing the species which is 'performing the worst', with 'performance' in this case being equated to a species' relative population (step 2). Adding a new species introduces novelty into the system. Note that although the actual connections of a new node with other nodes are created randomly, the new node has the same average connectivity as the initial set of nodes. Thus the new species is not biased in any way towards increasing the complexity of the chemical organization. Step 2 and step 3 represent the interaction of the system with the external environment. The third feature of the model is dynamics of the system that depends upon the interaction among its components (step 1). The phenomena to be described in the following sections are all consequences of the interplay between these three elements – selection, novelty and an internal dynamics.

16.5   Self Organization

We now discuss the results of graph evolution. Figure 16.8 shows the total number of links in the graph versus time (n, the number of graph updates). Three runs of the model described in the previous section, each with $s = 100$ and different values of p are exhibited. Also exhibited is a run where there was *no selection* (in which step 2 is modified: instead of picking one of the nodes of $\mathcal{L}$, any one of the s nodes is picked randomly and removed from the graph along with all its links. The rest of the procedure remains the same). Figure 16.9 shows the time evolution of two more quantities for the same three runs with selection displayed in Figure 16.8.

The quantities plotted are the number of nodes with $X_i > 0$, s_1, and the Perron-Frobenius eigenvalue of the graph, λ_1. The values of the parameters p and s for the displayed runs were chosen to lie in the regime $ps < 1$. Much of the analytical work described below, such as estimation of various timescales, assumes that $ps \ll 1$. Figure 16.10 shows snapshots of the graph at various times in the run shown in Figure 16.9b, which has $p = 0.0025$. It is clear that without selection each graph update replaces a randomly chosen node with another which has on average the same connectivity. Therefore the graph remains random like the starting graph and the number of links fluctuates about its random graph value $\approx ps^2$. As soon as selection is turned on the behaviour becomes more interesting. Three regimes can be observed. First, the 'random phase' where the number of links fluctuates around ps^2 and s_1 is small. Second, the 'growth phase' where l and s_1 show a clear rising tendency. Finally, the 'organized phase' where l again hovers (with large fluctuations) about a value much higher than the initial random graph value, and s_1 fluctuates (again with large fluctuations) about its maximum value s. The time spent in each phase clearly depends on p, and we find it also depends on s. This behaviour can be understood by taking a look at the structure of the graph in each of these phases, especially the ACS structure, and using the results of sections 2 and 3.
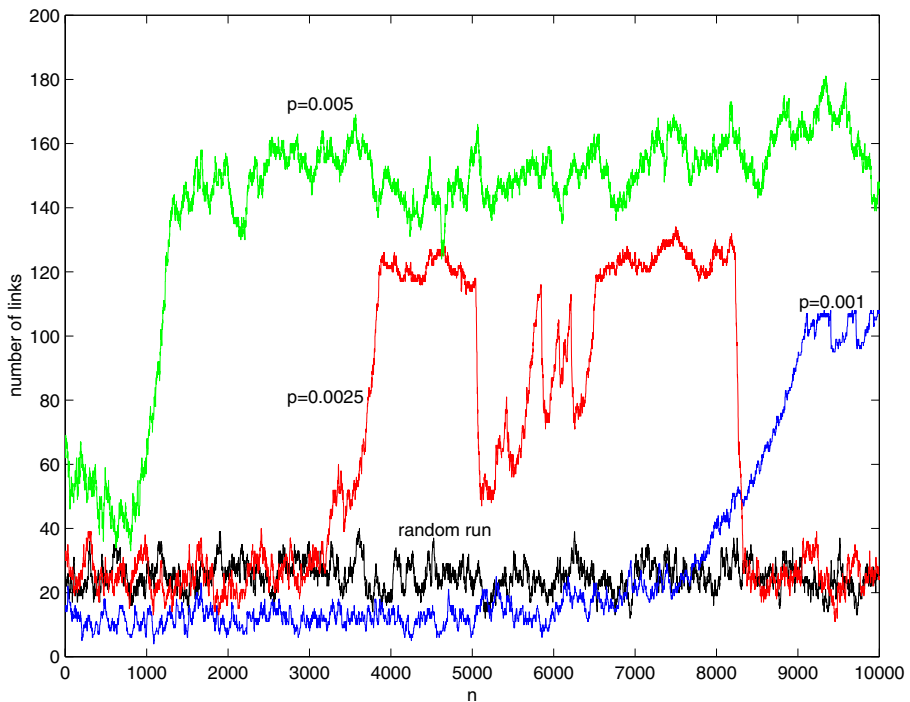
Figure 16.8: The number of links versus time (n) for various runs. Each run had $s = 100$. The black curve is a run with selection turned off; a random node is picked for removal at each graph update. The other curves show runs with selection turned on and with different p values: Blue $p = 0.001$, Red $p = 0.0025$, Green $p = 0.005$.

16.5.1   The random phase

Initially, the random graph contains no cycles, and hence no ACSs, and its Perron-Frobenius eigenvalue is $\lambda_1 = 0$. We have seen in section 3 that for such a graph the attractor will have nonzero components for all nodes which are at the ends of the longest paths of nodes, and zero for every other node. (In Figure 16.10a, there are two paths of length 4, which are the longest paths in the graph. Both end at node 13, which is therefore the only populated node in the attractor for this graph.) These nodes, then, are the only nodes protected from elimination during the graph update. However, these nodes have high relative populations *because they are supported by other nodes*, while the latter (supporting) nodes do not have high relative populations. Inevitably within a few graph updates a supporting node will be removed from the graph. When that happens a node which presently has nonzero X_i will no longer be at the end of the longest path and hence will get $X_i = 0$. For example node 34, which belongs to $\mathcal{L}$, is expected to be picked for replacement within $\approx O(s)$ graph update time steps. In fact it is replaced in the 8th time step. After that node 13 becomes a singleton and joins the set $\mathcal{L}$. Thus no structure is stable when there is no ACS. Eventually, all nodes are removed and replaced, and the graph remains random.
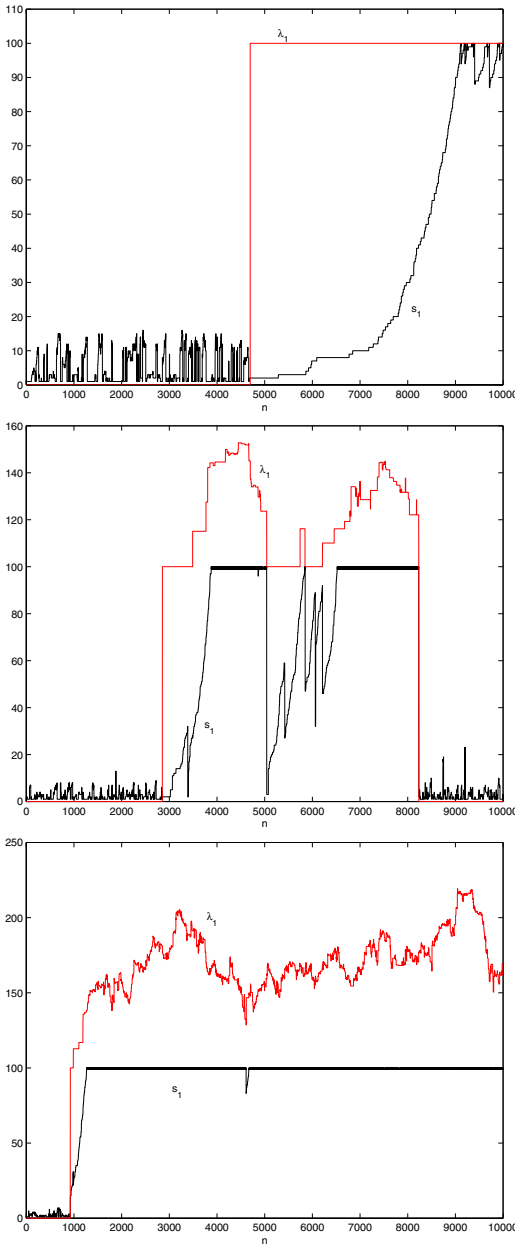
Figure 16.9: Number of populated nodes, s_1, (black curve) and the Perron-Frobenius eigenvalue of the graph, λ_1, (red curve) versus time, n, for the same three runs shown in Figure 16.8. Each run has $s = 100$ and $p = 0.001, 0.0025$ and 0.05 respectively. The λ_1 values shown are 100 times the actual value.
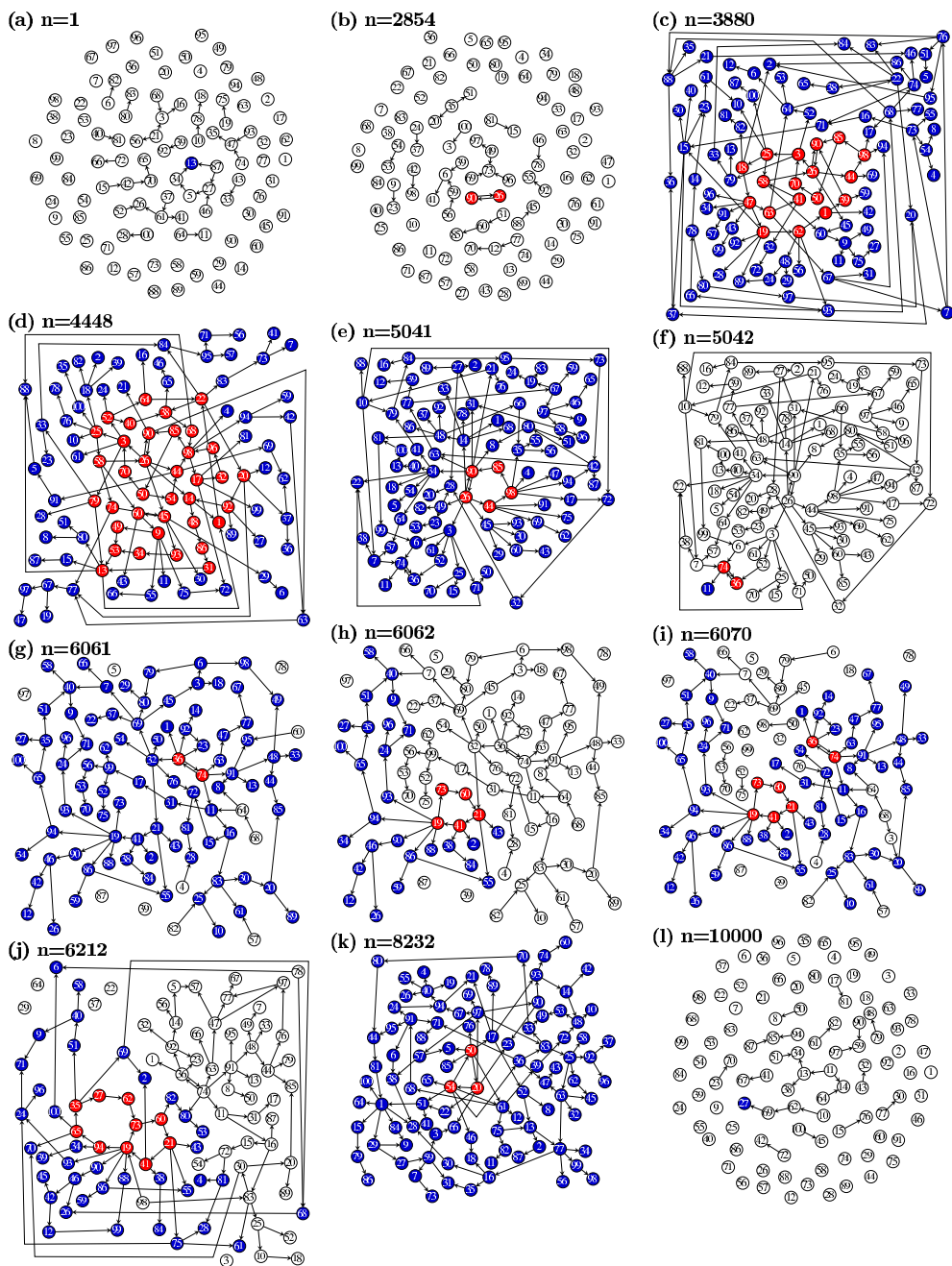
(a) n=1

(b) n=2854

(c) n=3880

(d) n=4448

(e) n=5041

(f) n=5042

(g) n=6061

(h) n=6062

(i) n=6070

(j) n=6212

(k) n=8232

(l) n=10000

Figure 16.10: Snapshots of the graph at various times for the run shown in Figure 16.9b with $s = 100$ and $p = 0.0025$. See text for a description of the major events. In all graphs, white nodes are those with $X_i = 0$. All coloured nodes have $X_i > 0$. In graphs which have an ACS, the red nodes are core nodes and the blue nodes are periphery nodes.

Note that the inital random graph is likely to contain no cycles when p is small ($ps \ll 1$). If larger values of p are chosen, it becomes more likely that the initial graph will contain a cycle. If it does, there is no random phase; the system is then in the growth phase, discussed below, right from the initial time step.

16.5.2   The growth phase

At some graph update an ACS is formed by pure chance. The probability of this happening can be closely approximated by the probability of a 2-cycle (the simplest ACS) forming by chance, which is $p^2 s$ (= the probability that in the row and column corresponding to the replaced node in C, any matrix element and its transpose both turn out to be unity). Thus the average time of appearance of an ACS is $1/p^2 s$. In the run whose snapshots are displayed in Figure 16.10, a 2-cycle between nodes 26 and 90 formed at $n = 2854$. This is a graph which consists of a 2-cycle and several other chains and trees. For such a graph we have shown in Example 3 in section 3 that the attractor has non-zero X_i for nodes 26 and 90 and zero for all other nodes. The dominant ACS consists of nodes 26 and 90. Therefore these nodes cannot be picked for removal at the graph update and hence a graph update cannot destroy the links that make the dominant ACS. *The autocatalytic property is guaranteed to be preserved until the dominant ACS spans the whole graph.*

When a new node is added to the graph at a graph update, one of three things will happen:

1. The new node will not have any links from the dominant ACS and will not form a new ACS. In this case the dominant ACS will remain unchanged, the new node will have zero relative population and will be part of the least fit set. For small p this is the most likely possibility.

2. The new node gets an incoming link from the dominant ACS and hence becomes a part of it. In this case the dominant ACS grows to include the new node. For small p, this is less likely than the first possibility, but such events do happen and in fact are the ones responsible for the growth of complexity and structure in the graph.

3. The new node forms another ACS. This new ACS competes with the existing dominant ACS. Whether it now becomes dominant, overshadowing the previous dominant ACS or it gets overshadowed, or both ACSs coexist depends on the Perron Frobenius eigenvalues of their respective subgraphs and whether (and which) ACS is downstream of the other. It can be shown that this is a rare event compared with possibilities 1 and 2.

Typically the dominant ACS keeps growing by accreting new nodes, usually one at a time, until the entire graph is an ACS. At this point the growth phase stops and the organized phase begins. As a consequence it follows that λ_1 *is a nondecreasing function of n as long as $s_1 < s$* [16].

Time scale for growth of the dominant ACS.

If we assume that possibility 3 above is rare enough to neglect, and that the dominant ACS grows by adding a single node at a time, we can estimate the time required for it to span the entire graph. Let the dominant ACS consist of $s_1(n)$ nodes at time n. The probability that the new node gets an incoming link from the dominant ACS and hence joins it is ps_1. Thus in Δn graph updates, the dominant ACS will grow, on average, by $\Delta s_1 = ps_1 \Delta n$ nodes. Therefore

$s_1(n) = s_1(n_a)exp((n - n_a)/\tau_g)$, where $\tau_g = 1/p$, n_a is the time of arrival of the first ACS and $s_1(n_a)$ is the size of the first ACS (=2 for the run shown in Figure 16.10). Thus s_1 is expected to grow exponentially with a characteristic timescale $\tau_g = 1/p$. The time taken from the arrival of the ACS to its spanning is $\tau_g \ln(s/s_1(n_a))$. This analytical result is confirmed by simulations (see Figure 16.11).

In the displayed run, after the first ACS (a 2-cycle) is formed at $n = 2854$, it takes 1026 time steps, until $n = 3880$ for the dominant ACS to span the entire graph (Figure 16.10c). This explains how a autocatalytic network structure and the positive feedback processes inherent in it can bootstrap themselves into existence from a small seed. The small seed, in turn, is more or less guaranteed to appear on a certain time scale ($1/p^2s$ in the present model) just by random processes.
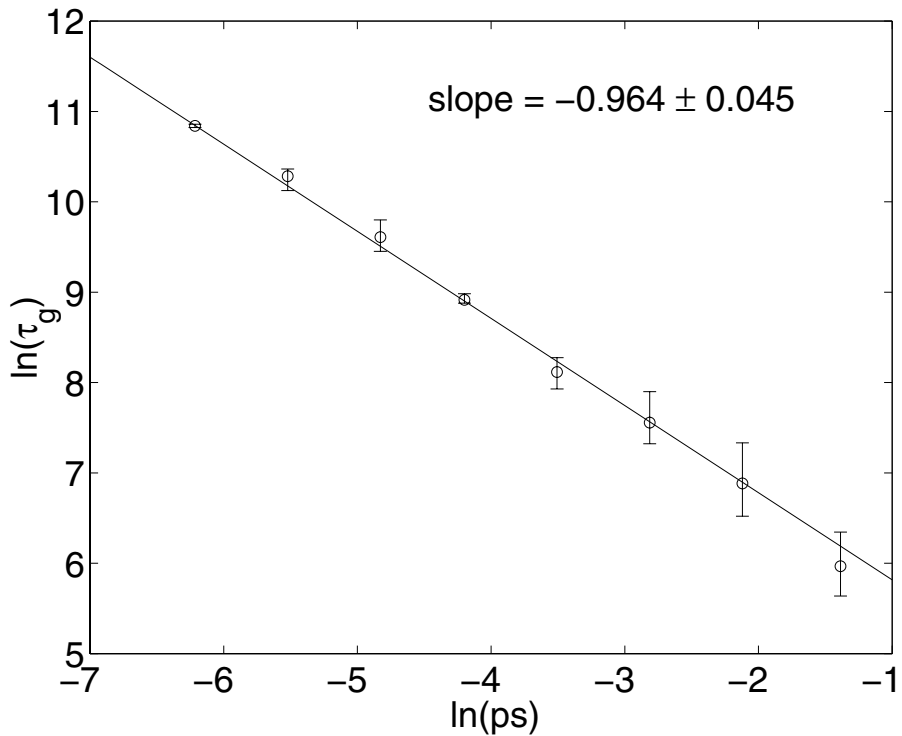
Figure 16.11: Each data point shows the average of τ_g (the growth timescale for an ACS) over 5 different runs with $s = 100$ and the given p value. The error bars correspond to one standard deviation. The solid line is the best linear fit to the data points on a log-log plot. Its slope is consistent with the analytically predicted slope -1 (see the discussion of the growth phase in section 5.)

A measure of the 'structure' of the evolved graph.

A fully autocatalytic graph is a highly improbable structure. Consider a graph of s nodes and let the probability of a positive link existing between any pair of nodes be p^*. Such a graph has on average $m^* = p^*(s-1)$ incoming or outgoing positive links per node since links from a node to itself are disallowed. For the entire graph to be an ACS, each node must have at least one incoming link, i.e. each row of the matrix C must contain at least one positive element. Hence the probability, P, for the entire graph to be an ACS is

$$
\begin{aligned}
P &= \text{probability that every row has at least one positive entry} \\
  &= [\text{probability that a row has at least one positive entry}]^s \\
  &= [1 - (\text{probability that every entry of a row is zero})]^s \\
  &= [1 - (1 - p^*)^{s-1}]^s \\
  &= [1 - (1 - m^*/(s-1))^{s-1}]^s
\end{aligned}
$$

Note from Figure 16.8 that at spanning the number of links is $O(s)$. Thus the average degree m^* at spanning is $O(1)$. We have found this to be true in all the runs we have done where the initial average degree (at $n = 1$) was $O(1)$ or less.

For large s and $m^* \sim O(1)$, $P \approx (1 - e^{-m^*})^s \sim e^{-\alpha s}$, where α is positive, and $O(1)$. Thus a fully autocatalytic graph is exponentially unlikely to form if it were being assembled randomly. In the present model nodes are being added completely randomly but the underlying population dynamics and the selection imposed at each graph update result in the inevitable arrival of an ACS (in, on average, $\tau_a = 1/p^2 s$ time steps) and its inevitable growth into a fully autocatalytic graph in (on average) an additional $\sim \tau_g \ln s$ time steps.

It is a noteworthy feature of self-organization in the present model that an organization whose a priori probability to arise is exponentially small, $\sim e^{-\alpha s}$, arises inevitably in a rather short time, $\sim \frac{1}{p} \ln s$ (for large s). Why does that happen? First a small ACS of size $s_1(n_a) \sim O(1)$ forms by pure chance. The probability of this happening is not exponentially small; it is in fact quite substantial. Once this has formed, it is a cooperative structure and is therefore stable. Its appearance ushers in an exponential growth of structure with a time scale $\tau_g = 1/p$. Hence a graph whose 'structuredness' (measured by the reciprocal of the probability of its arising by pure chance) $= e^{\alpha s}$ arises in only $\frac{1}{p} \ln s$ steps.

As mentioned in the introduction, one of the major puzzles in the origin of life is the emergence of very special chemical organizations in a relatively short time. We hope that the mechanism described above, or its analogue in a sufficiently realistic model, will help in addressing this puzzle. The relevance of this mechanism for the origin of life is discussed in ref. [21]. We remark that other models of self-organization (e.g. the well-stirred hypercycle) do not seem to be able to produce complex structured organizations from a simple starting network (see ref. [23]).

Another graph theoretic measure of the structure of the evolved graph is 'interdependency' among the nodes, discussed in [16, 21]. Like the links and s_1, the interdependency is low in the random phase, then rises in the growth phase to a value that is about an order of magnitude higher.

16.5.3   The organized phase

Once an ACS spans the entire graph the effective dynamics again changes although the microscopic dynamical rules are unchanged. At spanning, for the first time since the formation of the initial ACS, a member of the dominant ACS will be picked for removal. This is because at spanning all nodes by definition belong to the dominant ACS and have non zero relative populations; one node nevertheless has to be picked for removal. Most of the time the removal of the node with the least X_i will result in minimal damage to the ACS. The rest of the ACS will remain with high populations, and the new node will keep getting repeatedly removed and replaced until it once again joins the ACS. Thus s_1 will fluctuate between s and $s - 1$ most of the time. However, once in a while, the node which is removed happens to be playing a crucial role in the graph structure despite its low population. Then its removal can trigger large changes in the structure and catastrophic drops in s_1 and l. Alternatively it can sometimes happen that the new node added can trigger a catastrophe because of the new graph structure it creates. The catastrophes and the mechanisms which cause them are the subject of the next section.

16.6   Catastrophes and recoveries in the organized phase

Figure 16.12 shows the same run as that of Figure 16.9b for $n = 1$ to $n =50,000$. In this long run one can see several sudden, large drops in s_1: *catastrophes* in which a large fraction of the s species become extinct. Some of the drops seem to take the system back into the random phase, others are followed by *recoveries* in which s_1 rises back towards its maximum value s. The recoveries are comparatively slower than the catastrophes, which in fact occur in a single time step.

   In order to understand what is happening during the catastrophes and subsequent recoveries we begin by examining the possible changes that an addition or a deletion of a node can make to the core of the dominant ACS.

Deletion of a node

We have already seen how the deletion of a node can change the core – recall the discussion of keystone nodes in section 3: the removal of a keystone node results in a zero overlap between the cores of the dominant ACS before and after the removal. A zero core overlap means that a single graph update event (in which one of the least populated species is replaced by a randomly connected one) has caused a major reorganization of the dominant ACS: the cores of the dominant ACS before and after the event (if an ACS still exists) have not even a single link in common. We will call such events *core-shifts*.

   In an actual run a keystone node can only be removed if it happens to be one of the nodes with the least X_i. However the core nodes are often 'protected' by having higher X_i. Why is that?

   X is an eigenvector of C with eigenvalue λ_1. Therefore, when $\lambda_1 \neq 0$ it follows that for nodes of the dominant ACS, $X_i = (1/\lambda_1) \sum_j c_{ij} X_j$. If node i of the dominant ACS has only one incoming link (from the node j, say) then $X_i = X_j/\lambda_1$; we can say that X_i is 'attenuated' with respect to X_j by a factor λ_1. The periphery of an ACS is a tree like structure emanating
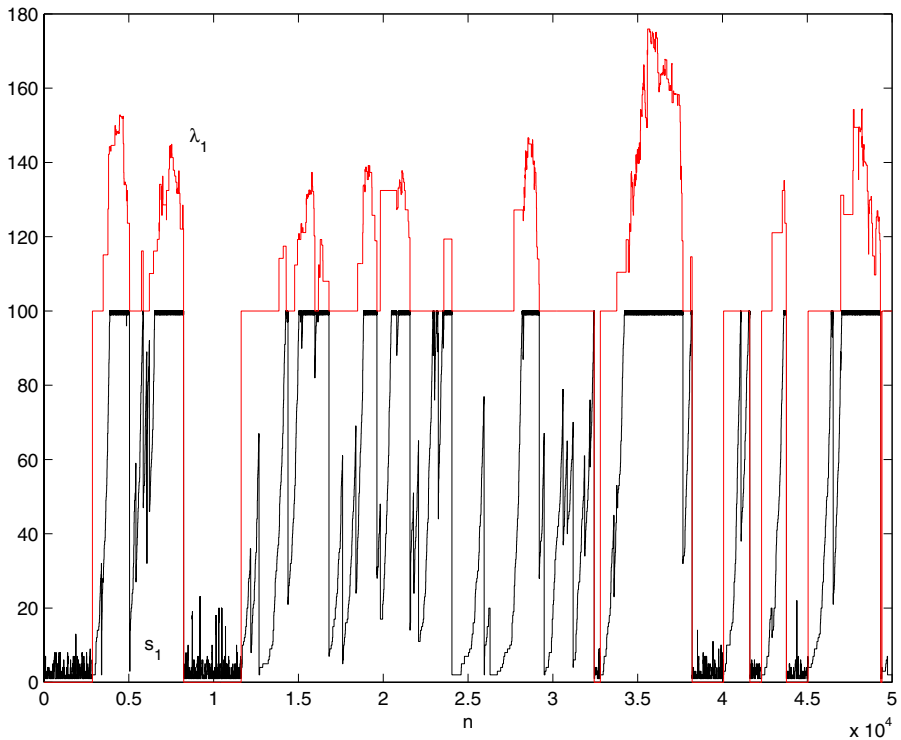
Figure 16.12: The same run displayed in Figure 16.9b over a longer timescale, till $n = 50000$. This displays repeated rounds of crashes and recoveries.

from the core, and for small p most periphery nodes have a single incoming link. For instance the graph in Figure 16.10c, whose $\lambda_1 = 1.31$, has a chain of nodes $44 \rightarrow 45 \rightarrow 24 \rightarrow 29 \rightarrow 52 \rightarrow 89 \rightarrow 86 \rightarrow 54 \rightarrow 78$. The farther down such a chain a periphery node is, the lower is its X_i because of the cumulative attenuation. For such an ACS with $\lambda_1 > 1$ the 'leaves' of the periphery tree (such as node 78) will typically be the species with least X_i while the core nodes will have larger X_i.

However, when $\lambda_1 = 1$ there is no attenuation. Recall that Proposition 1(iii) shows that at $\lambda_1 = 1$ the core must be a cycle or a set of disjoint cycles, hence each core node has only one incoming link within the dominant ACS. All core nodes have the same value of X_i. As one moves out towards the periphery $\lambda_1 = 1$ implies there is no attenuation, hence each node in the periphery that receives a single link from one of the core nodes will also have the same X_i. Some periphery nodes may have higher X_i if they have more than one incoming link from the core. Iterating this argument as one moves further outwards from the core, it is clear that at $\lambda_1 = 1$ the core is not protected and in fact will always belong to the set of least fit nodes if the dominant ACS spans the graph. We have already seen in section 3 that when $\lambda_1 = 1$ and the core is a single cycle every core node is a keystone node. Thus when $\lambda_1 = 1$ the organization is fragile and susceptible to core-shifts caused by the removal of a keystone node.

Addition of node

We now turn to the effects of the addition of a node to the dominant ACS. We will use the notation $C'_n \equiv C_{n-1} - k$ for the graph of $s - 1$ nodes just before the new node at time step n is brought in (and just after the least populated species k is removed from C_{n-1}). Q'_n will stand for the core of C'_n. In the new attractor the new species k may go extinct, i.e., X_k may be zero, or it may survive, i.e., X_k is non-zero. If the new species goes extinct then it remains in the set of least fit nodes and clearly there is no change to the dominant ACS. So we will focus on events in which the new species survives in the new attractor.

Innovations

We define an *innovation* to be a new node for which X_k in the new attractor is nonzero, i.e. a new node which survives till the next graph update [23]. This may seem to be a very weak requirement, yet we will see that it has nontrivial consequences. A description of various types of innovations and their consequences, with examples, is given in [30]. Here we present a graph-theoretic classification of innovations (in terms of a hierarchy, see Figure 16.13).

   Remarks to Fig. 16.13: All classes of events except the leaves of the tree are subdivided into two exhaustive and mutually exclusive subclasses (represented by the two branches emanating downwards from the class). The number of events in each class pertain to the run of Figure 16.9b with a total of 9999 graph updates, between $n = 1$ (the initial graph) and $n = 10000$. In that run, out of 9999 node addition events, most (8929 events) are not innovations. The rest (1070 events), which are innovations, are classified according to their graph theoretic structure. The classification is general; it is valid for all runs. X_k is the relative population of the new node in the attractor configuration of (16.1) that is reached in step 1 of the dynamics (see Section 4) immediately following the addition of that node. N stands for the new irreducible subgraph, if any, created by the new node. If the new node causes a new irreducible subgraph to be created, N is the maximal irreducible subgraph that includes the new node. If not, $N = \phi$ (where ϕ stands for the empty set). Q_{in} is the core of the graph just before the addition of the node (just before step 3 of the dynamics in Section 4) and Q_{fin} the core just after the addition of the node. The six leaves of the innovation subtree are numbered from 1 to 6 and correspond to the classes discussed in Section 6. The impact of each kind of innovation on the system dynamics is discussed in the text and in more detail in [30]. Some classes of events happen rarely (e.g., classes numbered 5 and 6) but have a major impact on the dynamics of the system. The precise impact of all these classes of innovations on the system over a short time scale (before the next graph update) as well as their probable impact over the medium term (upto a few thousand graph updates) can be predicted from the graph theoretic structure of N and the rest of the graph at the moment these innovations appear in a run.

   The innovations which have the least impact on the populations of the species and the evolution of the graph on a short time scale (of a few graph updates) are ones which do not affect the core of the dominant ACS, if it exists. Such innovations are of three types (see boxes 1-3 in Figure 16.13):

1. Random phase innovations.   These are innovations which occur in the random phase when no ACS exists in the graph, and they do not create any new ACSs. These innovations are typically short lived and have little short term or long term impact on the structure of the graph.
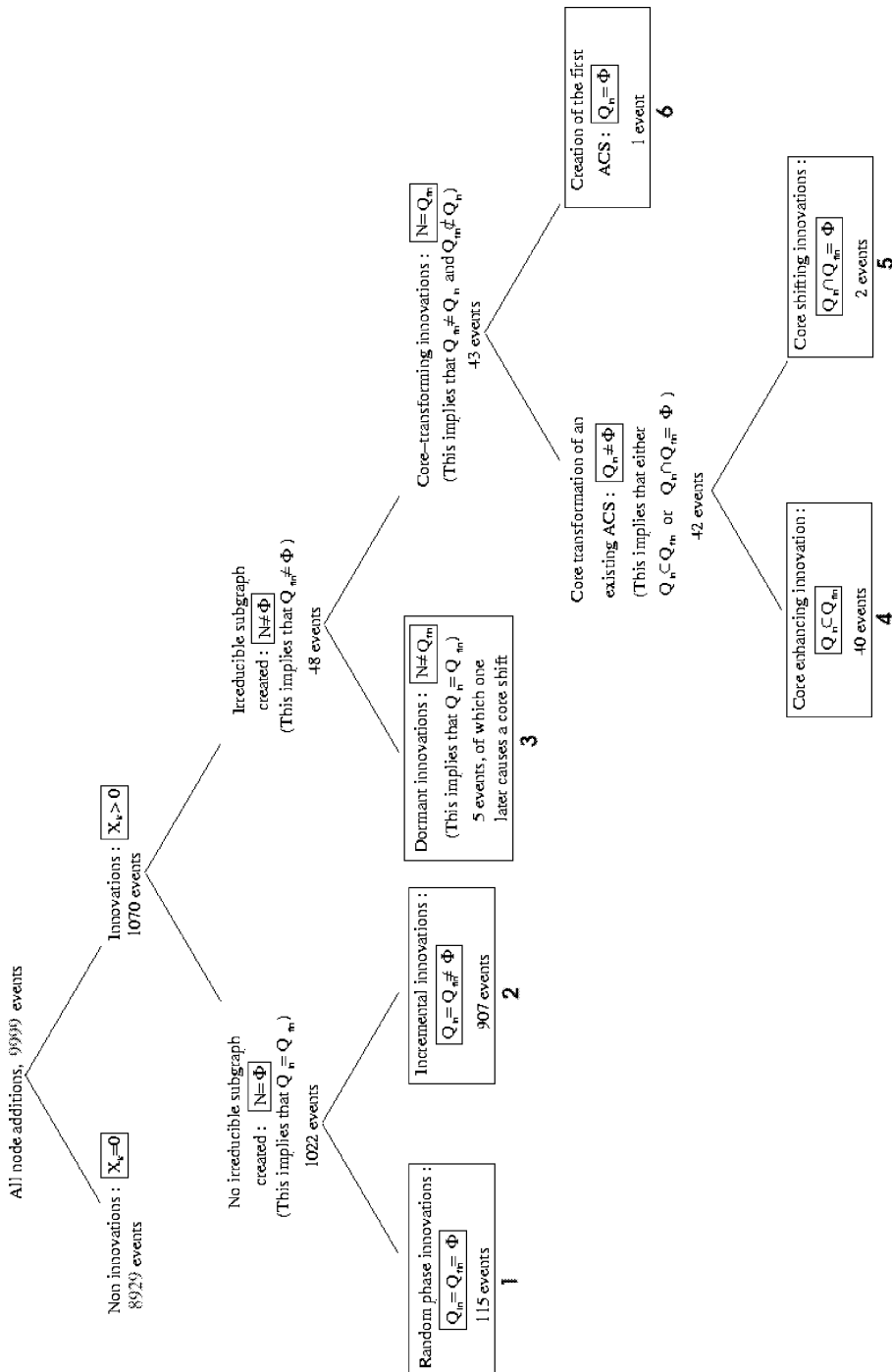
Figure 16.13: A hierarchy of innovations. Each node in this binary tree represents a class of node addition events. Each class has a name; the small box contains the mathematical definition of the class.

2. Incremental innovations.   These are innovations which occur in the growth and organized phases, which add new nodes to the periphery of the dominant ACS without creating any new irreducible subgraph. In the short term they only affect the periphery and are responsible for the growth of the dominant ACS. In a longer term they can also affect the core as chains of nodes from the periphery join the core of the dominant ACS.

3. Dormant innovations.   These are innovations which occur in the growth and organized phases, which create new irreducible subgraphs in the periphery of the dominant ACS. These innovations also affect only the periphery in the short term. But they have the potential to cause core-shifts later if the right conditions occur (discussed in the next subsection).

Innovations which do immediately affect the core of the existing dominant ACS are always ones which create a new irreducible subgraph. They are also of three types (see boxes 4-6 in Figure 16.13):

4. Core enhancing innovations.   These innovations result in the expansion of the existing core by the addition of new links and nodes from the periphery or outside the dominant ACS. They result in an increase of λ_1 of the graph.

5. Core-shifting innovations.   These are innovations which cause an immediate core-shift often accompanied by the extinction of a large number of species.

6. Creation of the first ACS.   This is an innovation which creates an ACS for the first time in a graph which till then had no ACSs. The innovation moves the system from the random phase to the growth phase, triggering the self organization of the system around the newly created ACS.

Innovations of types 4, 5 and 6 which affect the core of the dominant ACS will be called *core-transforming innovations*. These innovations cause a substantial change the vector of relative populations in a single graph update. Innovations of type 5 and 6 also make a qualitative change in the structure of the graph and significantly influence subsequent graph evolution. The following theorem makes precise the conditions under which a core transforming innovation can occur.

Core transforming Theorem

Let N (or N_n at time step n) denote the maximal new irreducible subgraph which includes the new species. One can show that N_n will become the new core of the graph, replacing the old core Q_{n-1}, whenever either of the following conditions are true:
(a) $\lambda_1(N_n) > \lambda_1(Q'_n)$ or,
(b) $\lambda_1(N_n) = \lambda_1(Q'_n)$ and N_n is 'downstream' of Q'_n (i.e., there is a path from Q'_n to N_n but not from N_n to Q'_n.)

Such an innovation will fall into category 4 above if $Q_{n-1} \subset N_n$. However, if Q_{n-1} and N_n are disjoint, we get a core-shift and the innovation is of type 5 if Q_{n-1} is non-empty and type 6 otherwise.

16.6.1   Catastrophes, core-shifts and a classification of proximate causes

The large sudden drops visible in Figure 16.12 are now discussed. Our first task is to see if the large drops are correlated to specific changes in the structure of the graph. Let us focus on those events in which more than 50% of the species go extinct. There were 701 such events out of 1.55 million graph updates in a set of runs with $s = 100, p = 0.0025$. Figure 16.14 shows a histogram of core overlaps $Ov(C_{n-1}, C_n)$ for these 701 events. 612 of these have zero core overlap, i.e., they are core-shifts.
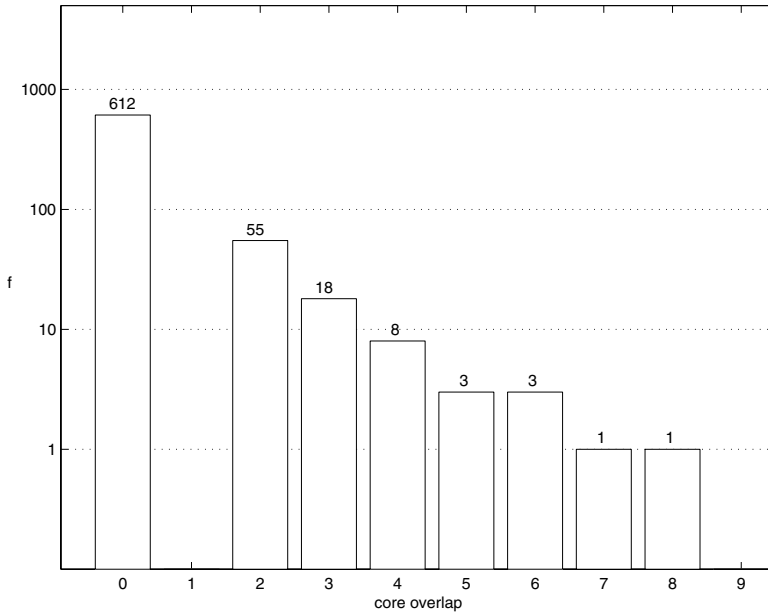
Figure 16.14: Large crashes are predominantly core-shifts. A histogram of core overlaps for the 701 events where s_1 dropped by more than $s/2$ observed in various runs with $s = 100$ and $p = 0.0025$, totalling 1.55 million iterations.

If we now look at only those events in which more than 90% of the species went extinct then we find 235 such events in the same runs, out of which 226 are core-shifts. Clearly most of the large extinction events happen when there is a drastic change in the structure of the dominant ACS – a core-shift.

Classification of core-shifts

Using the insights from the above discussion of the effects of deletion or addition of a node, we can classify the different mechanisms which cause core-shifts. Figure 16.15 differentiates between the 612 core-shifts we observed amongst the 701 crashes.

They fall into three categories [23]: (i) complete crashes (136 events), (ii) takeovers by core-transforming innovations (241 events), and (iii) takeovers by dormant innovations (235 events).
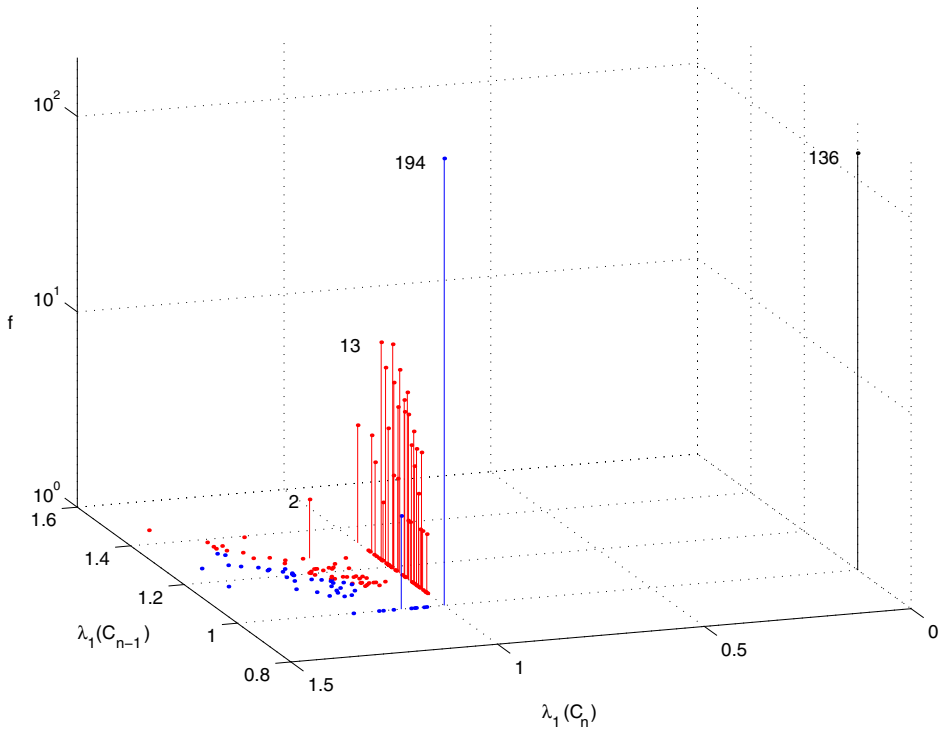
Figure 16.15: Classification of core-shifts into three categories. The graph shows the frequency, f, of the 612 core-shifts observed (see Figure 16.14) in a set of runs with $s = 100$ and $p = 0.0025$ vs. the λ_1 values before, $\lambda_1(C_{n-1})$, and after, $\lambda_1(C_n)$, the core-shift. Complete crashes (black; $\lambda_1(C_{n-1}) = 1$, $\lambda_1(C_n) = 0$), takeovers by core-transforming innovations (blue; $\lambda_1(C_n) \geq \lambda_1(C_{n-1}) \geq 1$) and takeovers by dormant innovations (red; $\lambda_1(C_{n-1}) > \lambda_1(C_n) \geq 1$) are distinguished. Numbers alongside vertical lines represent the corresponding f value.

Complete crashes

A *complete crash* is an event in which an ACS exists before but not after the graph update. Such an event takes the system into the random phase. A complete crash occurs when a keystone node is removed from the graph. For example at $n = 8232$ the graph had $\lambda_1 = 1$ and its core was the simple 3-cycle of nodes 20, 50 and 54. As we have seen above, when the core is a single cycle every core node is a keystone node and is also in the set of least fit nodes. At $n = 8233$ node 54 was removed thus disrupting the 3-cycle. The resulting graph had no ACS and λ_1 dropped to zero. As we have discussed earlier, graphs with $\lambda_1 = 1$ are the ones which are most susceptible to complete crashes. This can be seen in Figure 16.15: every complete crash occurred from a graph with $\lambda_1(C_{n-1}) = 1$.

Takeovers by core-transforming innovations

An example of a takeover by a core-transforming innovation is given in Figures 16.10g,h. At $n = 6061$ the core was a single loop comprising nodes 36 and 74. Node 60 was replaced by a new species at $n = 6062$ creating a cycle comprising nodes $60, 21, 41, 19$ and 73, downstream from the old core. The graph at $n = 6062$ has one cycle feeding into a second cycle that is downstream from it. We have already seen in section 3 (see the discussion of Example 4) that for such a graph only the downstream cycle is populated and the upstream cycle and all nodes dependent on it go extinct. Thus the new cycle becomes the new core and the old core goes extinct resulting in a core-shift. This is an example of condition (b) for a core-transforming innovation. For all such events in Figure 16.15, $\lambda_1(Q'_n) = \lambda_1(C_{n-1})$ since k happened not to be a core node of C_{n-1}. Thus these core-shifts satisfy $\lambda_1(C_n) = \lambda_1(N_n) \geq \lambda_1(Q'_n) = \lambda_1(C_{n-1}) \geq 1$ in Figure 16.15.

Takeovers by dormant innovations

We have earlier discussed dormant innovations, which create an irreducible structure in the periphery of the dominant ACS which does not affect its core at that time. For example the 2-cycle comprising nodes 36 and 74 formed at $n = 4696$. At a later time such a dormant innovation can result in a core-shift if the old core gets sufficiently weakened.

In this case the core has become weakened by $n = 5041$, when it has $\lambda_1 = 1.24$. The structure of the graph at this time is very similar to the graph in Figure 16.7a. Just as node 3 in Figure 16.7a was a keystone node, here nodes 44, 85, and 98 are keystone nodes because removing any of them results in a graph like Figure 16.7b, consisting of two 2-cycles, one downstream from the other.

Indeed at $n = 5041$, node 85 is hit and the resulting graph at $n = 5042$ has a cycle (26 and 90) feeding into another cycle (36 and 74). Thus at $n = 5042$ nodes 36 and 74 form the new core with only one other downstream node, 11, being populated. All other nodes become depopulated resulting in a drop in s_1 by 97. A dormant innovation can takeover as the new core only following a keystone extinction which weakens the old core. In such an event the new core necessarily has a lower (but nonzero) λ_1 than the old core, i.e., $\lambda_1(C_{n-1}) > \lambda_1(C_n) \geq 1$ (see Figure 16.15).

Note that 85 is a keystone node, and the graph is susceptible to a core-shift *because* of the innovation which created the cycle 36-74 earlier. If the cycle between 36 and 74 were absent, 85 would *not* be a keystone species by our definition, since its removal would still leave part of the core intact (nodes 26 and 90).

16.6.2   Recoveries

After a complete crash the system is back in the random phase. In $O(s)$ graph updates each node is removed and replaced by a randomly connected node, resulting in a graph as random as the initial graph. Then the process starts again, with a new ACS being formed after an average of $1/p^2 s$ time steps and then growing to span the entire graph after, on average, $(1/p) \ln(s/s_0)$ time steps, where s_0 is the size of the initial ACS that forms in this round (typically $s_0 = 2$).

After other catastrophes, an ACS always survives. In that case the system is in the growth phase and immediately begins to recover, with s_1 growing exponentially on a timescale $1/p$. Note that these recoveries happen because of innovations (mainly of type 2 and 4, and some of type 3).

16.6.3   Correlation between graph theoretic nature of perturbation and its short and long term impact

In previous sections we have analysed several examples of perturbations to the system. These can be broadly placed in two classes based on their effect on s_1:

(i) 'Constructive perturbations': these include the birth of a new organization (an innovation of type 6), the attachment of a new node to the core (an innovation of type 4) and an attachment of a new node to the periphery of the dominant ACS (an innovation of type 2).

(ii) 'Destructive perturbations': these include complete crashes and takeovers by dormant innovations (both caused by the loss of a keystone node), and takeovers by core-transforming innovations (innovations of type 5). Note that the word 'destructive' is used only in the sense that several species go extinct on a short time scale (a single graph update in the present model) after such a perturbation. In fact, over a longer time scale (ranging from a few to several hundred graph updates in the run of Figure 16.9b), the 'destructive' takeovers by innovations generally trigger a new round of 'constructive' events like incremental innovations (type 2) and core enhancing innovations (type 4).

Note that the maximum upheaval is caused by those perturbations that introduce new irreducible structures in the graph (innovations of type 4, 5 and 6) or those that destroy the existing irreducible structure. For example the creation of the first ACS at $n = 2854$ triggered the growth phase, a complete change in the effective dynamics of the system. Other examples of large upheavals are core-shifts caused by a takeover by a core-transforming innovation at $n = 6061$, takeover by a dormant innovation at $n = 5041$, and a complete crash at $n = 8233$. In sections 2 and 3 we have mentioned that irreducibility is related to the existence of positive feedback and cooperation, and the 'magnitude' of the feedback is measured by λ_1. While the present model is a highly simplified model of evolving networks, we expect that this qualitative feature, namely, the correlation between the dynamical impact of a perturbation and its 'structural' character embodied in its effect on the 'level of feedback' in the underlying graph, will hold for several other complex systems.

16.7   Concluding remarks

In this article we have attempted to show that a certain class of dynamical systems, those in which graphs coevolve with other dynamical variables living on them (in our example, living on the nodes of the graph), possess rich dynamical behaviour which is analytically and computationally tractable. Even in the highly idealized model discussed here, this behaviour is reminiscent of what happens in real life — birth of organizational structure characterized by interdependence of components, cooperation of parts of the organization giving way to competition, robust organizations becoming fragile, crashes and recoveries, innovations causing growth as well as collapse, etc.

From the point of view of the origin of life problem the main conclusions are:

(i) The model shows the emergence of an organization where none exists: a small ACS emerges spontaneously by random processes and then triggers the self-organization of the system.

(ii) A highly structured organization, whose timescale of forming by pure chance is exponentially large (as a function of the size of the system), forms in this model in a very short timescale that grows only logarithmically with the size of the system. In [21] we have speculated that this timescale may be ~ 100 million years for peptide based ACSs, which is in the same ballpark as the timescale on which life is believed to have originated on the prebiotic earth.

We remark that this speculation is not necessarily in conflict with, and is possibly complementary to, some other approaches to the origin of life:

(i) Complex autocatalytic organizations of polypeptides could enter into symbiosis with the autocatalytic citric acid cycle proposed in [31]. The latter would help produce, among other things, amino acid monomers needed by the former; the former would provide catalysts for the latter.

(ii) It is conceivable that membranes (possibly lipid membranes, which have been argued to have their own catalytic dynamics [32]) could form in regions where autocatalytic sets of the kind discussed here existed, thereby surrounding complex molecular organization in an enclosure. These 'cells' may have contained different parts of the ACS, thereby endowing them with different fitnesses. Such an assembly could evolve.

(iii) It is also conceivable that such molecular organizations formed an enabling environment for self replicating molecules such as those needed for an RNA world.

Testing some of these possibilities is a task for future models and experiments. Furthermore, the mathematical ideas and mechanisms discussed here might be relevant for these other approaches also.

The present model has a number of simplifying features which depart from realism but enhance analytical tractability. One is the linearity of the populations dynamics on a fixed graph. Equation (16.1) is nonlinear, but since it originates via a nonlinear change of variables from a linear equation, equation (16.2), its attractors can be easily analysed in terms of the underlying linear system. The attractors are always fixed points, and are just the Perron Frobenius eigenvectors of the adjacency matrix of the graph. This allows us to use (static) graph theoretic results for analysis of the dynamics.

In this context it is helpful to note that while the population dynamics in the present model is essentially linear as long as the graph is fixed, the model feeds the result of the population dynamics into the subsequent graph update (the least populated node is removed). Thus over long time scales over which the graph changes, the 'coupling constants' c_{ij} in equation (16.1) are not constant but implicitly depend upon the x_i, thus making the evolution highly nonlinear. By virtue of the simplifying device of widely separated time scales for the graph dynamics and the population dynamics (the population variables reach their attractor before the graph is modified), what we have is piecewise linear population dynamics. It is essentially linear

between two graph updates, and nonlinear over longer time scales because of the intertwining of population dynamics and graph dynamics. This nonlinearity is essential for all the complex phenomena described above, while the short time scale linearity is an aid in analysis. It would be interesting to explore complex phenomena in models in which the short term population dynamics is also inherently nonlinear. This naturally arises in prebiotic chemistry when the concentration of the reactants (which are assumed buffered here) are dynamical variables in addition to the catalysts and products, as well as in several other fields.

The present model describes a well-stirred reactor; there are no spatial degrees of freedom. This precludes a discussion of the origin of spatial structure and its consequences alluded to in section 1. It is worthwhile to extend the model in that direction. Another issue is the generation of novelty. Here the links of the new node are drawn from a fixed probability distribution. In real systems this distribution depends upon the (history of) states of the system. A further direction for generalization consists in letting the two time scales of the population and graph dynamics, separated by hand in the present model, be endogenous.

Acknowledgements

References

[1] Albert, R. and Barabasi, A.-L. (2002) Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74**, 47 *(www.arXiv.org/abs/cond-mat/0106096)*.

[2] Dorogovtsev, S. N. and Mendes, J. F. F. (2002) Evolution of networks, *Adv. Phys.* **51** 1079 *(www.arXiv.org/abs/cond-mat/0106144)*.

[3] Strogatz, S. H. (2001) Exploring complex networks, *Nature* **410**, 268-276.

[4] Watts, D. J. (1999) *Small Worlds: The dynamics of Networks between Order and Randomness* (Princeton Univ. Press, Princeton).

[5] Dyson, F. (1985) *Origins of Life* (Cambridge Univ. Press Cambridge, UK).

[6] Farmer, J. D., Kauffman, S. and Packard, N. H. (1986) Autocatalytic replication of polymers, *Physica* **D22** 50-67.

[7] Bagley, R. J., Farmer, J. D. and Fontana, W. (1991) Evolution of a metabolism, in *Artificial Life II*, eds. Langton, C. G., Taylor, C., Farmer, J. D. and Rasmussen, S. (Addison Wesley, Redwood City), pp. 141-158.

[8] Kauffman, S. A. (1993) *The Origins of Order* (Oxford Univ. Press).

[9] Bak, P. and Sneppen, K. (1993) Punctuated equilibrium and criticality in a simple model of evolution, *Phys. Rev. Lett.* **71**, 4083-4086.

[10] Fontana, W. and Buss, L. (1994) The arrival of the fittest: Toward a theory of biological organization, *Bull. Math. Biol.* **56**, 1-64.

[11] Harary, F. (1969) *Graph Theory* (Addison Wesley, Reading, MA, USA).

[12] Bang-Jensen, J. and Gutin, G. (2001) *Digraphs: Theory, Algorithms and Applications* (Springer-Verlag, London).

[13] Seneta, E. (1973) *Non-Negative Matrices* (George Allen and Unwin, London).

[14] Berman, A. and Plemmons, R. J. (1994) *Non-negative matrices in the mathematical sciences* (SIAM, Philadelphia).

[15] Rothblum, U. G. (1975) Algebraic eigenspaces of nonnegative matrices, *Linear Algebra and Appl* **12**, 281-292.

[16] Jain, S. and Krishna, S. (1999) Emergence and growth of complex networks in adaptive systems, *Computer Physics Comm.* **121-122**, 116-121.

[17] Eigen, M. (1971) Self-organization of matter and the evolution of biological macro-molecules, *Naturwissenschaften* **58**, 465-523.

[18] Kauffman, S.A. (1971) Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems, *J. Cybernetics* **1**, 71-96.

[19] Rossler, O. E. (1971) A system theoretic model of biogenesis, *Z. Naturforschung* **26b**, 741-746.

[20] Jain, S. and Krishna, S. (1998) Autocatalytic sets and the growth of complexity in an evolutionary model, *Phys. Rev. Lett.* **81**, 5684-5687.

[21] Jain, S. and Krishna, S. (2001) A model for the emergence of cooperation, interdependence and structure in evolving networks, *Proc. Natl. Acad. Sci. (USA)* **98**, 543-547.

[22] Jain, S. and Krishna, S. (2002) Crashes, recoveries and 'core-shifts' in a model of evolving networks, *Phys. Rev. E* **65**, 026103, *www.arXiv.org/abs/nlin.AO/0107037*.

[23] Jain, S. and Krishna, S. (2001) Large extinctions in an evolutionary model: the role of innovation and keystone species, *Proc. Natl. Acad. Sci. (USA)* **99**, 2055-2060, *www.arXiv.org/abs/nlin.AO/0107038*.

[24] Paine, R. T. (1969) A note on trophic complexity and community stability, *Am. Nat.* **103**, 91-93.

[25] Pimm, S. L. (1991) *The Balance of Nature? Ecological Issues in the Conservation of Species and Communities* (Univ. of Chicago Press, Chicago).

[26] Jordán, F., Takácks-Sánta, A. and Molnár, I. (1999) A reliability theoretical quest for keystones, *OIKOS* **86**, 453-462.

[27] Solé, R. V. and Montoya, J. M. (2000) Complexity and fragility in ecological networks, *www.arXiv.org/abs/cond-mat/0011196*.

[28] Joyce, G. F., Schwartz, A. W., Miller, S. L. and Orgel, L. E. (1987) The case for an ancestral genetic system involving simple analogues of the nucleotides, *Proc. Natl. Acad. Sci. (USA)* **84**, 4398-4402.

[29] Joyce, G. F. (1989) RNA evolution and the origins of life, *Nature* **338**, 217-223.

[30] Jain, S. and Krishna, S. (2002) Constructive and destructive effects of 'innovation' in evolving networks, Preprint 2002.

[31] Morowitz, H. J., Kostelnik, J. D., Yang, J. and Cody, G. D. (2000) The origin of intermediary metabolism, *Proc. Natl. Acad. Sci. (USA)* **97**, 7704-7708.

[32] Segré, D., Ben-Eli, D., Deamer, W. D. and Lancet, D. (2001) The lipid world, *Origins of Life and Evol. of the Biosphere* **31**, 119-145.