Autocatalytic and Genetic Nets

Presented by Janet Snape

CS 790R Seminar
Computational Models of Complex Systems
Instructor, René Doursat
University of Nevada, Reno
March 17, 2005

References

- *Kauffman, S. A. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. Journal of Theoretical Biology,* ***22****: 437-467.*

- *Kauffman, S. A. (1986) Autocatalytic sets of proteins. Journal of Theoretical Biology,* ***119****: 1-24.*

- *Kauffman, S. A. (1995)* ***At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*****.** *Oxford University Press.*

Overview

- **Introduction**
- **Genetic Nets**
- **Autocatalytic Sets of Proteins**
- **Origins of Life**
- **Questions and answers**

Introduction

- The origin of the very ability to evolve!
- Results of Kauffman's 30 years of work.
- New chemical creation hypothesis.
    - IF number of different molecules > threshold, THEN autocatalytic metabolism occurs
- Autocatalytic metabolism
    - A self-maintaining and self-sustaining network of reactions

Genetic Networks

- Hypothesis (1969):
  - ❖ Organisms are randomly constructed molecular automata.
- Gene modeled as binary device.
- Kauffman invented Boolean Networks to explore the origins of life.
- Results: Each gene is directly affected by two or three other genes.

Kauffman - 1969

- There is an understanding that a living organism is an interconnected network of chemical reactions. Kauffman believes that if original proto-organisms built their reaction nets randomly, then he believes we should build a theory for the metabolic behavior of these systems.
- Kauffman uses his hypothesis to initiate evolution.

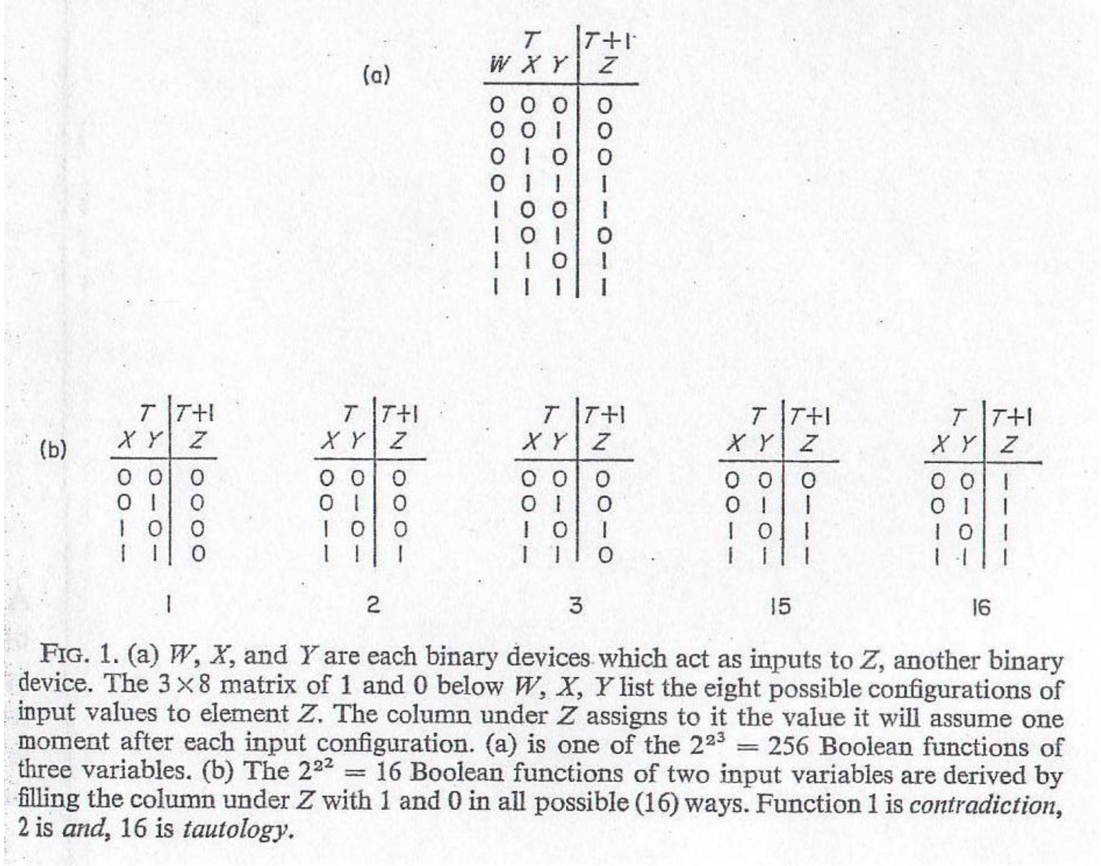
Kauffman's Evidence

- Large networks of binary elements (genes) behave with simplicity, stability, and order comparable to that in living organisms.
- Behavior cycles parallel and predict the time required for cell replication.
- The number of distinguishable modes of behavior of one randomly constructed net accurately predicts the number of cell types in an organism which embodies a genetic net of the same size.
- Genetic nets are like cells in that they are capable of differentiating directly from any one mode of behavior to at most a few of its other modes.

Kauffman’s Genetic Model

- Uses boolean functions (*contradiction*, *and*, *tautology*)
  - ❖ See Fig. 1, Slide #9
- Must contain one or more behavior cycles.
- Totally connected nets, K=N
  - ❖ Each element receives an input from ALL elements with the state randomly sampled from infinite supply of the 2^N distinct states of the net.
  - ❖ Chaotic / impossible
- One connected nets, K=1
  - ❖ Each element receives just one input. (Fig. 2(c) Slide #10)
  - ❖ chaotic / behavior cycles do NOT compare with living organism
- Two connected nets, K=2
  - ❖ Each element receives just two inputs from other elements
  - ❖ Order / embody short stable cycle / homeostatic

Boolean Functions



(a)

	T		T+1
W	X	Y	Z
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	1

(b)

T		T+1
X	Y	Z
0	0	0
0	1	0
1	0	0
1	1	0

1

T		T+1
X	Y	Z
0	0	0
0	1	0
1	0	0
1	1	1

2

T		T+1
X	Y	Z
0	0	0
0	1	0
1	0	1
1	1	0

3

T		T+1
X	Y	Z
0	0	0
0	1	1
1	0	1
1	1	1

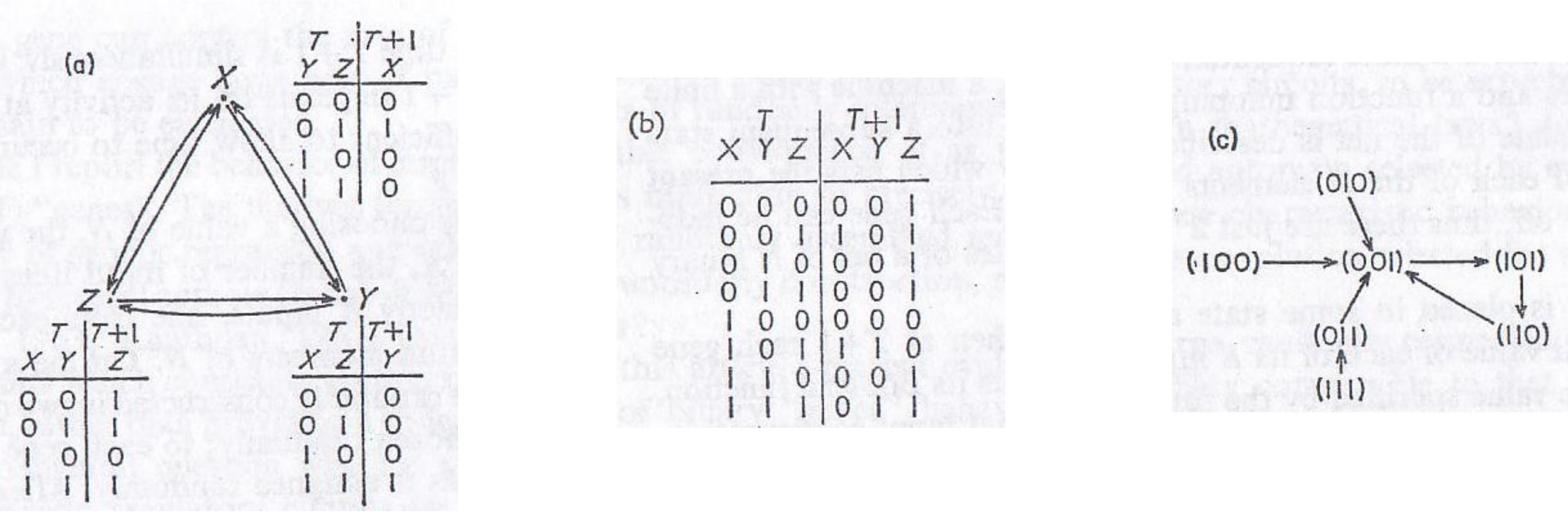
15

T		T+1
X	Y	Z
0	0	1
0	1	1
1	0	1
1	1	1

16

FIG. 1. (a) W, X, and Y are each binary devices which act as inputs to Z, another binary device. The 3×8 matrix of 1 and 0 below W, X, Y list the eight possible configurations of input values to element Z. The column under Z assigns to it the value it will assume one moment after each input configuration. (a) is one of the $2^{2^3} = 256$ Boolean functions of three variables. (b) The $2^{2^2} = 16$ Boolean functions of two input variables are derived by filling the column under Z with 1 and 0 in all possible (16) ways. Function 1 is *contradiction*, 2 is *and*, 16 is *tautology*.

Connected Nets



(a)

T		T+1
Y	Z	X
0	0	0
0	1	1
1	0	0
1	1	0

T		T+1
X	Y	Z
0	0	1
0	1	1
1	0	0
1	1	1

T		T+1
X	Z	Y
0	0	0
0	1	0
1	0	0
1	1	1

(b)

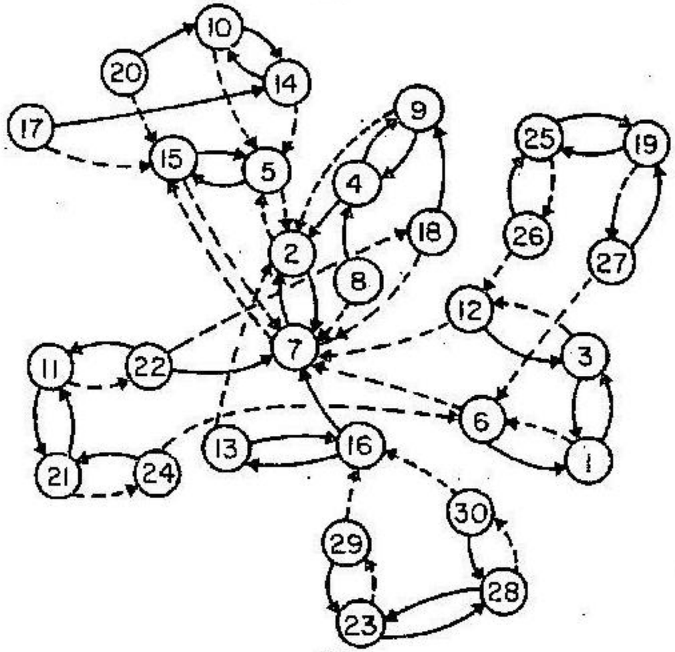
T			T+1		
X	Y	Z	X	Y	Z
0	0	0	0	0	1
0	0	1	1	0	1
0	1	0	0	0	1
0	1	1	0	0	1
1	0	0	0	0	0
1	0	1	1	1	0
1	1	0	0	0	1
1	1	1	0	1	1

FIG. 2. (a) A net of three binary elements, each of which receives inputs from the other two. The Boolean function assigned to each element is shown beside the element. (b) All possible states of the 3-element net are shown in the left 3×8 matrix below T. The subsequent state of the net at time $T+1$, shown in the matrix on the right, is derived from the inputs and functions shown in (a). (c) A kimatograph showing the sequence of state transitions leading into a state cycle of length 3. All states lie on one confluent. There are three run-ins to the single state cycle.

Markov Chain and Noise Perturbation

- Thus…Kauffman’s belief from the modeling, is that the cycle stabilization correlates to Schrodinger’s 1944 molecular specificity guaranteed by quantum stabilization required for the precision of biosynthesis in living organisms.
- Kauffman hypothesizes that living genetic nets are randomly formed which is consistent with random modifications of protein structure induced by mutation, etc. IF each element is directly affected by about the same low number of other elements as are macromolecules in living organisms, then Kauffman believes this supports the hypothesis that living metabolic nets are randomly constructed.
- See Fig. 10 Slide #12

Markov Chain



(b)

FIG. 10. (a) A matrix listing the 30 cycles of one net and the total number of times one unit of perturbation shifted the net from each cycle to each cycle. The system generally returns to the cycle perturbed. Division of the value in each cell of the matrix by the total of its row yields the matrix of transition probabilities between modes of behavior which constitute a Markov chain. The transition probabilities between cycles may be asymmetric. (b) Transitions between cycles in the net shown in (a). The solid arrows are the most probable transition to a cycle other than the cycle perturbed, the dotted arrows are the second most probable. The remaining transitions are not shown. Cycles 2, 7, 5 and 15 form an ergodic set into which the remaining cycles flow. If *all* the transitions between cycles are included, the ergodic set of cycles becomes: 1, 2, 3, 5, 6, 12, 13, 15, 16. The remainder are transient cycles leading into this single ergodic set.

Kauffman vs. Hof & Sparrow

- Kauffman uses his Boolean Network model to predict cell replication (characteristic cyclic phenomena).
- He compares it to the Hof & Sparrow (1963) study of cell replication time in higher organisms. See Fig. 11 Slide #14
    - ❖ Linear function of DNA content per nucleus
- Kauffman’s cell replication. See Fig. 12 Slide #15
    - ❖ Function of the number of genes per cell -- throughout a wide range of phyla.

Kauffman vs. Hof & Sparrow

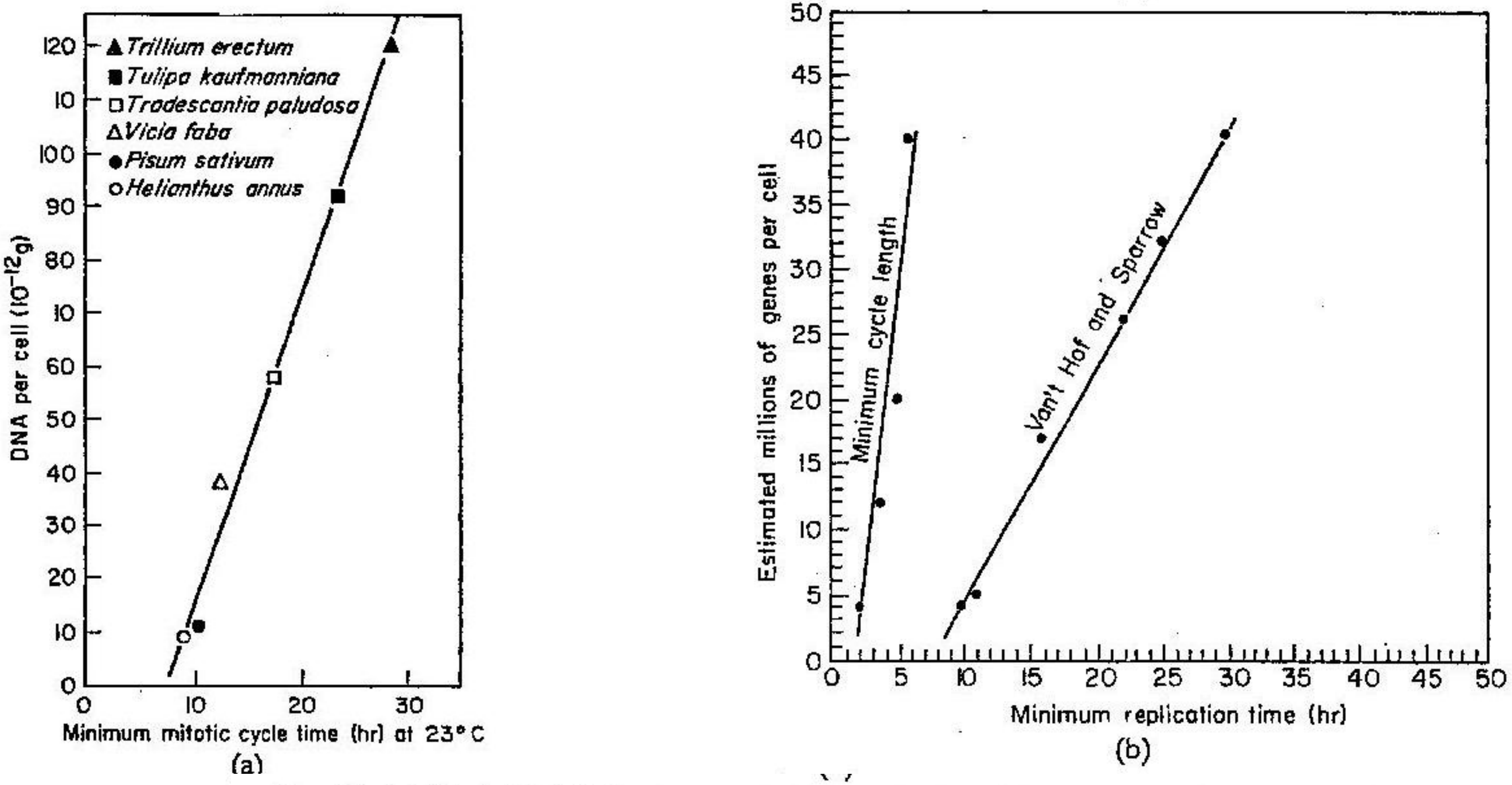


FIG. 11. (a) Van't Hof & Sparrow's (1963) Fig. 3 showing minimum cell replication time as a function of the DNA per nucleus for several plant species. (b) Projected cycle time in nets of 2 to 40 million binary genes using all 16 Boolean functions of two input variables, compared to Van't Hof and Sparrow's plot. In the range where Van't Hof and Sparrow report a linear relation, the binary net model predicts values which are nearly linear. Reduction in the number of elements assigned tautology or contradiction should raise expected cycle lengths and shift the nearly linear slope of the theoretical data to correspond closely with Van't Hof and Sparrow's data.

Cell Replication

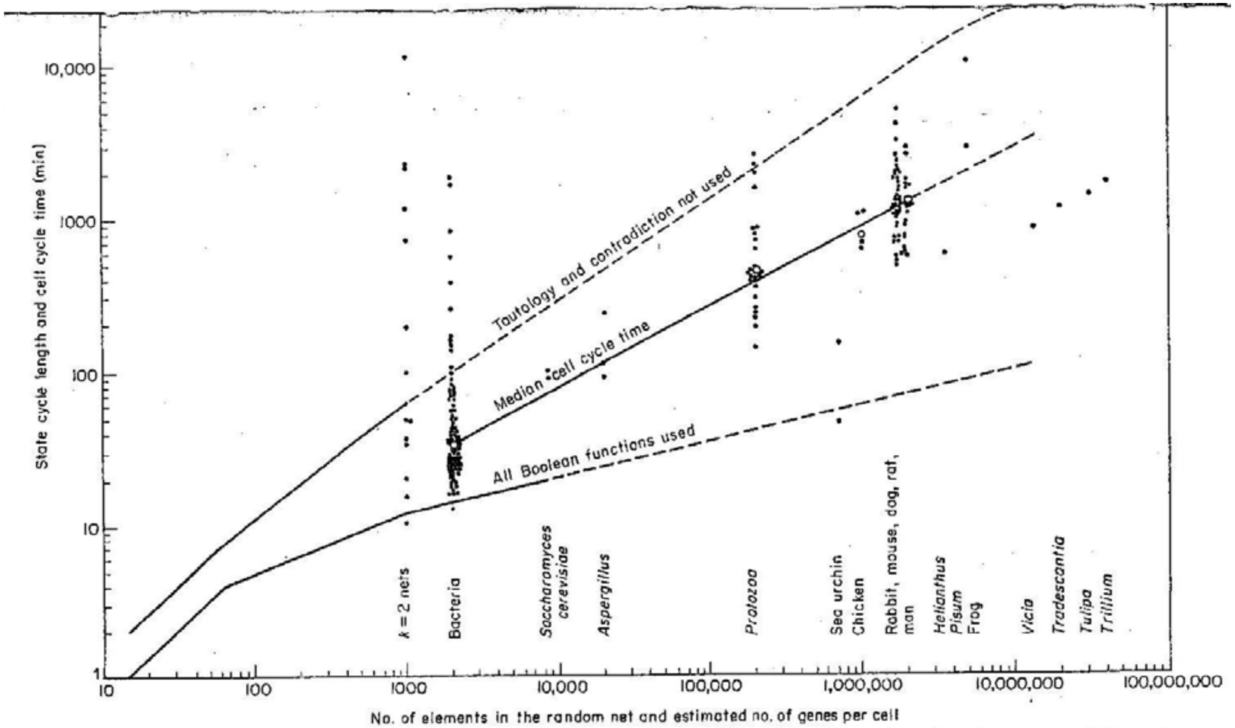


FIG. 12. Logarithm of cell replication time in minutes plotted against the logarithm of the estimated number of genes per cell for various single cell organisms, and various cell types in several metazoan organisms. The data for the plants, *Vicia faba*, *Pisum*, etc. are the *minimum* replication times described by Van't Hof & Sparrow (1963). The solid line through the biologic data connects the median replication times of bacteria, protozoa, chicken, mouse and dog and rabbit, and man. Data from binary nets of 1024 elements using neither nor contradiction are included for comparison. Median cycle lengths in binary nets with and without tautology and contradiction, as a function of the number of elements in the net, are superimposed on the biologic data. Scale: 2×10^6 genes = 6×10^{-12} g DNA per cell.

Kauffman’s Cell Differentiation

- Known facts:
  - Genome can behave in more than one mode.
  - There are mechanisms which can insure the appropriate assignment of a mode to a particular cell.
  - **Jacob and Monod won Nobel Prize mid-1960s for their ground-breaking article on cell differentiation.**
- Kauffman contends that differential activity of genes can be deduced from large randomly constructed genetic nets because:
  - The binary elements behaved in multiple distinct modes.
  - Different state cycles are isolated from each other for no state can be on two cycles.
  - As we can identify one cell type with one state cycle, Kauffman demonstrates the occurrence of multiple modes of behavior in his genetic system.
  - See Slide #17

Cell Differentiation

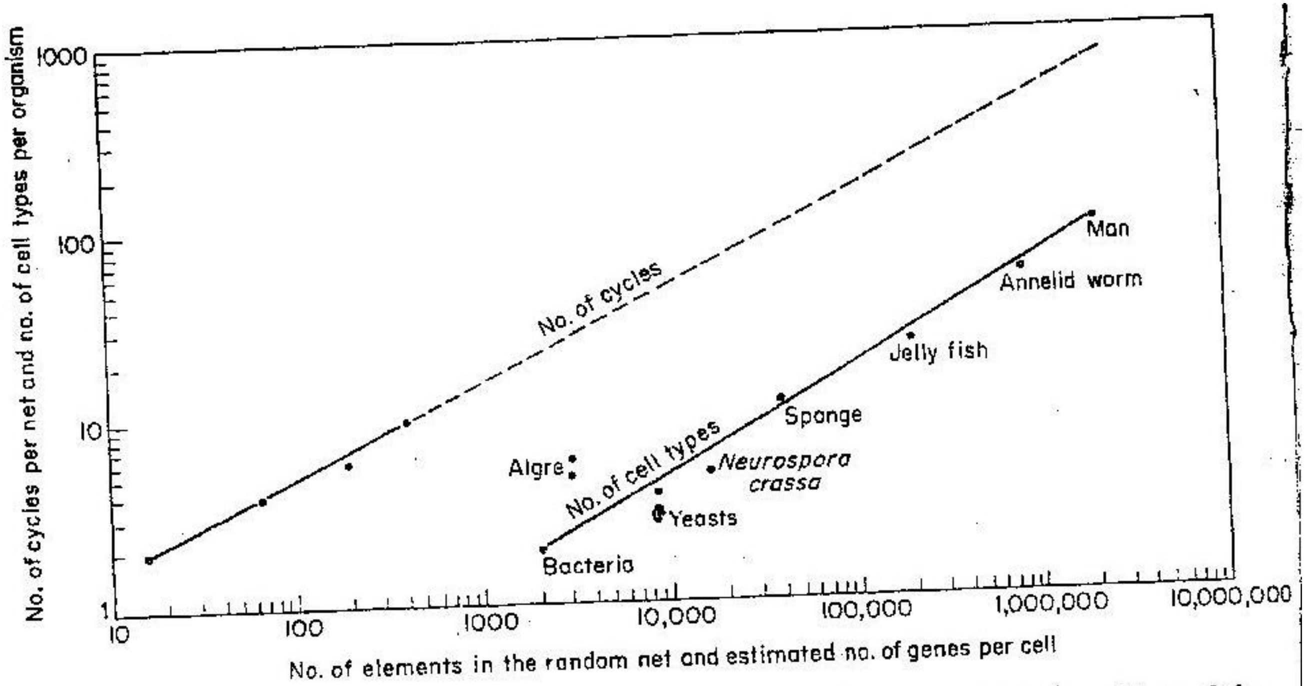


FIG. 13. The logarithm of the number of cell types is plotted against the logarithm of the estimated number of genes per cell, and the logarithm of the median number of state cycles is plotted against logarithm *N*. The observed and theoretical slopes are about 0·5. Scale: 2×10^6 genes per cell $= 6 \times 10^{-12}$g DNA per cell.

Boolean Genetic Networks
1980s

- Early 1980s brought on renewed interest in the theoretical possibilities of autocatalytic sets of proteins.
- In biology, a gene specifies a protein.
- Proteins are components of all organic bodies composed of 20+ amino acids linked in a genetically linear sequence into one or more long polypeptide chain.

Autocatalytic Sets of Proteins
1986

- After Kauffman models his boolean genetic networks for several years, he is led to hypothesize autocatalytic sets of proteins.

Kauffman’s Catalysis Theory (1986)

- Catalysis
  - ❖ the causing or accelerating of a chemical change by the addition of a catalyst.
- Catalyst
  - ❖ a substance that causes or accelerates a chemical reaction without itself being affected
- Theory
  - ❖ With a set of amino acid monomer and polymer species up to some max length, M, the potential number of possible peptides are huge for end condensation, cleavage, and transpeptidation exchange, and reverse reactions.
  - ❖ As M increases, the ratio of reactions among possible polypeptides to polypeptides rises rapidly, making the existence of autocatalytic subsets assured for any fixed probability of catalysis.

Assumption

- Each polymer has an independent probability to catalyze any reactions.
- Kauffman tests the probability that an arbitrary protein-like polymer catalyzes an arbitrary reaction.
  - ❖ Also calculates and tests the minimal size polymer sets required to nucleate collective self-replication.

Theory Implications

- This theory could have substantial implications for the origin of life.
- Theory suggests the emergence of self-replicating systems are a self-organizing collective property of complex protein systems in prebiotic evolution.

Main Model Feature

- Law of minimum complexity
  - “To obtain connected catalyzed transformations as an emergent collective property, a sufficient complexity is needed. Smaller systems fail to achieve catalytic closure.”

Thoughts

- Do autocatalytic sets undergo selective adaptation?
- Are autocatalytic sets able to catalyze its own and only its own formation?

Simple Autocatalytic Set

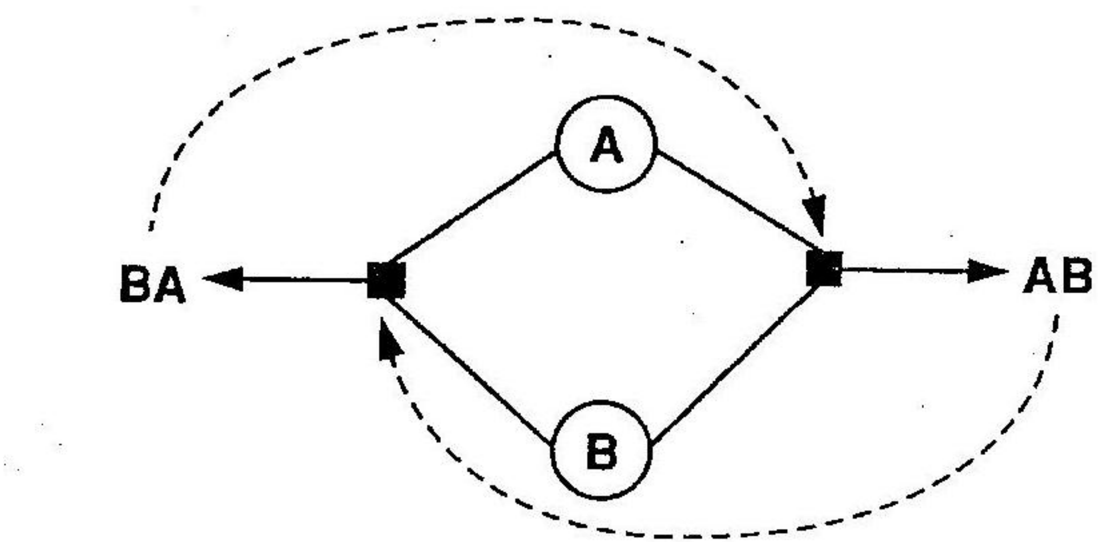


Figure 3.1 *A simple autocatalytic set. Two dimer molecules, AB and BA, are formed from two simple monomers, A and B. Since AB and BA catalyze the very reactions that join As and Bs to make the dimers, the network is autocatalytic: given a supply of "food" molecules (As and Bs), it will sustain itself.*

Monomers and Polymers

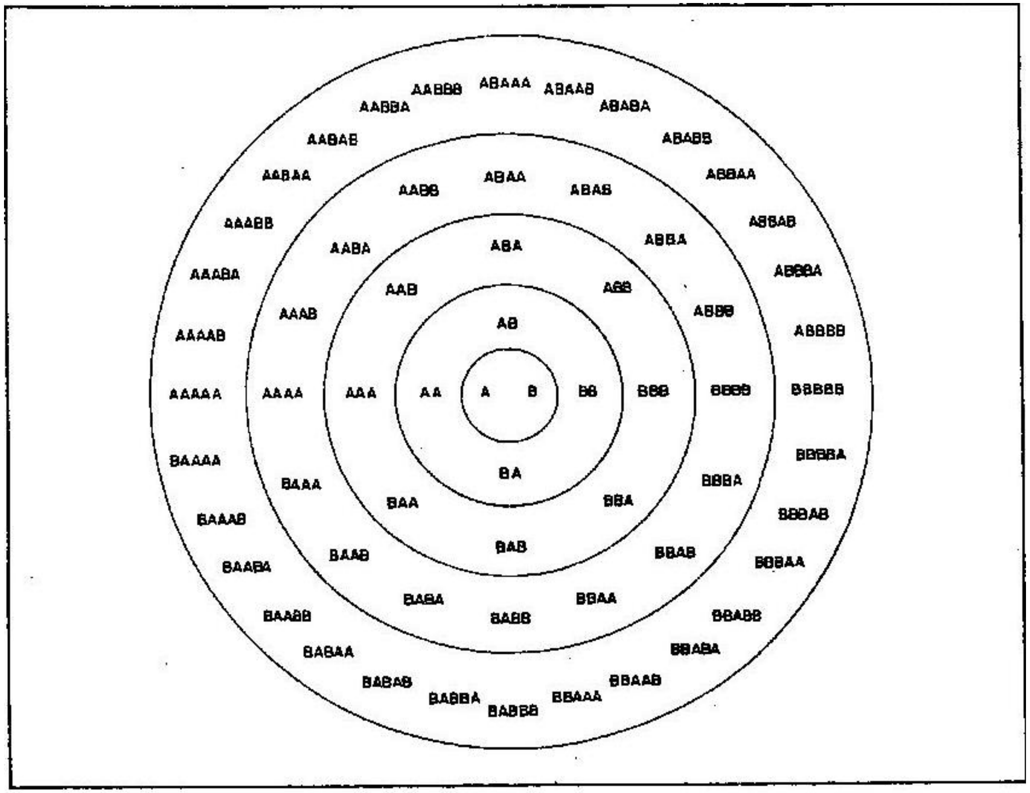


FIG. 1. Monomers A and B and polymers to length 5.

Condensation

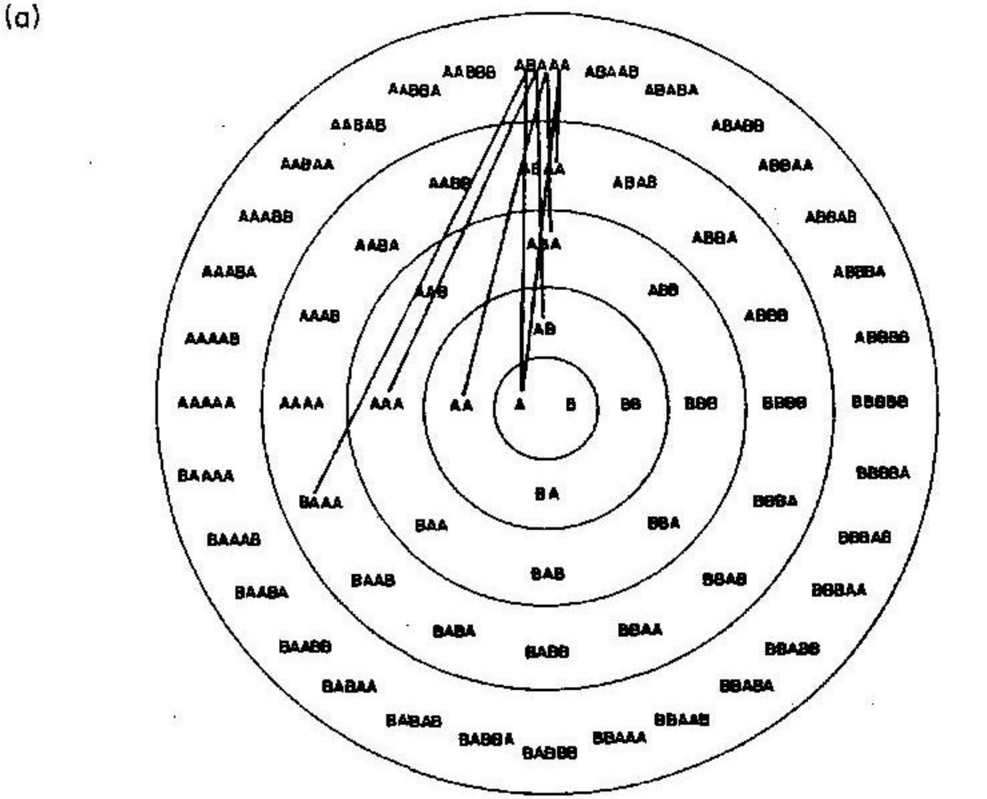


FIG. 2. (a) Pairs of lines meet at the juncture of successive monomers in ABAAA, and lead to those smaller pairs of polymers which form ABAAA by condensation. (b) Lines connect dimer BA with A and B, forming BA by condensation, and with each larger polymer which yields BA by cleavage.

Cleavage

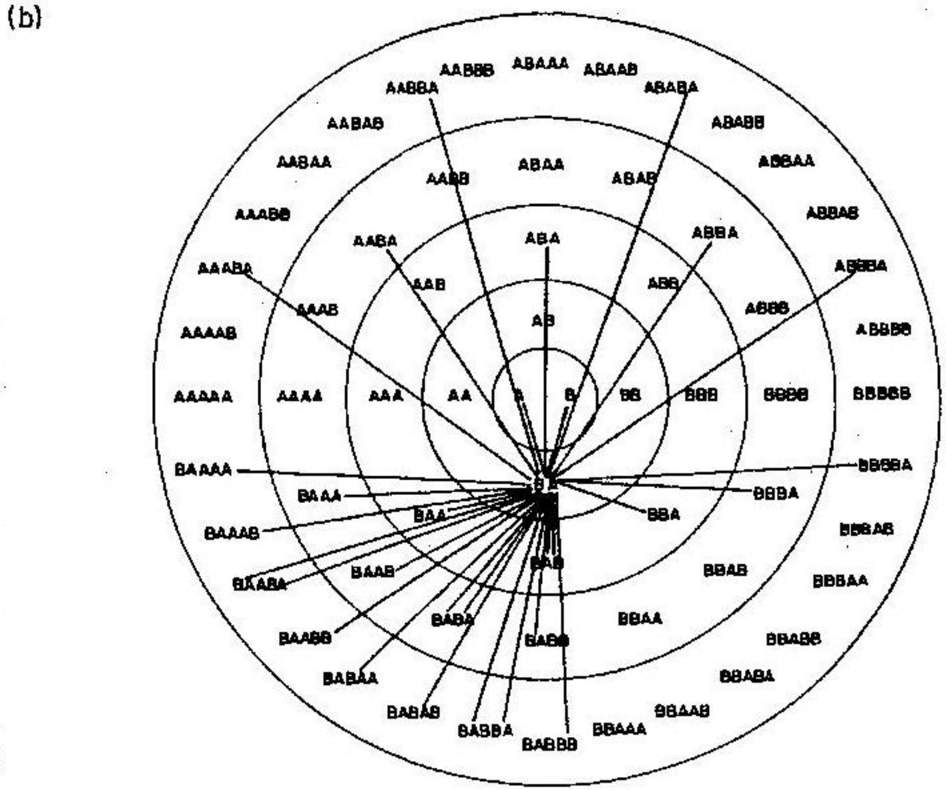


FIG. 2. (a) Pairs of lines meet at the juncture of successive monomers in ABAAA, and lead to those smaller pairs of polymers which form ABAAA by condensation. (b) Lines connect dimer BA with A and B, forming BA by condensation, and with each larger polymer which yields BA by cleavage.

New Era of Experimental Work

- Cloning technology
- Discovery of molecular structure of genes renews interest in the origin of life.

Ontogeny - Origins of Life

- **Processes involved**
  - **Cell differentiation (Jacob and Monod won Nobel Prize mid-1960s)**
  - **Morphogenesis**
- **Hypothesis**
- **From buttons and threads to chemicals**
- **Phase transition**
- **Boolean network types**
  - **Order**
  - **Chaos**
  - **Edge of chaos (networks are both stable and flexible)**
- **Homeostasis**
- **Holism**
  - **Property of autocatalytic sets**
- **Sources of order**
  - **Selection**
  - **“Order for Free” (Self-organization that arises naturally)**

Hypothesis

- Life started with the spontaneous replication of molecules (autocatalytic nets).
- Huge problem with hypothesis:
  - ❖ Why do all living organisms seem to have a minimal complexity below which it is impossible to go?
  - ❖ Answer: Just because!
- Kauffman’s theory
  - ❖ “Matter must reach a certain level of complexity in order to spring into life. This threshold is not an accident of random variation and selection; I hold that it is inherent to the very nature of life.”

Buttons and Threads

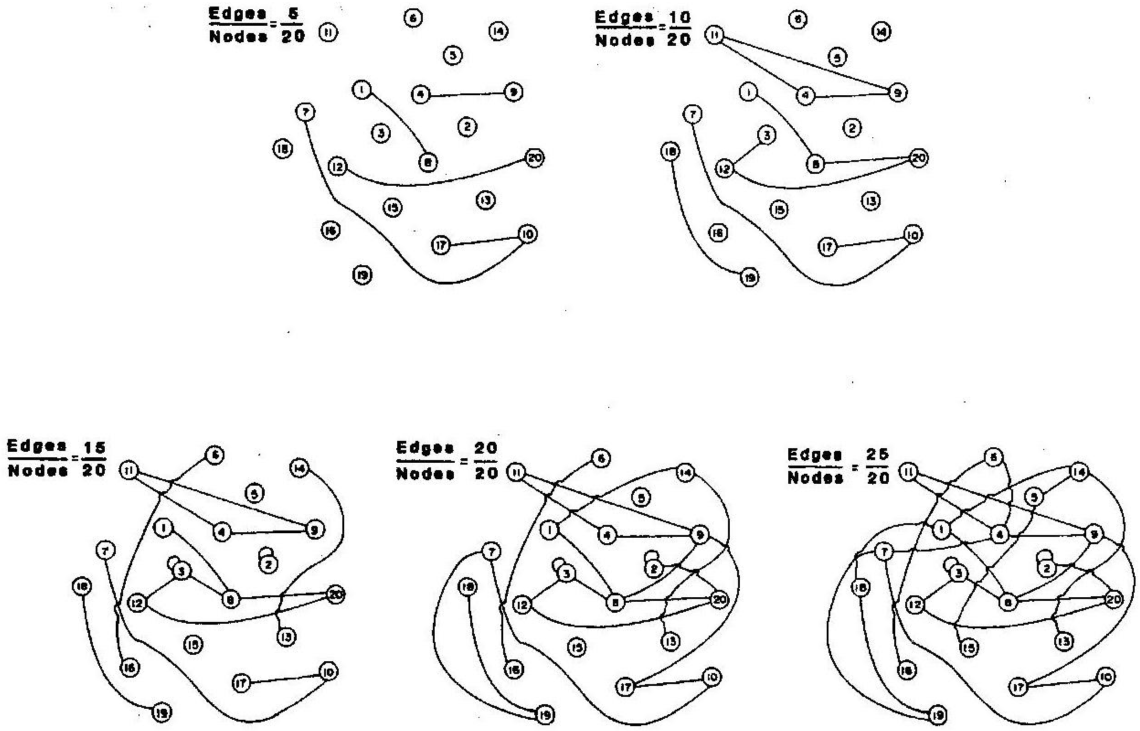


Figure 3.3 *Crystallization of connected webs. Twenty "buttons" (nodes) are connected at random by an increasing number of "threads" (edges). For large numbers of buttons, as the ratio of threads to buttons increases past a threshold of 0.5, most points become connected in one giant component. As the ratio passes 1.0, closed pathways of all lengths begin to emerge.*

Phase Transition

- Threshold where network stability and flexibility are achieved through a balanced state.
- Edge of chaos
- Most complex behaviors are just near the phase transition.

Phase Transition

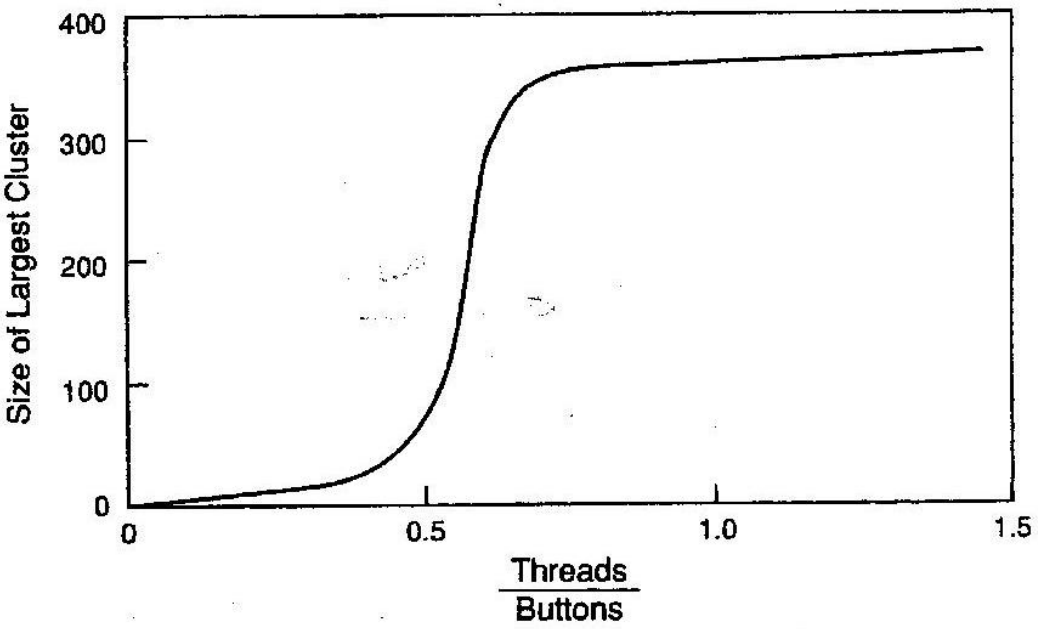
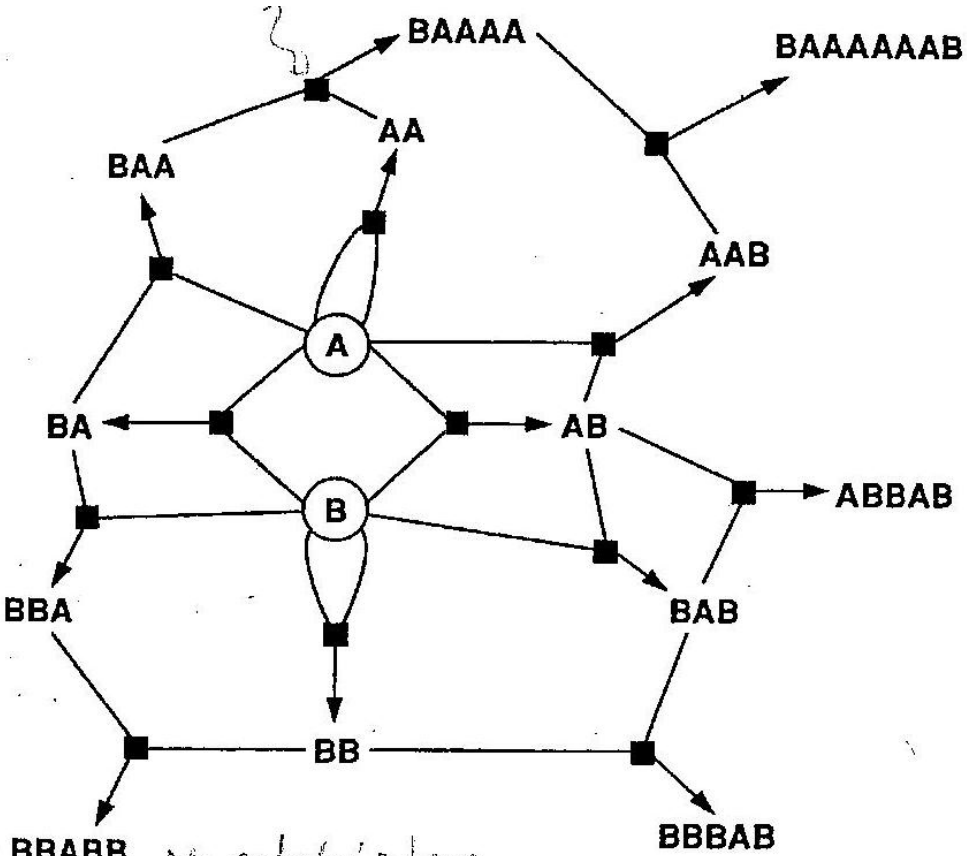


Figure 3.4 *A phase transition. As the ratio of threads (edges) to buttons (nodes) in a random graph passes 0.5, the size of the connected cluster slowly increases until it reaches a "phase transition" and a giant component crystallizes. (For this experiment, the number of threads ranges from 0 to 600, while the number of buttons is fixed at 400.)*

Reaction Graph
No Catalysts



No catalysts here

Figure 3.5 *From buttons and threads to chemicals. In this hypothetical network of chemical reactions, called a reaction graph, smaller molecules (A and B) are combined to form larger molecules (AA, AB, etc.), which are combined to form still larger molecules (BAB, BBA, BABB, etc.). Simultaneously, these longer molecules are broken down into simple substrates again. For each reaction, a line leads from the two substrates to a square denoting the reaction; an arrow leads from the reaction square to the product. (Since reactions are reversible, the use of arrows is meant to distinguish substrates from products in only one direction of the chemical flow.) Since the products of some reactions are substrates of further reactions, the result is a web of interlinked reactions.*

Reaction Graph - Catalysts

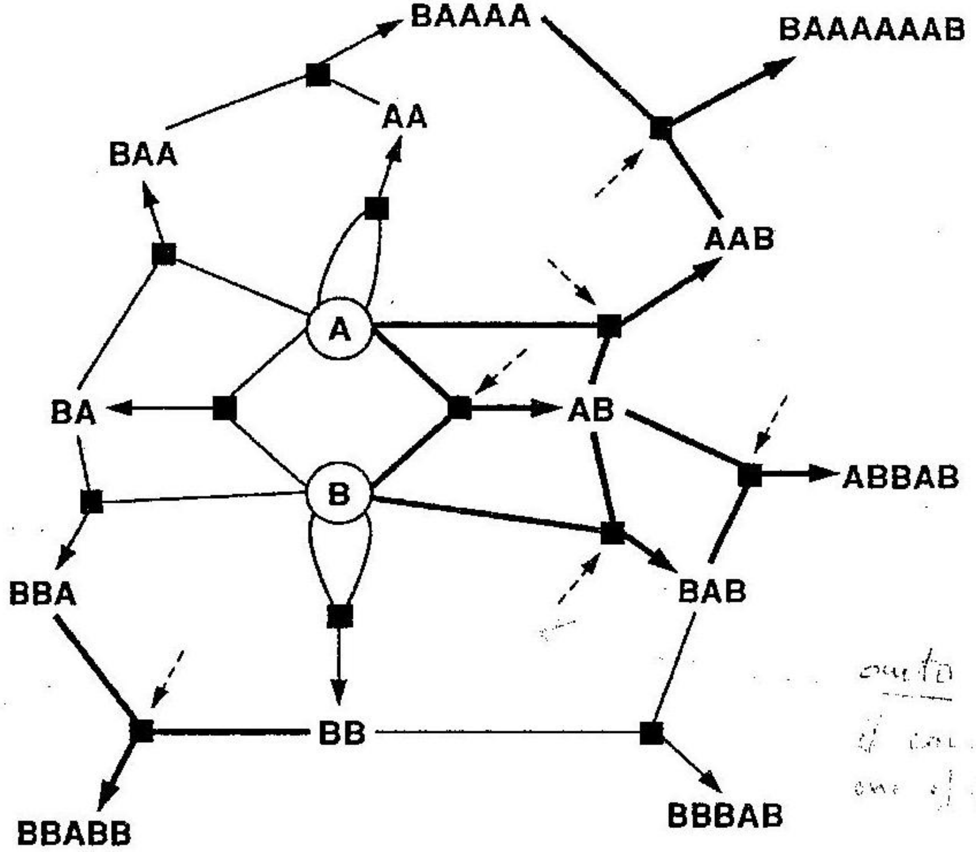


Figure 3.6 *Molecules catalyzing reactions. In Figure 3.5, all the reactions were assumed to be spontaneous. What happens when we add catalysts to speed some of the reactions? Here the reaction squares indicated by dashed-line arrows are catalyzed, and the heavy, darker lines connect substrates and products whose reactions are catalyzed. The result is a pattern of heavy lines indicating a catalyzed subgraph of the reaction graph.*

Boolean Networks

- Invented by Kauffman
- Include attractors, each draining some basin of attraction.

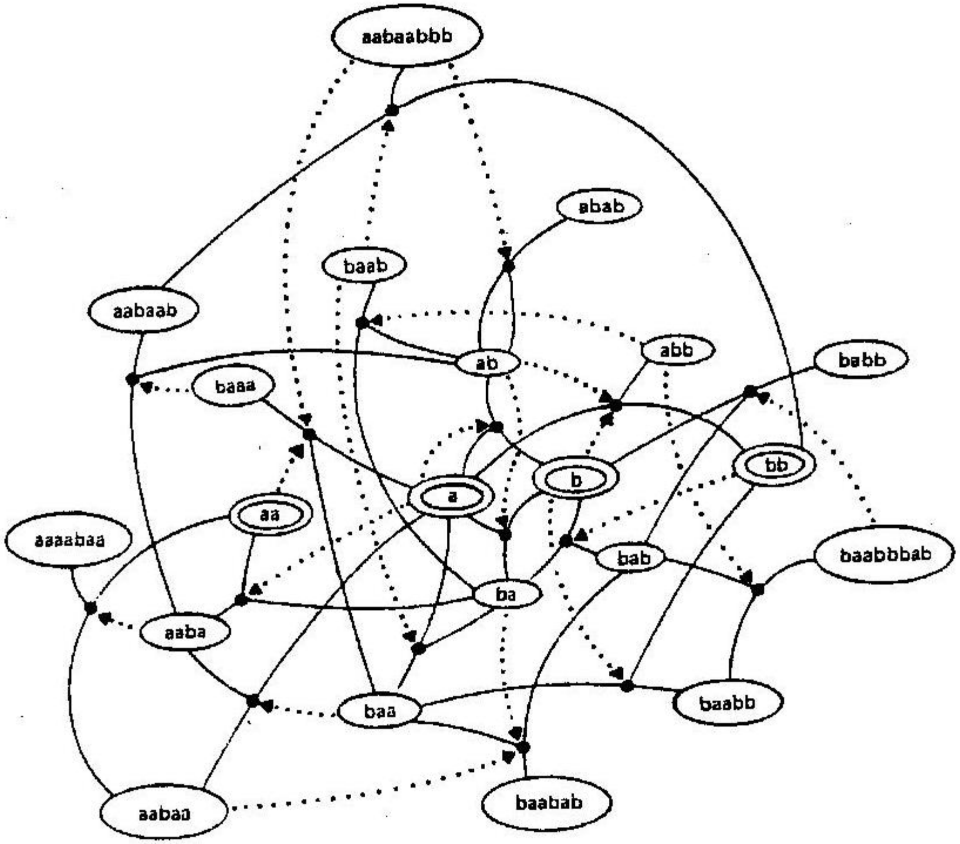
Homeostasis

- The tendency of cell types to remain the same, owing to the coordinated response of its parts, following perturbations (any situation or stimulus tending to disturb its normal condition or function).
- In other words, an autocatalytic net is resistant to small perturbations.
- Attractors are the ultimate source of homeostasis, and ensure a stable system.
- Most perturbations leave the system in the same basin of attraction. Thus…the system returns to the same state cycle from which it was perturbed and we have homeostatic stability.

Holism

- The theory that whole entities, as fundamental components of reality, have an existence other than as the mere sum of their parts.
- In other words, life emerged whole, at once, not piecemeal.

Autocatalytic Set
Catalysis



◎ = food set

○ = other chemicals

⅄ = reactions

◂······ = action of catalysts

Figure 3.7 *An autocatalytic set. A typical example of a small autocatalytic set in which food molecules (a, b, aa, bb) are built up into a self-sustaining network of molecules. The reactions are represented by points connecting larger polymers to their breakdown products. Dotted lines indicate catalysis and point from the catalyst to the reaction being catalyzed.*

Conclusion

- Random networks are formed.
- Behavior
  - ❖ Order and stability
  - ❖ Cell replication
  - ❖ Cell differentiation
  - ❖ Possibility of a general theory of metabolic behavior.

Questions and Answers

- Open discussion...