Ancestors: Commentaries on *The Global Dynamics of Cellular Automata* by Andrew Wuensche and Mike Lesser (Addison-Wesley, 1992)

Harold V. McIntosh
Departamento de Aplicación de Microcomputadoras
Instituto de Ciencias, Universidad Autónoma de Puebla
Apartado Postal 461, (72000) Puebla, Puebla, México.

July 20,1993

Abstract

The collection of commentaries on the book: Andrew Wuensche and Mike Lesser, *The Global Dynamics of Cellular Automata,* Addison-Wesley, 1992 (ISBN 0-201-55740-1) which were posted on CA-MAIL during June and July, 1993, is reproduced with the correction of misspellings and adaptation to TEX format. Citations to some of the references mentioned have been included.

1 ancestors (1)

Last fall and this spring various discussions have taken place regarding Andrew Wuensche (`<100020.2727@COMPUSERVE.COM>`) and Mike Lesser's new book, "The Global Dynamics of Cellular Automata." Not having had a copy of the book to refer to has precluded making any commentary on its subject matter, but a copy has now arrived in Puebla (Puebla has a splendid Colonial Cathedral, is near to a World-Famous Pyramid (Cholula), but you will look in vain to find a University Bookstore).

To begin with, the book is a real work of art, with something like 200 pages of carefully drawn evolution diagrams, for binary rules with 3 and 5 neighbors. All the symmetry classes of the former, and mostly the totalistic rules of the latter are shown, for rings of up to 15 cells. All in all, a tremendous collection of data, a vastly expanded version of Holly Peck's Table 13 in Wolfram's "Theory and Applications of Cellular Automata[1] ."

[1]Stephen Wolfram, Theory and Applications of Cellular Automata, World Scientific Press, Singapore, 1986

There is no telling how often, or how many people, have thumbed through that appendix, looking for examples of something or other.

Just as many delicacies serve as appetizers, not constituting the whole meal, this valuable collection may serve more to whet our interest, while it satisfies our curiously, than to offer us the Final Word. But then, no one wants to propose that someone publish a 10,000 page atlas, just to keep our interest going!

One of the first things which come to mind are the theories of random graphs of Erdös, Bollobás[2] and others. Evolution diagrams are trees rooted on cycles, so we know beforehand that there will be connected components (the different cycles) and no loops otherwise. We also know that the graphs must have the symmetry of the ring, so that there will be cyclical and reflective repetition of structures.

Within those constraints, the statistics which can describe the graphs are: average branching ratio, average length of transients, maximum and minimum values of these two quantities, and their variances. According to random graph theory, links should distribute fairly uniformly over the nodes (insofar as constraints allow, and one constraint is — one and only one out-link per node). More than that, the distribution is Poisson-like, so that the actual number of links is rarely the exact average, but nearby according to the well-known formula.

As a first reaction, based on my own experience, it might be interesting to comment on a study of (2,1) Rule 22, which is a sort of one-dimensional version of Life, which we made several years ago. Somehow, it did not seem that the rings became interesting until their circumference reached 20; from that point on several alternative structures showed up having the same period, and structures began to have a greater variety in general. Indeed, it was at this point, with the help of several incisive observations on the part of Robert Wainright, that we discovered how de Bruijn diagrams could be used to deduce the possible periodic configurations of a one-dimensional automaton.

We carried out a complete analysis of cycles up to rings of circumference 34. Two things happened; first, 2^{34} is 16 billion, and the analysis took months on two or three microcomputers running in parallel (2MHz 8080's). So adding another cell would have taken twice as long still. But vestiges of a second phenomenon began to appear; for shorter rings, periods in the tens, maybe hundreds showed up. But at n=34, periods began to run in the thousands and beyond. That could be confirmed, because certain configurations for Rule 22 are very regular, and could be checked explicitly. In fact, one suspected that a further great jump might be waiting at N=66, and at related values thereafter.

Actually, the literature contains some other instances where rather extensive results are available, namely for rules like the exclusive or's (which are equivalent to finite fields), where matrix theory gives pretty complete results.

What this means with respect to the Atlas, is that in spite of the wealth of data which it contains, it may just be skimming the surface of a large reservoir of interesting automata. The foregoing comments suggest that it may be fairly adventurous to extrapolate from small rings; but if one is forewarned, the small rings can still be used to good advantage.

The Atlas contains much more than just the evolutionary diagrams; one of its most valuable features may be the comparisons which it suggests, and documents to a good extent, between automata whose rules are similar. One always suspects that similar rules

[2] Béla Bollobás, Random Graphs, Academic Press, London, 1985

should produce similar automata. Lenore Levine[3] has described an interesting sequence of rules, which deviate less and less from Rule 128 - OR, from a topological point of view. With the programs which accompany the Atlas, such ideas can be tried out, and their results evaluated.

More commentary will follow.

2 ancestors (2)

Here we continue an analysis of Andrew Wuensche and Mike Lesser's new book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1). It is a wonderful compilation of data, and we have already commented on its complete coverage of evolution diagrams, for (2,1) automata on rings of circumference up to 15, and of selected (2,2) automata.

The first reaction of a person who intends to compile evolution diagrams might be to make a systematic list of all the configurations, say of the 2^{15} configurations on a 15-member ring; 2^{15} is only 32K, and is still relatively manageable. An evolution graph can be created by applying the evolution rule to each configuration, observing the configuration into which it evolves, and noting the result in an array of links.

Many interesting statistics of the evolution diagram are available as matrix elements, vector products, or quadratic forms of this matrix. For example. the trace of its powers can be made to yield the number of loops with the corresponding length. Multiplying by a row of 1's gives a vector containing the number of out links for each node. And so on.

Many people won't even go to that much trouble; if the cycles of the automaton are all that is wanted, you just follow the evolution until the configuration repeats, and note where the loop began. By treating the configurations as binary numbers, and abandoning a search when it is found that the numbers aren't ordered satisfactorily, search time can be reduced; likewise there are sometimes advantages to generating the configurations in Gray Code order.

None of this is what Wuensche and Lesser recommend; rather, they explain a method for calculating ancestors, rather than descendants. Obviously one is not going to get the upper branches of an evolution tree by working downward, and it is impractical to compare each evolution chain with all the others to see if it has the highest nodes. The solution is to be able to work in both directions: start somewhere, go to the bottom, then work upward, identifying everything connected to the bottom. For the next tree, don't look at anything that belongs to the first one. And, in the bargain, there is no messy matrix algebra.

Calculating ancestors in a one dimensional automaton is pretty straightforward. Decide on your configuration, and start somewhere, say at the left end. By the rule table, you know what neighborhoods are going to give the first cell, so make a list of them. Now look at the second cell, whatever it is. It has its own ancestral neighborhoods, but they overlap the neighborhoods of the first cell. Discard the ones which don't match, and extend your list. But do it systematically, trying to extend the first neighborhood of the first cell in all possible ways before turning to its second neighborhood. This is a nice recursive process

[3] Lenore Levine, Regular Language Invariance under One-Dimensional Cellular Automaton Rules, Complex Systems 6 163-178 (1992).

with a simple search strategy. In fact, it is a game of dominoes, or rather, the analysis of all possible domino games.

As extensions are made, the lists may proliferate, maintain themselves, or die out. Eventually the right end is reached; when the configuration lies on a ring, the question is whether the first cell can still be used. If so, an ancestral configuration has been found. This, amongst other things, is what we learn how to do in Chapter 3 of "Global Dynamics." How many ancestors there are altogether is related to the Z-parameter, in ways that we propose to examine later on.

This is the place at which it seems that the use of some matrix theory might be useful. Once again, the de Bruijn diagram is important, and the matrix which describes its connectivity. Think of a (2,1) automaton, and "halves" of neighborhoods. There are four, 00, 01, 10, and 11. Let these be nodes in a graph, and let the links assert that the parts overlap to make a full neighborhood. Then 00 links to 01, and even to 00, but not to 10 nor 11. In fact, there are always two out links and two in links at each node, and we might as well label them according to the full neighborhoods. Thus link 010 joins node 01 to node 10 to form the neighborhood 010, which evolves into a 1 according to rule 22.

Make up two matrices, one for neighborhoods which evolve into 0's, one for neighborhoods which evolve into 1's. For Rule 22 we get:

		00	01	10	11			00	01	10	11
	00	1	0	.	.		00	0	1	.	.
for 0:	01	.	.	0	1	1:	01	.	.	1	0
	10	0	1	.	.		10	1	0 .	.	.
	11	.	.	1	1		11	.	.	0	0.

The dots in these matrices are always zeroes, indicating halves which won't join to make a neighborhood in the first place (dominoes whose spots don't match). The 0's and 1's don't mean that is the cell you get, but are boolean no's and yes's, that you will get a 0 if you are using the 0-matrix, or a 1 if you are using the 1-matrix, when that neighborhood evolves.

What makes these matrices really useful, is that when they are multiplied, they tell what kinds of chains can be formed; the rules for multiplying matrices (and the fact that zeroes and ones are being used) just end up counting the number of paths from the row node to the column node. And if a product matrix is zero, that tells that there aren't any paths at all - no ancestor, a poor little orphan, and presto! the Garden of Eden has a new resident.

Try multiplying out the matrices for 10101001. Worse yet, look at the 1-matrix for the Zero-Rule.

More commentary will follow.

3 ancestors (3)

We continue our commentary on Andrew Wuensche and Mike Lesser's new book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1). In previous episodes, the role of diagrams in describing the evolution of automata has been mentioned, especially the evolution diagram, (of which their computer program generates the nice examples shown in the Atlas), and the de Bruijn diagram, which they do not mention, but which can be used to calculate ancestors.

For Wolfram's (2,1) Rule 22, this leads to a pair of 4×4 matrices,

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note that both are partially diagonal, although in different ways. For A this says that a zero neighborhood is a counterimage of zero, but then if you try to extend it, you can't ever have any ones. Of course; ones are expansive for Rule 22. The other component has maximum eigenvalue 1.459, which is the growth factor for long configurations which vanish instantly. The configurations which it generates had better not have pairs of zeroes. The greatest growth factor for any rule is 2.0, the maximum row or column sum for these 4x4 de Bruijn matrices.

The second matrix, B, has a cube root of the unit matrix in its upper left hand corner. This means that you can never use the neighborhood 111 in a counterimage of pure ones, and that the fragments 01, 10, and 00 must always run in cyclic order. The eigenvalues of this corner have absolute value 1, so the number of counterimages of pure 1's is always the same, no matter how long or short the ring.

In general, it is arbitrary configurations whose ancestors are sought, not pure strings. However, we know that there is a norm for matrices, related to the absolute value of their largest eigenvalue, following which the norm of a product is always less than the product of the norms (but equal when the factors commute). This means that the state with the largest fraction of ancestors is always going to get the lion's share of the ancestors, exponentially following that majority's share of the configuration. So it is, that configurations in Rule 22 have few ancestors unless they have lots of zeroes, as we see by turning to page 96 of the Atlas. .

The de Bruijn pair for the famed stochastic Rule 30 is

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note that Rule 30 is a one-bit mutant of Rule 22, differing only in the behavior of neighborhood 011, and thus the assignment of the link 01 - 11. However, the Perron Eigenvalues of both matrices are 1, so that the number of ancestors will be expected to remain constant for all configuration lengths. Compare the attraction basins on page 144 of the Atlas.

Now we have set the stage for thinking of the possible relation of the Z-parameter and Langton's lambda parameter to the matrices derived from the de Bruijn diagram. Nevertheless we should introduce the characters and outline the plot before proceeding with the show.

The nodes in the de Bruijn diagram are called "start strings" in section 3.4.1 of the Atlas, left or right according to the direction of arrows connecting them. In all ancestor calculations, the possible variety of start strings lends an ambiguity which tends to persist;

reversible rules are only possible when this ambiguity can be shown to be irrelevant, even when all the other factors are favorable.

When every row (alternatively, every column) of a de Bruijn diagram contains a single 1, we have one of the "limited preimage rules." By Gerschgorin's theorem, the eigenvalues of such matrices are bounded by 1 (and according to Perron, the maximum eigenvalue is exactly 1), all of which is in accordance with the experience that counterimages not proliferate excessively, thereby justifying the adjective chosen in the Atlas.

Note that the eigenvalues represent an asymptotic rate of growth, and that boundary conditions have to be taken into account. Thus there are Gardens of Eden in the Atlas even when growth factors are unity. Generally we would expect that if some populations grow, others languish, to maintain constant the total number of configurations. But in finite systems, the constraints are more exacting, and can produce an occasional vanishment which might not be found in an infinite system.

Cyclic boundary conditions correspond to the diagonal of the de Bruijn matrices, because the "start string" is the same as the "stop string." Individual matrix elements correspond to choosing one particular start, and one particular stop. For infinite chains all elements must be considered, whereas for configurations which are "quiescent at infinity," the (q,q) element is the relevant one (q the quiescent state).

It should be explained that the theory here outlined is inherent in the Atlas' list of references, particularly in the work of Erica Jen[4] there cited, and in articles of Stephen Wolfram[5], also cited. The principal difference is that Jen's work is phrased in terms of recursion relations, not matrix theory (although she exhibits them and describes their use). Furthermore, the nicest theorems do not seem to result from matrix theory alone. The advantage of a matrix-oriented point of view is that it leads naturally to eigenvalues and eigenvectors, or whatever it is that they signify in the particular application. Here, it is rates of growth in the number of counterimages with respect to the length of the ring of cells. Eigenvectors are less important (the principal eigenvector can be scaled to get positive real probabilities), but they would have exact rates of growth.

More commentary will follow.

4 ancestors (4)

Commentary on Andrew Wuensche and Mike Lesser's new book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1) continues. We have discussed how to form a de Bruijn diagram for a cellular automaton rule, its connectivity matrix, and how evolution splits the matrix into submatrices, which can be used to count ancestors. The rows and columns of the de Bruijn matrix are labelled by Wuensche and Lersser's "start strings," but the columns are really "stop strings."

To fix ideas, consider the two matrices for Wolfram's (2,1) Rule 252, which is associated with rules 3, 17, 63, 119, 238, 192, and 236 in the Atlas, on pages 128 and 129.

[4]E. Jen, Enumeration of Preimages of Cellular Automata, Complex systems 3 421-456 (1989).

[5]Stephen Wolfram, Computation theory of cellular automata, Communications in Mathematical Physics 96 15-57 (1984).

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

At first sight, the basins for Rules 3 and 252 look completely different, but an examination of these matrices shows why it is sensible to consider them jointly (actually, to pair 252 and 192) - their only difference lies in exchanging the A and the B matrices. Therefore, they will have identical statistics, with 1's and 0's exchanged. In other words, ancestors will proliferate similarly, although their cycles and periods may be different.

The other two operations which the authors consider, reflection and complementation, also have their repercussions. For example, reflection exchanges the start neighborhoods 01 and 10, while complementation does this, exchanges 00 with 11, and exchanges A with B. None of these things changes norms or eigenvalues. Persons familiar with matrix theory will see that the matrices representing these exchanges will commute with the A, B pair, and that they will constitute a symmetry group. A sort of supersymmetry can be achieved by making a bigger matrix having A and B as submatrices, but we will not make further use of the possibility.

There is an extensive theory of positive matrices, just as there is of integer matrices. Matrices in general can be associated with graphs; their nodes are the indices, which are linked according to whether the matrix elements are zero or not. This means that nonzero elements in the product of matrices are associated with chains in the diagram. The relationship is useful when some properties are more evident in one context than in the other. For example, the matrix is partially diagonal when the graph consists of two disjoint parts. Blocks of zeroes correspond to attractors (into which links may enter but cannot leave) and dispersors (links leave but do not enter). Conversely, statistics concerning the graph often have nice matrix formulas, and in general properties of graphs can be worked out in a computer whenever they have a formulation via matrix algebra.

Two important properties of matrices are, on the one hand, their eigenvalues and eigenvectors, and on the other, their norms. The two are related, but the relationship is more complicated when the matrices are not symmetric, so that attractors and dispersors can be present. The norm is not a perfect "absolute value" because the norm of a product is only less, not necessarily equal, to the product of norms. For purposes of analysis and calculating limits, the inequality is entirely satisfactory, but it is less favorable when exact counts are required.

The matrices A and B count the ancestors of a single cell, and catalog them according to the start and stop strings constituting the neighborhood. It is a fundamental reality of cellular automata theory that there are always more cells amongst the neighborhoods than the number of cells being considered; this shows itself when we use a matrix rather than a scalar to do our counting. If A and B count the ancestors of a single cell, we expect their products to count the ancestors of a sequence of cells. A matrix is still called for, because marginal cells always remain, however long the chain.

The de Bruijn matrices for (2,1) automata have both norm 2, and largest eigenvalue 2, and these quantities are always larger than those of any of the fragments into which the matrices decompose. The value 2 corresponds to doubling the total number of configurations every time a single cell is added to the automaton.

Consider the formulas for counting (within the limits of pure ASCII). If $\mathbf{M}$ is the connectivity matrix of a graph, let $\mathbf{u}$ be a vector of ones, and $\mathbf{i}$ be a unit vector in the i^{th} coordinate. Let ^{T} designate trasnspose. Then $\mathbf{i}^T\mathbf{Mj}$ is the ij^{th} element, the number of links from i to j, while $\mathbf{i}^T\mathbf{Mi}$ is the number of loops starting at i. Their sum is the Trace of $\mathbf{M}$, yielding the number of loops altogether (with a possible multiplicity according to their length). The product $\mathbf{u}^T\mathbf{Mu}$ is the number of links altogether, no restriction. $\mathbf{q}^T\mathbf{Mq}$ is the number of paths beginning and ending with a quiescent state, $\mathbf{q}$.

All three of these formulae can be written as traces, in the form Tr($\mathbf{GM}$), with a suitable metric matrix $\mathbf{G}$. To count everything, $\mathbf{G}$=$\mathbf{uu}^T$; for periodicboundary conditions, $\mathbf{G}$=$\mathbf{I}$, the unit matrix, and for configurations quiescent at infinity, $\mathbf{G}$=$\mathbf{qq}^T$. Among other things, this means that the choice of a boundary condition is not very essential to a calculation. It only enters at the last moment, in the selection of the metric matrix. But the essential qualities of the matrix, as represented by its eigenvalues and eigenvectors, expressing such things as rates of growth, are not affected by the boundary condition.

Suppose that we want to count configurations. We must add A and B, which always results in the de Bruijn matrix D (which has a rather simple characteristic equation - $D^{k+1} = kD^k$, because $D^k = uu^T$, $Duu^T = kuu^T$). So, for cyclic boundary conditions in a ring of circumference s, $Tr(D^s) = 2^s$, which is so unsurprising that it might be considered boring. But it is almost the only result that we are going to get free.

Counting is good for getting averages. But suppose we want variances. Then it is necessary to sum squares. That is where matrix theory is really going to shine. We need $(Tr(GM))^2$, which turns out to be $Tr(GM \otimes GM)$, which in turn is $Tr((G \otimes G)(M \otimes M))$. Extracting the constant factor $G \otimes G$, we have to calculate $Tr(M \otimes M)$. Here $\otimes$ indicates a tensor product, which is a way of compounding matrices that will be found in books on matrix theory[6].

A certain amount algebraic manipulation ends us up with a formula for the second moment when the de Bruijn matrix is split into A and B; namely we need a trace involving $(A \otimes A + B \otimes B)^s$, (rather than $(A+B)^s$) in a ring of circumference s. These commentaries aren't the place for mathematical derivations, but the details are available in a pair of preprints[7] that anyone can have by sending a mailing address (including zip code, city, country, etc).

A tensor square has an interpretation as a graph; it is nothing other than the graph of ordered pairs taken from the graph of the original matrix. (There is also a symmetrized tensor square, and an unordered-pair graph). This relationship is intimately related to Niall Graham (`niall@nmsu.edu`)'s assertion of 21 Jun 93 14:55:27:

```
>
> A 1-dim CA is reversible iff the pair digraph of its
> associated finite automaton is acyclic.
>
```

More commentary will follow.

---

[6]For example, F. R. Gantmacher, The Theory of Matrices, Chelsea Publishing Company, New York, 1959.

[7]1) Harold V. McIntosh, Linear Cellular Automata via de Bruijn diagrams (May, 1991), and 2) Harold V. McIntosh, Reversible Cellular Automata (January, 1991).

5 ancestors (5)

Commentary on Andrew Wuensche and Mike Lesser's new book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1) continues. For the moment we are laying a groundwork of matrix and graph theory, rather than discussing the book explicitly; reference to the book will still provide examples for statements which will be made.

Much of the theory of cellular automata, and especially of one dimensional cellular automata, can be described with the aid of de Bruijn diagrams and their associated connectivity matrices. The calculation of ancestors (for a binary automaton) can be accomplished by using a pair of these matrices, one for each of the two states into which a neighborhood can evolve.

To write down a de Bruijn matrix quickly, and to interpret one rapidly, use Wolfram's scheme for enumerating automata. Suppose that the rule is

$$\begin{array}{cccccccc} 111 & 110 & 101 & 100 & 011 & 010 & 001 & 000 \\ h & g & f & e & d & c & b & a. \end{array}$$

The corresponding positions in the connectivity matrix are

$$\begin{bmatrix} a & b & . & . \\ . & . & c & d \\ e & f & . & . \\ . & . & g & h \end{bmatrix}$$

where all elements marked by dots are filled with zeroes. The length of the "steps" in the matrix correspond to the number of states in the automaton, while the number of steps reflects the size of the neighborhood, as does the number of "flights of stairs;" the two are always equal.

Large de Bruijn diagrams are hard to draw, having so many nodes and links. The best visualization we have found is just to draw a circle, divide the circumference into the requisite number of nodes, and treat them as though they were k-adic numbers modulo the total number of start strings. All but the simplest are still quite congested. Artistically, they have a pleasing structure.

From the basic de Bruijn diagram, others may be derived. One is the subset diagram, whose elements are subsets of nodes from the primitive diagram. the concept was introduced in the 1950's by E. L. Moore, who was interested in experiments which could identify the state of an automaton, or to place it in an arbitrarily prescribed state. Its relation to calculating ancestors is a consequence of offering a sure and easy way to ascertain whether the primitive diagram contains a given path or not.

To create a subset diagram, link two nodes if there is a link from at least one node in the source (tail) subset to all the nodes in the destination (head) subset. Or in other words, take a source subset and run through all the nodes to which its members are linked; that collection is the destination. When there are no such links, the subset is connected to the empty set. That way a uniform quota of links is guaranteed for each node in the subset diagram, so that arbitrary paths can always be found; but getting trapped at the empty set is always a possibility.

With respect to ancestors, not caring about the start string means beginning at the full set. Any path leading from there to the empty set implies the lack of an ancestor, and thus a non-empty Garden of Eden. Many conclusions can be drawn, for example, the shortest ancestorless string, the longest loopless string with ancestors, and so on. And, of course, if there is no path at all, there is no Garden of Eden.

Since the subset diagram is naturally ordered, not finding a Garden of Eden leads to some natural questions: what is the largest set that still leads to the null set? what is the smallest set still reachable from the full set? and so on. G. A. Hedlund[8] and his followers have studied these questions in much detail.

Although the subset diagram reveals the existence of ancestors, it is not very helpful in identifying them, because it is not so easy to backtrack.

For Wolfram's (2,1) Rule 22, the subset diagram has 16 elements, which we may rank by size from the full set to the empty set.

Connection matrix:

$$\left[\begin{array}{c|cccc|cccccc|cccc|c}
1 & 1 & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ \hline
. & 1 & 1 & . & . & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & 1 & . & . & . & . & 1 & . & . & . & . & . & . & . \\
1 & . & . & . & . & 1 & . & . & . & . & . & . & . & . & . & . \\
. & . & . & . & 1 & . & 1 & . & . & . & . & . & . & . & . & . \\ \hline
. & . & . & . & . & . & . & 1 & 1 & . & . & . & . & . & . & . \\
. & . & . & . & . & 2 & . & . & . & . & . & . & . & . & . & . \\
. & . & . & 1 & . & . & . & . & . & . & . & . & 1 & . & . & . \\
. & . & . & . & . & . & 1 & . & . & 1 & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & 1 & . & . & 1 & . & . \\
. & . & . & . & 1 & . & . & . & . & . & . & 1 & . & . & . & . \\ \hline
. & . & . & . & . & . & . & . & . & . & . & 1 & 1 & . & . & . \\
. & . & . & . & . & . & . & . & . & . & . & . & . & 1 & 1 & . \\
. & . & . & . & . & . & . & . & . & . & . & 1 & 1 & . & . & . \\
. & . & . & . & . & . & . & . & . & . & 1 & . & . & . & . & 1 \\ \hline
. & . & . & . & . & . & . & . & . & . & . & . & . & . & . & 2
\end{array}\right]$$

Noteworthy details: The 4x4 unit-class submatrix in the lower right hand corner is almost a de Bruijn diagram, because most of its nodes have continuations with either a 0 or a 1, but the start string 11 leads to both 110 and 111 to form a neighborhood which evolves to 0, so it links to a two-element subset via 0 and the empty set via 1. That explains the next-to-last row. This is also the only direct linkage to the empty set in the matrix.

In fact, the eighth is the smallest power of the matrix with a direct link from the full set to the empty set, corresponding to the "poison word" 10101001 (and of course, its reflection, 10010101). This is easier to appreciate by drawing the diagram, without a computer program which can display the matrix and its powers. For any Rule, by the 16th power, a decision will necessarily have been reached, as to whether the (full, empty) matrix element is zero or not. (But, the de Bruijn diagram only had 4 nodes)

Primitive diagrams have another kind of derivative, the pair diagram (and more generally, the n-tuple diagram); both ordered pairs and unordered pairs may be considered.

[8] G. A. Hedlund, Endomorphisms and automorphisms of the shift dynamical system, Mathematical Systems Theory 3 320-375 (1969).

Consider ordered pairs; (x,y) is linked to (X,Y) if x is linked to X AND y is linked to Y. A pair diagram is a kind of greatest lower bound between its constituents (a greatest common denominator, if you wish). Unordered pairs might be more convenient if both members were taken from the same set, such pairs are also subsets.

The most evident application of a pair diagram lies in testing whether or not the same sequence of nodes can be found in two different places in the primitive diagram, and therefore is related to establishing uniqueness. If two paths exist, lay them out side by side, pair their corresponding links, and get the path in the pair diagram. Conversely, any path in the pair diagram may be decomposed into its constituents.

A more subtle application of the pair matrix lies in calculating variance. If the powers of the connection matrix of a graph reveal the number of paths of the corresponding length between nodes, then the powers of the pair matrix reveal the square of this number, or in other words, the second moment. From there to the variance is an exercise in elementary statistics. The application to counting ancestors lies in the observation that if some configurations have more ancestors, then others have fewer, which must result in a non-zero variance. Still, the connection is not easy to prove.

For Wolfram's (2,1) Rule 22, the pair diagram has 16 elements, thus a 16x16 connection matrix,

$$\left[\begin{array}{cccc|cccc|cccc|cccc}
1 & * & . & . & * & 1 & . & . & . & . & . & . & . & . & . & . \\
. & . & * & 1 & . & . & 1 & * & . & . & . & . & . & . & . & . \\
 & 1 & . & . & 1 & * & . & . & . & . & . & . & . & . & . & . \\
. & . & 1 & 1 & . & . & * & * & . & . & . & . & . & . & . & . \\
\hline
. & . & . & . & . & . & . & . & * & 1 & . & . & 1 & * & . & . \\
. & . & . & . & . & . & . & . & . & . & 1 & * & . & . & * & 1 \\
. & . & . & . & . & . & . & . & 1 & * & . & . & * & 1 & . & . \\
. & . & . & . & . & . & . & . & . & . & * & * & . & . & 1 & 1 \\
\hline
* & 1 & . & . & 1 & * & . & . & . & . & . & . & . & . & . & . \\
. & . & 1 & * & . & . & * & 1 & . & . & . & . & . & . & . & . \\
1 & * & . & . & * & 1 & . & . & . & . & . & . & . & . & . & . \\
. & . & * & * & . & . & 1 & 1 & . & . & . & . & . & . & . & . \\
\hline
. & . & . & . & . & . & . & . & 1 & * & . & . & 1 & * & . & . \\
. & . & . & . & . & . & . & . & . & . & * & 1 & . & . & * & 1 \\
. & . & . & . & . & . & . & . & * & 1 & . & . & * & 1 & . & . \\
. & . & . & . & . & . & . & . & . & . & 1 & 1 & . & . & 1 & 1
\end{array}\right]$$

Stars designate locations where 1's could appear, but don't, in Rule 22. Note how the structure of the tensor product makes the overall matrix look like a de Bruijn diagram, as well as each submatrix. There are better arrangements of the nodes, and consequently of the indices; Wuensche and Lesser's clustering techniques ought to be followed more closely. That is, the complementary automaton is lurking in this matrix, if one only knows how to perceive it.

More commentary will follow.

6 ancestors (6)

Commentary on Andrew Wuensche and Mike Lesser's new book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1) continues. We have digressed into some issues of matrix theory and graph theory in the expectation that a better understanding of the foundations of cellular automaton theory will help with understanding some of the topics of the book, such as Langton's parameter, Z, and perturbations (mutations).

To maintain a connection with the book, consider again the A and B matrices for Wolfram's (2,1) Rule 22:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

For a chain of three cells, we have eight products of these three matrices, corresponding to the ancestors of 000, 001, 010, and so on:

$$AAA = \begin{bmatrix} 1\,0\,0\,0 \\ 0\,1\,1\,1 \\ 0\,0\,1\,1 \\ 0\,1\,1\,2 \end{bmatrix}, \; AAB = \begin{bmatrix} 0\,1\,0\,0 \\ 1\,0\,0\,0 \\ 0\,0\,0\,0 \\ 1\,0\,1\,0 \end{bmatrix}, \; ABA = \begin{bmatrix} 0\,1\,0\,0 \\ 0\,0\,0\,0 \\ 0\,1\,0\,0 \\ 1\,0\,0\,0 \end{bmatrix}, \; ABB = \begin{bmatrix} 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,0 \\ 0\,1\,0\,0 \end{bmatrix},$$

$$BAA = \begin{bmatrix} 0\,0\,1\,1 \\ 0\,0\,0\,1 \\ 1\,0\,0\,0 \\ 0\,0\,0\,0 \end{bmatrix}, \; BAB = \begin{bmatrix} 0\,0\,0\,0 \\ 0\,0\,1\,0 \\ 0\,1\,0\,0 \\ 0\,0\,0\,0 \end{bmatrix}, \; BBA = \begin{bmatrix} 0\,1\,0\,0 \\ 1\,0\,0\,0 \\ 0\,0\,0\,1 \\ 0\,0\,0\,0 \end{bmatrix}, \; BBB = \begin{bmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,0 \end{bmatrix},$$

From these matrices we draw conclusions about the number of ancestors by looking at the (0,0) elements (quiescent-at-infinity configurations), traces (periodic configurations) or summing all the elements in the matrix (unrestricted configurations):

cells	quiescent	periodic	general	(= types of ancestors)
000	1	5	10	
001	0	0	4	
010	0	0	3	
011	0	0	3	
100	0	0	4	
101	0	0	2	
110	0	0	3	
111	1	3	3	
—	-	-	–	
sum	2	8	32	

Turning to page 96 of the Atlas, we confirm that s=8 (sum of the "periodic" column) and mp=5 (largest number in the "periodic" column). The only branching ratios are 3 and

5 (and of course, 0), confirmed by examining the diagram at level 3. The fact that g=6 (GOE configurations) follows from the fact that six of the matrices have zero diagonal, and hence trace zero.

None of the matrices is identically zero, so there are no poison words of length 3; however, the fact that the multiplicities are not uniform goes against the theorem (not yet discussed) about nonuniform multiplicities implying a Garden of Eden.

These details relate to a tiny fraction of the information contained in the Atlas, but they should suffice to establish the fact that as long as one is prepared to multiply A and B matrices, several different types of ancestors can be counted. Also, if one is prepared to work with symbolic matrices rather than numerical matrices, quite explicit ancestors can be calculated.

Of course, what is really wanted are general theorems about the matrices, so that all their products DON'T have to be calculated explicitly.

Considerable space would be consumed by listing all eight tensor squares, but the next point can probably be made with just one of them: small

$$AAA \otimes AAA \quad = \left[\begin{array}{cccc|cccc|cccc|cccc}
1 & . & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\
. & 1 & 1 & 1 & . & . & . & . & . & . & . & . & . & . & . & . \\
. & . & 1 & 1 & . & . & . & . & . & . & . & . & . & . & . & . \\
. & 1 & 1 & 2 & . & . & . & . & . & . & . & . & . & . & . & . \\
\hline
. & . & . & . & 1 & . & . & . & 1 & . & . & . & 1 & . & . & . \\
. & . & . & . & . & 1 & 1 & 1 & . & 1 & 1 & 1 & . & 1 & 1 & 1 \\
. & . & . & . & . & . & 1 & 1 & . & . & 1 & 1 & . & . & 1 & 1 \\
. & . & . & . & . & 1 & 1 & 2 & . & 1 & 1 & 2 & . & 1 & 1 & 2 \\
\hline
. & . & . & . & . & . & . & . & 1 & . & . & . & 1 & . & . & . \\
. & . & . & . & . & . & . & . & . & 1 & 1 & 1 & . & 1 & 1 & 1 \\
. & . & . & . & . & . & . & . & . & . & 1 & 1 & . & . & 1 & 1 \\
. & . & . & . & . & . & . & . & . & 1 & 1 & 2 & . & 1 & 1 & 2 \\
\hline
. & . & . & . & 1 & . & . & . & 1 & . & . & . & 2 & . & . & . \\
. & . & . & . & . & 1 & 1 & 1 & . & 1 & 1 & 1 & . & 2 & 2 & 2 \\
. & . & . & . & . & . & 1 & 1 & . & . & 1 & 1 & . & . & 2 & 2 \\
. & . & . & . & . & 1 & 1 & 2 & . & 1 & 1 & 2 & . & 2 & 2 & 4
\end{array}\right]$$

Note the following comparisons:

	(0,0) element	trace	overall sum
AAA	1	5	10
AAA $\otimes$ AAA	1	25	100.

In each case, the corresponding quantity is squared in the tensor square. Examining the structure of the matrix, it isn't hard to see why. However, this should give a graphic illustration of why the tensor powers participate in the calculation of moments, and how the tensor square (the connectivity matrix of the pair diagram) will eventually be involved in calculating the variance. Note that if we want to establish zero variance, it is only necessary to compare the square of the first moment with the second moment.

What is needed is $(A \otimes A + B \otimes B)^n$; when the power is expanded and all the terms collected, one finds one term for the square of the number of ancestors of each configuration. To get the third moment, take $(A \otimes A \otimes A + B \otimes B \otimes B)^n$, whereas if there were three states, there would be a C matrix, with a second moment expressed in terms of $(A \otimes A + B \otimes B + C \otimes C)^n$, and so on.

Because of the powers, we are interested in the rate of growth of the terms in parentheses in the last paragraph, which boils down to finding their largest eigenvalues or more generally, finding estimates or bounds for them.

More commentary will follow.

7 ancestors (7)

Commentary related to Andrew Wuensche and Mike Lesser's book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1) continues. As part of the general background for the commentary, we still need to describe the uniform multiplicity theorem.

To summarize what has been discussed so far:

1) There is a diagram, actually at least a century old, which gained much prominence in shift-register theory, called the de Bruijn diagram, which manages to subsume a great deal of the theory of linear cellular automata (and is a starting point for higher dimensions). That is, provided that it is labelled appropriately and interpreted satisfactorily.

This diagram just tells how to connect Wuensche and Lesser's 'start strings' together, and is practically a recipe for playing dominoes, which is another profitable way to interpret cellular automata — as a tiling problem.

2) For purposes of calculating ancestors, the de Bruijn diagram, or more appropriately, its connectivity matrix, has to be split into two parts. (For a binary automaton, that is, which is what is being discussed.) If symbols are put in the right places, and the symbolism of regular expressions is used, multiplying the matrices yields explicit ancestors. The practice would be more useful than it is if the complexity of the expressions did not grow exponentially. But that is the nature of reality, and decimal notation for numbers hides the fact that the size of products indeed grows exponentially. Moreover, in performing arithmetic, products and sums are consolidated into single numbers at each stage, whereas simplifying symbolic expressions as you go along never helps much.

For the purposes of counting, numerical de Bruijn fragments are quite satisfactory, but the problem remains of working with (noncommutative) matrices rather than numbers, affording a good chance for using ones numerical knowledge about matrices, and especially about non-negative matrices.

3) From the de Bruijn diagram, two more diagrams can be constructed, each of which illuminates the theory in its own way. The first is the subset diagram, which reveals what kind of paths the underlying diagram contains. It is the same as an exhaustive search, but it prescribes a systematic way to carry out the search. For automata, there are two different diagrams, due to the fact that the start strings can be extended either to the left or to the right. Not all rules of evolution are symmetric by reflection, so the difference is significant.

Applied to the calculation of ancestors, the subset diagram reveals at a glance whether there are ancestors or not. Due to working with subsets, rather comprehensive vision is required for the glance to work; the graph is rather large even for modest automata. Some things are fairly easy to read out of the diagram, but others require work; for configurations which actually have ancestors, it is far easier to multiply the aforementioned matrices than to decipher the subset path.

4) The pair diagram is much more modest than the full subset diagram; with ordered pairs its information is much more explicit. Its principal application lies in detecting and resolving ambiguity. There is a part of the pair diagram in which the two members of the pair are equal, and of course it reproduces the original diagram. However, whenever there is a mapping of the automaton to itself, there are pairs of the form (x,f(x)), whose subset has to show up in the pair diagram. A good example is Wolfram's (2,1) Rule 90, which is the same rule when all the cells are complemented. Both the diagonal and the antidiagonal of the de Bruijn pair diagram match the underlying diagram.

Rule 90 has quite a personality. Amongst other things, the even cells and the odd cells go their own merry ways, quite independently of one another, except for alternating generations.

The opposite of ambiguity is uniqueness, for which the pair diagram also serves. Suppose that the pair diagram has no loops except within its diagonal (pairs of the form (x,x)). Since the complement of the diagonal is finite, any path originating there must exit in fewer steps than there are pairs in the complement (otherwise it would enter one of those disallowed loops); the only place to go is onto the diagonal.

Suppose that a path leaves the diagonal. For the same reason as before, it must either terminate or reenter the diagonal in finitely many steps (again, the size of the complement). If all access to the diagonal is one-way, and if a finite configuration has an ancestor at all, it has to be unique except for a certain amount of leader or trailer. Following up these two quibbles will lead to the type of detailed analysis that we want to dispose of in the most general way, not arguing case by case.

The connection diagram of the pair matrix is also the second moment matrix for counting ancestors. It thereby relates statistics of the automaton, specifically variance in the number of ancestors, to numerical properties of the de Bruijn matrix. Again, general theorems are desired, rather than case-by-case analyses.

It seems to be hard, nay impossible, to get the desired proofs from within matrix theory, which is to say, by deducing limits on eigenvalues or norms (which are growth factors) from the size and arrangement of the matrix elements; this in spite of the fact that the results seem almost "obvious."

Rather, the de Bruijn matrices are matrices with positive elements and a norm, which could be, the sum of their elements. As such they form a ring, and rings have ideals, namely an algebraic structure. An ideal is simply a subset which persists under addition and multiplication; there are different kinds of ideals according to the handedness of the multiplication.

These matrices all have different norms, some are bigger, others are smaller. Consider those of minimum norm (which could well be zero). Such matrices are candidates for an ideal. The same for those of maximum norm (which, if it were infinity, would not be very helpful).

Suppose w is a word, N(w) the product of de Bruijn fragments counting its ancestors, that $u^T N(w) u$ is an extremal number (for all finite words), and that a is a single cell which we will add to the chain. N(wa)=N(w)N(a) is the new ancestor matrix. NOW, average over all the one-cell extensions; we have to divide by 2 (the number of states, 2 for binary automata) to get 1/2(N(w)N(0)+N(w)N(1)). N(0) is the earlier A matrix, N(1) the B matrix. Taking out a common factor we get 1/2N(w)D, because D is the sum of the de Bruijn fragments. We need 1/2(u^TN(w)Du), but Du=2u!. So we have u^TN(w)u back, which is still that extremal value. HOW can an AVERAGE be EXTREMAL? Only if everything being averaged is equal, and we see that the value is the same for all long chains.

Cleaning up details (the upper bound is actually finite, making it equal to the lower bound, therefore not zero and equal to the value for even single cells) we finally have the Uniform Multiplicity Theorem: Unless every configuration has the same number of ancestors as every other, there must be some configurations without any ancestors at all.

This soup is still not free of flies; how is it possible for there to be unique ancestors, and hence reversible automata, if all the configurations have to have four ancestors (the average is 4, so zero-variance means that all are 4) to avoid that one of them has none?

More commentary will follow.

8 ancestors (8)

Commentary related to Andrew Wuensche and Mike Lesser's book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1) continues. Having described the line of reasoning leading to the Uniform Multiplicity Theorem, we turn to an analysis of variance, or equivalently, of the second moment of the ancestor distribution. The reason for the interest lies in the relation between zero variance and zero Garden of Eden.

Every (2,1) automaton has a pair of 4x4 matrices which can be used to count ancestors and whose tensor squares count squares of numbers of ancestors; these are the A and B matrices of previous commentaries. Individual terms in the expansion of $(A+B)^n$ yield the number of ancestors associated with the monomials in the expansion; A+B itself is the de Bruijn matrix D, whose powers can be calculated explicitly. Each one is double its predecessor, and in the end there is an AVERAGE of four ancestors per configuration, whatever its length. How well individual terms of the sum conform to this average is an object of study.

For the second moment, powers of $A \otimes A + B \otimes B$ are required; this is not the same as $(A+B) \otimes (A+B)$; and therein lies a tale. What we need are eigenvalues, not forgetting the discrepancy between spectral norm and spectral radius for certain matrices. A widely used, and one of the best, estimates of the eigenvalues of a matrix is Gerschgorin's theorem, which has some alternative forms. One says that the eigenvalue is contained in a disk in the complex plane whose radius is the sum of the absolute values of the elements of some row. Not knowing which row leads to superposing the disks for each row and saying that the eigenvalue is lurking somewhere within. All of them.

Obvious variants use columns instead of rows, others center the disk on the (complex) diagonal elements, calculating the radius from the remainder of the row. It is also possible to average the rows, and it is possible to apply statistical concepts to the rows and eigenvectors themselves. Here it is useful to work with non-negative matrices, because all the numbers in the matrices can be used directly without absolute values. Furthermore, the eigenvector whose eigenvalue dominates the growth rate is non-negative, or can be normalized to be so.

Elementary statistics teaches that the average is an ideal origin for a set of data, within which the variance provides an ideal scale. Following this precept, the elements of a vector might be decomposed into an an average plus a residual. Write the column sums in a matrix as $C_i = \gamma + c_i$, the normalized eigenvector as $x_i + 1/n + u_i$ (for an nxn matrix), whose Perron eigenvalue (the largest one) is λ. Then there is a Statistical Gershgorin Theorem which asserts

$$\lambda = \gamma + nvar(c)var(x)\cos\theta$$

where θ is an angle involved in the derivation (correlation between c and x), but whose cosine is bounded between -1 and 1. For the matrices of our interest, γ is 1/n th the sum of their elements. For A and B, this is 1/4 the number of ancestors, and so a number ranging between 0 and 2. For $A \otimes A$, $B \otimes B$, and their sum, we have 1/16 the square of the number of ancestors. For $A \otimes A$ and $B \otimes B$, the value ranges between 0 and 4 (the square of 2), while for $A \otimes A + B \otimes B$ it ranges between 2 and 4.

Of the correction terms in this formula, n can be large, the variance in column sums can be modest; and the variance in x is small, the elements themselves never surpassing 1 and averaging 1/n. The formula itself is not something that anyone would think remarkable, and one mostly hopes that either the variances are small or that θ runs around 90 degrees. On the other hand, when the correction IS minor, it says that an 'average' number of ancestors is the eigenvalue which determines the growth rate.

Suppose that a is the sum of the elements in A, and b is the sum of the elements in B. We have a+b=8 always. To estimate the eigenvalue of $A \otimes A + B \otimes B$, we would then have $(a^2 + b^2)/16$, subject to the same constraint. The value is smallest when $a = b$, greatest when a=0 or b=0, symmetric between a and b.

The following table summarizes the results of a survey in which the maximum eigenvalue of each matrix was estimated by comparing the ratio of its third and fourth powers. The data was classified according to the value of a, with averages and variances for the estimated eigenvalue calculated individually for each value of a.

ab	number	min	max	ave	γ
08	1	4.000	4.000	4.000	4.000
17	8	2.938	3.381	3.102	3.125
26	28	2.400	2.800	2.580	2.500
35	56	2.132	2.695	2.253	2.125
44	70	2.000	2.480	2.155	2.000

In addition, the value for a=b (44) was split into two groups, according to whether the eigenvalue was 2 or not.

ab		number	mean	variance
44	$(\lambda = 2)$	30	2.000	0.000
44	$(\lambda > 2)$	40	2.272	0.102

The number in each of these categories is a binomial coefficient.

One may judge how well the estimate of γ matches the experimental data. It seems mostly better than 5%, and often better than 1%.

To turn the second moment into a variance, we need the relationship

$$\sigma^2 = ave(x^2) - (ave(x))^2.$$

Since the average is 2, and the second moment lies in the range 2-4, the standard deviation lies in the range 0-sqrt(2), with the assurance that some data are actually as far away from the mean as the standard deviation. Tchebycheff's Theorem is also pertinent, that less than $1/f^2$ of the data lies more than f standard deviations away from the mean.

The members of the group of 30 in the last table are candidates for reversible rules, and are the only (2,1) Rules for which there is no Garden of Eden. The other 40 Rules in the 44 class are balanced, in the sense that a=b and 0 has as many ancestral neighborhood as 1, namely 4 out of 8. That requirement is necessary but not sufficient.

The data which has been tabulated and discussed can also be presented as a histogram, but the limitations of a typescript prevent showing it on a printed page (although we could no doubt create an acceptable image if we really tried).

These results have been taken from two preprints which can be had upon request (with full mailing address), and were obtained by the use of the program LCAU21, for PC's, which is also available on request. Using it is non-trivial, however, due to its meagre documentation, especially in the ancestor option.

More commentary will follow.

9 ancestors (9)

Commentary related to Andrew Wuensche and Mike Lesser's book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1) continues. We have described a "statistical Gershgorin theorem" which implies a maximum eigenvalue for a matrix related to an average row (or column) sum with an error term which may or may not be easy to estimate.

Applied to (2,1) automata, it implies that the A and B matrices which have been introduced will have an eigenvalue between 1 and 2, namely n/4, where n is the least amongst number of ancestors of 0 (called a) or of 1 (called b), respectively. The result is not too surprising, but also refers to the growth in the number of ancestors of a string of pure 0's or pure 1's. We want the rate of growth of mixtures, without knowing too much about the mix except maybe its percentage composition.

Eigenvalue 1 means the number of ancestors remains constant as the length of the configuration grows; eigenvalue 2 means it doubles with each new cell. As we said, no surprise here.

Turning to the pair matrix $A \otimes A + B \otimes B$, which is also the second moment matrix, the same reasoning gives us eigenvalues in the range 2 to 4, namely $(a^2 + b^2)/16$. Whereas the average (first moment, computed from A+B), just doubles as the number of cells increments, the second moment AT LEAST doubles (when the ancestors are balanced) reaching a factor of 4, or quadrupling, in the cases of extreme unbalance in rules 0 or 255.

In addition, numerical experiments show that $(a^2 + b^2)/16$ is a good estimate of the rate of increase; for a given a the rate has an average close to the value in this formula with a small variance of its own (a variance in the variance, if you wish).

Turning the second moment into variance and thence into standard deviation, there are still some approximations to be made.

$$\sigma \quad = \quad \sqrt{g^2 \left(\frac{a^2 + b^2}{32}\right)^n - 4)}$$

the denominator 32 results from having to divide by 2^n, the number of configurations. The coefficient g is a correction due to the smaller eigenvalues of $A \otimes A + B \otimes B$, which interfere with the principal eigenvalue at first. Unless a=4, making us raise 1 to a power, subtracting 4 makes little difference.

So, except for a factor, σ is a number between 1 and 2, raised to the n/2 power (or, between 1 and 1.41 raised to the nth power).

Is this result credible? Is it useful? Rule 0 in a ten-cell automaton shows 1023 configurations with 0 ancestors, 1 with 4096 ancestors. This works out to a standard deviation of about 128. At the other extreme, Rule 150 has 4 ancestors per configuration, and zero standard deviation. For Rule 22, we have $(1.06)^{n/2}$, or about 3% increase for each additional cell. Six percent interest doubles your money in ten or twelve years, so we expect the standard deviation to double for each additional 20 to 25 cells in the configuration. The same would be expected for all the 112 (2,1) rules with unit imbalance, that is, ab=35.

ab	eigenvalue = growth of σ^2	distance needed to double σ
08	2.0	2
17	1.62	3
26	1.25	6
35	1.06	25
44	1.00	infinite

To make some sense of averages and sigmas, we need to have a feel for what an extremely skewed frequency distribution we are dealing with. No matter how long the configuration, the average number of ancestors is still 4; to get such an average, around half the data must be 4 or under, which means that if Wuensche and Lesser's basin diagrams (NAT's) were for unrestricted configurations (theirs are periodic), somewhere around half of the nodes would have 4 incoming links or less. Quite a few more would have slightly more than that.

The standard deviation is supposed to tell how far out from the average the data ranges, which will be mostly on the high side for ancestor data. To get a feel for typical values, consider an 8-cell configuration for (2,1) Rule 22 without boundary conditions:

ancestors	frequency	
0	2	.
1	57	***********
2	62	*************************
3	40	*************************
4	26	*********************
5	22	**********************
6	12	**************
7	6	********
8	8	*************
9	1	**
10	2	****
11	4	*********
12	4	*********
13	3	********
16	4	*************
	.....	
61	1	************
—-	—	
1024	256	

(1024 is gotten by multiplying number by frequency) The average is 4.0, sigma is 5.01; obviously that outlier is exerting an undue influence, but still 4-5=-1 to 4+5=9 does give a realistic approximation to where the data is. We are claiming that sigma will grow by about 3% for each new cell as the configuration is lengthened; at least from some point onward.

In fact, sigma behaves a bit more like $4.0exp(0.067n)$, (that is, with a multiplier of 1.069, or 7% growth) requiring about 5 iterations to double. So the use of a, b and $a^2 + b^2$ gives an approximation which is about so good, no more.

At least, there are some fairly general conclusions. One of the most important is that Langton's parameter is quite serviceable, although there are several things that could be called Langton's parameter, and their intended usage varies. All of them revolve around the idea of classifying the states by the fraction of neighborhoods which evolve into them, which is a portent of all kinds of things to come. One of them is the idea that as evolution progresses, an equilibrium must arise between the distribution of states and the distribution of ancestral neighborhoods.

Here we have seen that the statistics of ancestors depends on the neighborhood count by states in two ways. First, the dominant eigenvalue of the de Bruijn fragments is a direct function of this count - a multiple, in fact. Consequently the rate of growth for the number of ancestors of a string of like cells depends on the fraction of neighborhoods leading to that state. The biggest fraction always wins, in accordance with the principle that "Them as has, gets." Moreover, mixed strings predominate pretty much according to their mix of the dominant state.

Such general statements are always subject to refinement and correction, but the overall principle is pretty well justified.

The second aspect of the statistics of ancestors that has been discussed is their variance, which depends on the sum of squares of percentages - again, on Langton's parameter. Variance grows as the length of a string of cells grows, although never as fast as the number of configurations, the same which is true for the number of ancestors itself. Longer strings can have more (and less) ancestors than the average. The average is small, so much less is hard to come by; neverless the concept is interesting for reversible and "almost reversible" automata.

In applying this analysis to the Atlas, it should be borne in mind that, as a matter of statistics, there are only one quarter as many configurations satisfying periodic boundary conditions as there are without boundary conditions (and in turn one sixteenth as many which are "quiescent at infinity"). Thus the average number of ancestors for periodic configurations would be 1, not 4. Also, quantities may fluctuate more drastically for short configurations as the boundary conditions become more stringent.

In spite of this, growth factors apply equally for all kinds of boundary conditions. Moreover, once the powers of an irreducible matrix have come into equilibrium, fluctuations in the sizes of the matrix elements will also be immune to the boundary conditions, and they commence to dominate at the same stage as well.

More commentary will follow.

10 ancestors (10)

Commentary related to Andrew Wuensche and Mike Lesser's book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1) continues. While referring to this Atlas occasionally, a theory has been elaborated by which the statistical properties of the ancestor distribution for a one-dimensional cellular automaton may be deduced. Presentation of the theory having run its course, interpretation and comparison with the Atlas remains.

The theory which has been presented assigns a prominent role to a quantity which we have called γ, which is the average of either the row sums or the column sums of a matrix. For the connectivity matrices of de Bruijn matrices, this translates directly into numbers of ancestors, their squares, and their averages. This is reminiscent of earlier work of Christopher Langton, who used such averages for classifying automata; strictly his lambda compared the quiescent state to the rest, but there are only two states for binary automata. Wuensche and Lesser record lambda for all the automata in the Atlas, along with Z, a parameter of their own.

What is new in these commentaries is the relationship between variance and the parameters, namely the rate of growth in the standard deviation which depends upon $(a^2+b^2)/16$, where a is the number of ancestors of 0, b the number of ancestors of 1, and 16 is the dimension of the pair connection matrix.

Previously the rate of growth has been tabulated for (2,1) automata; here it is shown for (2,3/2) automata.

ab	number	γ	observed average	second moment	observed variance
0-16	1	4.000	4.000	16.00	0.000
1-15	16	3.531	3.515	12.36	0.088
2-14	120	3.125	3.116	9.73	0.132
3-13	560	2.781	2.797	7.84	0.133
4-12	1820	2.500	2.548	6.51	0.120
5-11	4368	2.281	2.361	5.58	0.104
6-10	8008	2.125	2.231	4.98	0.091
7- 9	11440	2.031	2.154	4.65	0.091
8- 8	12870	2.000	2.129	4.54	0.093
>2.0	12256	2.000	2.135	4.57	0.091
=2.0	614	2.000	2.000	4.000	0.000

In 1972 Amoroso and Patt[9] found some non-trivial reversible automata amongst the 614 with zero variance. By non-trivial, one discounts rules which work by shifting, complementing, or copying, which are the only reversible (2,1) Rules.

Another tabulation which we have made concerns (3,1/2) automata; here we have a, b, and c with lambda determined by $(a^2 + b^2 + c^2)/9$:

abc	number	min	max	σ	ave	γ
009	1	9.000	9.000	0.000	9.000	9.000
018	9	7.047	7.519	0.222	7.204	7.222
027	36	5.747	6.110	0.164	5.949	5.889
036	84	5.000	5.531	0.163	5.107	5.000
045	126	4.562	5.266	0.133	4.676	4.556
117	72	5.095	6.002	0.334	5.704	5.667
126	252	4.362	5.283	0.254	4.719	4.556
135	504	3.813	4.615	0.227	4.121	3.889
144	630	3.707	4.854	0.213	3.920	3.667
225	756	3.707	4.854	0.242	3.970	3.667
234	1260	3.259	5.000	0.248	3.602	3.222
333	1680	3.000	4.002	0.298	3.430	3.000
>0.0	1260			0.197	3.547	3.000
=0.0	420			0.000	3.000	3.000

In all of these (three) cases which we have presented, some common features can be observed. Each value of γ leads to a rather well defined cluster of growth rates, even though the value of γ itself corresponds to observation better for high values than for low values; nevertheless the discrepancy is gradual and monotonic.

At one time we thought that there was a gap between zero variance and the next lower value, but experience with additional (k,r) combinations has reduced our confidence in its

[9] S. Amoroso and Y. N. Patt, decision Procedures for Surjectivity and Injectivity of Parallel Maps for Tesselation Structures, Journal of Computer and System Sciences 6 448-464 (1972).

existence; it is most likely an artifact of small sample size (small k, small r). The general shape of the histogram ought to be roughly Gaussian, because of the binomial coefficients associated with given values of γ, which itself grows quadratically with a. Actually, because of $a^2 + b^2$, the purported gaussian is folded over in the middle, ab giving the same data as ba.

Since a Gaussian has a point of inflection at its own standard deviation, we would expect a noticeable division of γ into low values all of whose Rules have a slow growth rate in their ancestral variance, and those for large growth rates, up in the tail of the Gaussian. It shouldn't be hard to figure out this distribution function, but we haven't done it. What we do notice, is that the growth factor is pretty much the same for quite a few nearly balanced Rules, and that they are set off slightly from the exactly balanced Rules.

With respect to interpreting the Atlas, some additional work is called for. The analysis of variance which we have described ultimately translates into an average number of ancestors per configuration, and the growth of this average with respect to the length of the configuration. The Atlas only tabulates the maximum number of counterimages by basin; but it is the average number which follows more readily from our analysis. The average can be deduced by examining the images, but getting a good sample is going to be laborious.

On the other hand, it is instructive to make comparisons of the maximum number of counterimages, particularly as it depends on the lambda parameter.

More commentary will follow.

11 ancestors (11)

Commentary related to Andrew Wuensche and Mike Lesser's book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1) continues. Some observations on Langton's parameter, lambda, are in order.

We have described a 'Stastical Gerschgorin Theorem' (which is more of a formula than a theorem) which assigns a prominent role to the fraction of neighborhoods begetting each state in the enumeration of ancestors. These fractions enter into the calculation of moments with a correction term which experience shows to be small; if not always zero, its size is predictable and consistent.

If one calls such fractions 'Langton's parameter,' one has a solid basis for classifying automata according to such a parameter, whatever it is called. In other contexts, the fraction plays a role in calculating self-consistent probabilities, although there it yields a zero-order'th approximation to the fixed point.

As a predictor of automaton behavior, lambda has gained a mixed acceptance; Wuensche and Lesser introduce Z with the claim that it is a more sensitive indicator. The reason for this, among other things, is the bad company which Rules 18 and 126 are seen to be keeping in the example which follows. However, in browsing through an Atlas such as theirs, there is a tendency to see what one expects to see, particularly given the mass of data and their similarity to one another. So it behooves us to sharpen our tastes somewhat.

The situation may be likened to that prevailing in Botany before the advent of Carolus Linnaeus; there were tall trees and bushy trees and trees that kept their leaves through all seasons and those that shed their bark instead of their leaves, and those that smelled good and those that raccoons climbed in and others which monkeys preferred, and so on. Classifying them and every thing else according to the layout of their reproductive organs seemed rather prosaic, but in the end it brought order to a lot of chaos. And the monkeys even got to keep their tree (Araucaria araucana).

Calculating the average number of ancestors is like calculating the bushiness of our tree, in which case calculating their standard deviation amounts to observing whether this bushiness is strictly observed or whether it can vary considerably. Once again, examining a specific example may be helpful. Suppose lambda is 25% (lambda ratio = 0.5 in the Atlas). There are 56 (2,1) Rules with this ratio (including those with lambda = 75%), which the Atlas assigns to 11 clusters. The growth factors for the quiescent configuration and for the standard deviation in the number of ancestors are shown in the table below.

cluster typical Rule no.	ancestors of dominant cell	growth of standard deviation
3	1.618	1.653
5	1.615	1.648
6	1.470	1.587
9	1.339	1.579
10	1.167	1.581
12	1.100	1.588
18	1.614	1.634
24	1.100	1.588
33	1.466	1.578
36	1.618	1.653
126	1.618	1.653
nominal	1.500	1.582

A growth	Rule cluster
1.1	12 24
1.2	10
1.3	
1.4	9
1.5	6 33
1.6	3 5 18 36 126

The histogram on the side gives an idea of how the growth factor of the A matrix, the column in the table titled 'ancestors ...,' are distributed around their mean of nominal value of 1.500. The other column has a parallel distribution.

The range 1.1 to 1.6 should be compared with the nominal values of 1.25 for ab=35 Rules and 1.75 for ab=17 Rules; if there is any transgression, it is for the Rules with small growth rates.

What we have to decide by looking at the Atlas is, whether this is all true, and whether, by knowing it, there are some features of the basins there displayed which should attract our attention, maybe even stand out.

(Has anyone noticed that in the Atlas, although the custom is to place the panel showing evolution from a single cell on the left and that from a random initial configuration on the right, this has been reversed for Rule 4 on page 88? Such are the delights of trying to publish an exceedingly detailed book and getting everything right)

Up until now, we have been unwilling to identify "maximal growth rate" with "eigenvalue of A," because we have not proven rigorously that this is, in fact, the maximal growth rate. On the other hand, it is evident by inspection that the quiescent configuration will have the most ancestors, both according to the theory which has been presented in this series, and the empirical data comprising the Atlas; it only lacks a proof.

Another source of discrepancy is the fact that we have elaborated a general theory without boundary condition, whereas the Atlas is concerned exclusively with periodic configurations. Nevertheless, the nature of the general theory is such that all conclusions regarding multipliers, such as growth rates, apply equally to every variant of the boundary conditions which can be obtained by varying the metric matrix. This specifically includes periodic boundary conditions.

The one restraint which must be observed, is that it takes time for growth factors to reach the maximum eigenvalue of a matrix, so that conclusions should not be drawn for periodic configurations of a very short length. What constitutes 'very short' varies from Rule to Rule, but is closely related to the variation in size of the matrix elements within A, and especially to its pattern of zeroes.

One might wonder whether Garden of Eden configurations are included within this sweeping generalization, and the answer is yes, subject to the same precautions. The reason is that in a general automaton, poison words, and hence Gardens of Eden, arise because of incompatibilities - the requisite series of ancestral neighborhoods simply can't be found. In addition, likely ancestors may fail to meet boundary conditions.

For small rings, more ancestors will be lost because of boundary conditions. Recall that for Rule 22, eight is the shortest ring which has a poison word. But as rings grow longer, it is easier and easier to accommodate boundary conditions - two fragments which won't work separately may join together to compensate each other's deficiencies. Consequently for long rings, the Garden of Eden may be reduced by 1/4th, but otherwise its growth will follow that of the general theory.

Returning to "maximal growth rate" and being willing to equate it with the larger of the dominant eigenvalues of the A, B pair (which implies that the quiescent configuration (or pair of alternating uniform configurations if none is quiescent, or uniform ancestor of the quiescent configuration when the dominant eigenvalue does not belong to the latter), we could compare growth and eigenvalue for the 26 (lambda ratio = 0.5) clusters:

cluster: typical Rule no.	ancestors of dominant cell	ratios of largest basins according to Atlas	page reference
3	1.618	1.618	128
5	1.615	1.644, 1.592, 1.621	90
6	1.470	1.462	130
9	1.339	1.326, 1.300	134
10	1.167	many small basins	161
12	1.100	many small basins	136
18	1.614	1.544, 1.616	92
24	1.100	shifting rule	169
33	1.466	1.469	100
36	1.618	1.617	103
126	1.618	1.622	123

Except for three clusters with many small basins, the agreement is exemplary[10].

[10] It was later concluded that the three basins, with a typical growth rate of 1.100, were also in good agreement

More commentary will follow.

12 ancestors (12)

Commentary related to Andrew Wuensche and Mike Lesser's book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1) continues. The time has come to investigate Z.

We are basing our commentary on the A and B matrices into which the de Bruijn connectivity matrix for any binary rule splits. Consider Rule 193, featured in the text on page 40 of the Atlas; its AB pair is:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

For A, the vector of column sums is (1 2 1 1), of row sums it is (1 2 2 0). The column average is $5/4 = 1.25$, the second moment is $7/4 = 1.75$, giving a $\sigma(c) = 0.433$. For rows the average is also $5/4 = 1.25$, with second moment $9/4 = 2.25$ and $\sigma(r) = 0.81$ approximately. Using columns seems preferable.

Earlier we stated a 'Statistical Gershgorin Theorem' (erroneously writing var for the greek sigma that was in the original source):

$$\lambda \;=\; \gamma + n\sigma(c)\sigma(x)\cos(\theta)$$

where lambda is an eigenvalue (NOT Langton's parameter) of a matrix such as A, γ is $1/n^{th}$ the sum of the elements of the matrix (and IS double Langton's parameter when the majority state is quiescent). N is the number of rows in the matrix, here 4, the sigmas refer respectively to column sums and the eigenvector, and θ is the angle between their vectors of residuals. Interestingly, this angle must be EXACTLY 90 degrees for reversible automata[11].

Up until now, we have drawn some conclusions based on using γ while discarding the companion term; but all discrepancies noted are exclusively due to the correction. The discrepancies seem to follow a regular pattern, small but typically non-zero.

It will be noted that the A and B matrices have three kinds of rows (and columns as well). They can be zero, unit vectors, or contain a pair of ones. The general format is forced by the structure of the de Bruijn diagram, being just the number of out links (in links) per start string (stop string). Given more states than a binary automaton possesses, there will be more links, and so a greater variety of rows or columns; but the unit vectors signal situations in which just one single continuation is possible.

The unit vectors are significant, being what the Atlas calls deterministic; the fraction of them taking both A and B into consideration is Z. One is allowed to choose between rows or columns, so as to get the larger of the two numbers.

For Rule 193, there are altogether 2 unit rows and 6 unit columns. Therefore Z = 6/8, or 0.75, whereas the (Langton's) lambda is 5/8, or 0.625, yielding the lambda ratio 0.75 duly recorded in the Atlas, page 157.

[11] Ambiguous case when one of the standard deviations is zero.

Estimating the largest eigenvalue of A produces 1.324, which can be compared to the quotient of maximum preimaging for 15 and 14 member rings (p. 157), which is 68/51 = 1.333. Again, the agreement is exemplary.

Comparing fairly modest powers of A reveals the eigenvectors of this eigenvalue; normalized they are: column, (0.246, 0.323, 0.430, 0.000), row, (0.184, 0.323, 0.246, 0.246).

Residuals for the column sum are (-0.25, 0.75, -0.25, -0.25), residuals for the column eigenvector are (-0.004, 0.073, 0.180, -0.250). The inner product of these two vectors give directly the correction to γ, and is 0.072. Almost exactly what is expected, 1.250 + 0.072 = 1.322, the tiny discrepancy is surely due to the care (or lack thereof) taken in arriving at these numbers.

The L^2 norm of zero-average vectors is their standard deviation without averaging, which accounts for the factor n in the statistical theorem, so we quickly arrive at a value of $\cos\theta$ of 0.271, or a θ of about 74 degrees. We already knew $\sigma(c) = 0.433$; we readily calculate $\sigma(x) = 0.157$, so we have identified all the quantities in the formula.

$$1.322 \quad = \quad 1.250 + 4 * 0.433 * 0.157 * 0.271$$

There is an interesting, although possibly trivial, way to place Z in this formula. Create the supermatrix from A and B that was once mentioned:

$$\left[\begin{array}{cc} A & . \\ . & B \end{array}\right] = \left[\begin{array}{cccccccc} 0 & 1 & 0 & 0 & . & . & . & . \\ 0 & 0 & 1 & 1 & . & . & . & . \\ 1 & 1 & 0 & 0 & . & . & . & . \\ 0 & 0 & 0 & 0 & . & . & . & . \\ . & . & . & . & 1 & 0 & 0 & 0 \\ . & . & . & . & 0 & 0 & 0 & 0 \\ . & . & . & . & 0 & 0 & 0 & 0 \\ . & . & . & . & 0 & 0 & 1 & 1 \end{array}\right]$$

It is not such an artificial creation as might be feared: from graph theory it is the connectivity matrix of the least upper bound (union) of two graphs; one is the de Bruijn diagram with 0-ancestor links, the other is the de Bruijn diagram with 1-ancestor links. Nodes in a union are linked when there are links in either one (or both) of its two constituents.

The vector of column sums is (1 2 1 1 1 0 1 1). The average of column sums is 1, composed from 0's, 1's, and 2's; every 0 is paired with a 2. Thus the vector of residuals, (0 1 0 0 0 -1 0 0) will contain -1's, 0's, and 1's; the sum of their squares will be eight minus the number of columns which were unit vectors, so the average of the sum is 1-Z. Sigma squared is the second moment taken with respect to residuals, so $\sigma(c) = \sqrt{1-Z}$. When Z=1, as for reversible Rules, the standard deviation is zero; when Z=0, as for the Zero Rule, the standard deviation is 1.

The supermatrix now encompasses both maximum and minimum rates of growth; being partially diagonal its eigenvectors are those of its blocks, extended to the other block with zeroes, and the statistical theorem still applies to all eigenvalues. Because of the increased dimension, each eigenvector will be the same as before, but its mean will be reduced by half (to 1/8) because of all the new zeroes, while its second moment s^2 will be shifted by

1/64 and scaled by 1/2. Finally, the supermatrix has eight rows and always has eight ones distributed throughout. So the new formula is

$$\lambda \quad = \quad 1 + 8\sqrt{1-Z}\sqrt{0.5 * s^2 + .01561}\cos(\theta^*)$$

where we still have to contend with a new angle θ^* whose cosine we calculate to be 0.482 (about 61 degrees). The equality is nearly as good,

$$1.321 = 1.000 + 8.000 * 0.500 * 0.167 * 0.482$$

Just as the extended eigenvector is gotten by filling with zeroes, the extended vector of sums is gotten by adjoining 2u-x (making for the average of 1 and variance 1-Z), where u is a vector of 1's and x is the vector to be extended. The extended vectors of the Union Matrix therefore have the same inner product as the original vectors. Nevertheless, the residual vectors, from which θ^* is calculated, are slightly different; anyone who wanted an explicit relation between θ and θ^* could work out the algebra.

Critics may question the utility of the formulas that have been displayed; it remains to be seen to what extent the standard deviations and angles are consistent within families of automata, or whether they can be transferred from one automaton to another.

Nevertheless we have constructed a framework into which Langton's parameter and Z both fit, as algebraic quantities relating to the growth factor for maximal counterimaging.

More commentary will follow.

13 ancestors (13)

Commentary related to Andrew Wuensche and Mike Lesser's book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1) continues. At least two different parameters, lambda and Z, are useful in classifying cellular automata; but the general philosophy of parameters should be contemplated before deciding to rely on one or another of them, or on something else.

It is a natural idea to assign a single number to an automaton, or to an automaton rule, with the expectation that it will distinguish between automata, and perhaps also reveal some significant information about them. An extremely natural and obvious candidate is the fraction of transitions (or neighborhoods, if you wish) leading to one of the states, particularly in a binary automaton.

From the side of probability theory, one expects an automaton to evolve into an equilibrium in which the fraction of neighborhoods producing the state is the same as the relative proportion of that state. In its simplest application, this is mean field theory, but other approaches utilize varying degrees of sophistication for estimating the probabilities, making for correspondingly better estimates. The fraction of neighborhoods serves as an excellent starting point for these theories, and is often not too far from the equilibrium calculated by other theories.

From the side of ancestor theory, we have seen that there are some matrices arising from graph theory, whose spectral radii and spectral norms, between them, account for the proliferation of ancestors, which is the converse of the convergence seen during evolution. We have also seen that the ancestor fraction serves to give a reasonable estimate of the growth rate, not of ancestors from generation to generation, but of the number of ancestors for a single generation as a function of the length of a configuration.

Indirectly these estimates affect the characteristics of the evolution for many generations, which are thereby related to the underlying parameter. Thus a parameter may serve for more than classification, it may participate in some simple algabraic expression describing observable data produced by the automaton. When a parameter enjoys a general display of success, the public seems to become more demanding in its expectations for the parameter, requiring either a new and better parameter, or some supplementary parameters which explain discrepancies in the predictions of the original paramenter.

With respect to the calculation of ancestors, there is a good and solidly based theory from which to start. The de Bruijn diagram, and such of its subsidiaries as the subset diagram, pair diagram, union diagram, and so on, allow an exact calculation of ancestors, and statistics of ancestors, such as their average number, standard deviation, and, through their moments, their exact frequency distribution. In practice, the matrices are large, the calculations tedious, and above all, symbolic results are desired which apply to whole classes of automata, rather than just individual cases.

The procedure would be easier were it not for the discrepancy between spectral norm and spectral radius, but in general terms, estimating the largest eigenvalue of the larger of the de Bruijn fragments is sufficient to estimate the largest number of ancestors that any single configuration can possibly have, which is a quantity of interest that is tabulated in the Atlas.

The average number of ancestors is always necessarily 4, for (2,1) automata, but the actual number is influenced, especially for configurations of short length, by boundary conditions, and by the variance, which can be estimated by the same procedures using another matrix (the pair matrix). The actual distribution is highly skewed, because half of the configurations, in some sense, have 4 or less ancestors. This quantity includes zero ancestors - Garden of Eden configurations - and is just as inevitable as the average of 4.

Not only is the top half of the distribution highly skewed - spread over the range $(4, 2^n)$ - but it typically contains a few outliers with the bulk of the data closer to the mean; the standard deviation has to take this into account.

The outliers are generally the ancestors of the quiescent state, if there IS one. Modifications in this observation have to be made when there are TWO quiescent states, or none, or the state with the majority of ancestors is not the quiescent state (for non-binary automata the permutations increase quite rapidly with the number of states).

In previous commentaries, we have shown that both Langton's lambda and Wuensche and Lesser's Z can be incorporated into formulas expressing the rate of growth of "maximal preimaging" (beware - λ is γ and —em vice versa).

$$\begin{aligned} \text{(Langton) } \lambda &= \gamma + n\sigma(c)\sigma(x)\cos(\theta) \\ \text{(W\&L) } \lambda &= 1 + 2n\sqrt{1-Z}\sqrt{0.5 * s^2 + 0.01561}\cos(\theta^*) \end{aligned}$$

There is nothing especial to reccomeend these formulas except that they are extremely simple; in the Langton version, the second term is simply ignored. Nevertheless the results agree to about 10%, while the accurately computed eigenvalues agree quite well with data taken from the Atlas.

To satisfy curiosity regarding the composition of the discarded term, it is tabulated below for the ab values 08, 17, 26, and 35, according to Wuensche and Lesser's clusters, all of whose members obey the same statistics.

ab	rule	λ	=	γ	+	n	*	$\sigma(c)$	*	$\sigma(x)$	*	$\cos(\theta)$	θ	page
08	0	2.000	=	2.000	+	4	*	0.000	*	0.000	*	1.000	0	85
17	2	1.620	=	1.750	+	4	*	0.433	*	0.083	*	-0.897	154	126
	1	1.839	=	1.750	+	4	*	0.433	*	0.055	*	0.936	21	86
	4	1.754	=	1.750	+	4	*	0.433	*	0.049	*	0.056	87	88
26	3	1.617	=	1.500	+	4	*	0.5	*	0.156	*	0.378	68	128
	5	1.615	=	1.500	+	4	*	0.5	*	0.084	*	0.686	47	90
	6	1.470	=	1.500	+	4	*	0.5	*	0.165	*	-0.092	95	130
	9	1.338	=	1.500	+	4	*	0.5	*	0.178	*	-0.456	117	134
	10	1.166	=	1.500	+	4	*	0.5	*	0.271	*	-0.613	127	161
	12	1.100	=	1.500	+	4	*	0.5	*	0.330	*	-0.619	128	136
	18	1.613	=	1.500	+	4	*	0.5	*	0.153	*	0.375	68	92
	24	1.100	=	1.500	+	4	*	0.5	*	0.204	*	-1.000	180	169
	33	1.465	=	1.500	+	4	*	0.5	*	0.072	*	-0.241	103	100
	36	1.617	=	1.500	+	4	*	0.5	*	0.059	*	1.000	0	103
	126	1.617	=	1.500	+	4	*	0.5	*	0.059	*	1.000	0	123
35	193	1.324	=	1.250	+	4	*	0.433	*	0.158	*	0.279	74	157
	22	1.459	=	1.250	+	4	*	0.433	*	0.163	*	0.744	42	96
	37	1.380	=	1.250	+	4	*	0.433	*	0.086	*	0.845	32	104
	50	1.000	=	1.250	+	4	*	0.829	*	0.165	*	-0.454	117	106
	73	1.310	=	1.250	+	4	*	0.433	*	0.144	*	0.262	75	112
	94	1.200	=	1.250	+	4	*	0.433	*	0.065	*	-0.365	111	118
	7	1.465	=	1.250	+	4	*	0.433	*	0.169	*	0.738	42	132
	13	1.000	=	1.250	+	4	*	0.433	*	0.433	*	-0.333	109	138
	26	1.083	=	1.250	+	4	*	0.433	*	0.299	*	-0.333	109	140
	35	1.000	=	1.250	+	4	*	0.829	*	0.250	*	-0.301	108	146
	38	1.105	=	1.250	+	4	*	0.829	*	0.203	*	-0.229	103	148
	41	1.083	=	1.250	+	4	*	0.433	*	0.299	*	-0.333	109	150
	62	1.324	=	1.250	+	4	*	0.433	*	0.158	*	0.279	74	156
	11	1.062	=	1.250	+	4	*	0.433	*	0.333	*	-0.228	103	162
	14	1.083	=	1.250	+	4	*	0.433	*	0.345	*	-0.289	107	164
	25	1.083	=	1.250	+	4	*	0.829	*	0.299	*	-0.174	100	170
	28	1.000	=	1.250	+	4	*	0.433	*	0.259	*	-0.555	124	172
	19	1.618	=	1.250	+	4	*	0.829	*	0.156	*	0.710	45	94
44	29	1.000	=	1.000	+	4	*	0.707	*	0.144	*	0.000	90	174
	ab	1.617	=	1.250	+	4	*	1.089	*	0.160	*	0.526	58	174
	ba	1.618	=	1.250	+	4	*	0.829	*	0.156	*	0.710	45	174

In earlier commentaries, we have remarked on the excellent agreement between the eigenvalues and data in the Atlas. Even the three 26 cases which were judged to be marginal conform quite well to the predictions; we were much too hasty in our earlier conclusion, due to the large number of small basins and the closeness of the growth factors to 1.0.

Nevertheless, the agreement begins to suffer with the 35 and 44 automata, for reasons which answer an earlier question. 44 is balanced, yet does not always produce zero variance (a known failing). The answer is that the de Bruijn fragments can be strongly attractive and dissipative, hence the matrix might have the Jordan form, and the eigenvalue could differ markedly from the spectral norm.

It will be seen that the products AB and BA behave better than A or B alone, giving maximal preimaging consistent with the configurations 01010101.... shown in the Atlas, whilst the chains 0000... and 1111... may have rather few counterimages.

The single 44 example in the table illustrates what has happened. The spectral norm of the fragments is sqrt(2) = 1.414, but their eigenvalue is 1. While the eigenvalue underestimates general growth rates, the spectral norm overestimates them, given that the true result is $\sqrt{1.617}$.

More commentary will follow.

14 Reversible Automata

Discussion originating elsewhere concerning the construction of reversible automata and the existence of Gardens of Eden has pointed up the fact that by not taking up the AB = 44 case in any detail in our recently concluded series of commentaries "ancestors-(n)," we have missed a good opportunity to discuss the reversibility question. In order to make use of all the background developed in the earlier series, one could insert the present posting between ancestors-(13) and ancestors-(14).

The Uniform Multiplicity Theorem assures us that the number of ancestors of all configurations must be the same if an automaton is to be reversible. This may represent a surprise for someone who has relied on calculating the subset diagram to find out whether an automaton has a Garden of Eden, but once the theorem is known, far fewer automata have to be contemplated. For (2,1) automata, there are $8!/(4!\ 4!) = 70$ balanced Rules, for which AB = 44 in the nomenclature of the ancestor series; this number is much less than the 256 total, and is still further reduced into 16 of Wuensche and Lesser's clusters.

The rate of growth of "maximal preimaging" is related to, but not necessarily equal to, the eigenvalues of the A and B matrices. When A and B are different, one of them usually dominates, the other can be disregarded, and the largest eigenvalue of the dominant member is nearly always the number sought. When the number of nonzero elements in A and B is equal or nearly equal, both are candidates, although one of them may still dominate. If that happens, there WILL be a Garden of Eden, and the situation is similar to all the others, even though growth rates will be smaller than otherwise.

The expected eigenvalue, for both A and B, is 1; even when that is true, the eigenvalue of AB or BA may be greater than 1, so there is still a Garden of Eden, but maximal preimaging will occur for strings of 010101... rather than for strings of 000000.... or 111111... And even when AB and BA have unit eigenvalue there is still AAB, ABA, ... and so on. We were unsure that such would happen before completing this analysis, but examples will be seen in the table which follows.

rule	size	maximum eigenvalue of A	B	AB (BA)	pair	Atlas page	comment
23	(2)	1.466	1.466		2.490	98	Garden of Eden
51	(2)	1.000	1.000		2.000	108	identity
54	(4)	1.000	1.321		2.229	110	Garden of Eden
77	(2)	1.000	1.000	2.147	2.490	114	Garden of Eden
90	(2)	1.000	1.000		2.000	117	uniform
105	(2)	1.000	1.000		2.000	120	uniform
27	(8)	1.000	1.000	1.000	2.229	142	severe Jordan
30	(8)	1.000	1.000		2.000	144	uniform
45	(8)	1.000	1.000		2.000	152	uniform
58	(8)	1.000	1.000	1.618	2.265	154	Garden of Eden
15	(4)	1.000	1.000		2.000	166	shift
29	(4)	1.000	1.000	1.618	2.224	174	Garden of Eden
43	(4)	1.000	1.000	......	2.224	176	cubes nonuniform
46	(4)	1.000	1.000	BAB:2.000	2.510	179	Garden of Eden
57	(4)	1.000	1.000	1.618	2.205	180	Garden of Eden
60	(4)	1.000	1.000		2.000	183	uniform

To interpret these results, the actual de Bruijn fragments - the A and B matrices - have to be examined. There are 16 row-stochastic pairs among them, and 16 column-stochastic pairs. That they will have maximum eigenvalue 1 and u as an eigenvector is a foregone conclusion. At the same time there are 4 doubly stochastic pairs, which are counted doubly amongst the singly stochastic matrices. Altogether, then, there are 28 stochastic matrices, almost the full complement of 30 variance-zero Rules. The remaining two come from cluster 51 consisting of Rules 51 (complementation) and 204 (identity).

The great majority of the A and B matrices in this table have eigenvalue 1, but also, they all show the Jordan form to one degree or another. There are several instances where the eigenvalue of AB or BA exceeds 1, but a similar number require longer products before matrices with strict growth factors occur.

We need a definitive assessment of the relationship between spectral radius and spectral norm. On one hand, we know that the spectral norm is the square root of the spectral radius of the product of a matrix by its transpose, and that the spectral radius is less than or equal to the spectral norm. That is because the maximum amplification of any vector bears that relation to any specific amplification, of which the eigenvalue is one example. We also know that spectral norms are convenient because there are inequalities governing sums, scalar factors, and products which can be used in computing rigorous bounds, for all of which the spectral radius may be inadequate.

Nevertheless, there exists a representation of matrices in terms of their eigenvectors, which is Sylverster's formula. In the most general case, there are idempotents Gi constructed from eigenvectors, and ladder operators Ni, constructed form principal vectors, such that for a matrix A with eigenvalues λ_i, and any function (representable by a power series) f,

$$f(A) \quad = \quad \sum_i (f(\lambda_i)G_i + f'(\lambda_i)N_i + f''(\lambda_i)N_i^2 + ...).$$

The number of derivatives required is no larger than the dimension of the matrix; for estimating ancestors, the function f would be a large power. That leads to domination of the sum by the largest eigenvalue, but it will not eliminate the contributions of the derivatives, which may be of comparable magnitude to the leading term. If the largest eigenvalue is less than 1, we may correctly assume that the assemblage will tend to zero, but the fact of the matter is that for the ancestor application, the eigenvalue is 1 or larger, with especial interest in the case where it is exactly 1.

To see how this works, consider the following matrix and its powers:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} = 1^n \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + n * 1^{n-1} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

The situation is completely typical of matrices in the Jordan form, and can be observed at work in the A matrix for Rule 27 in the table above, as well as in the A matrix of Rule 46.

As for the de Bruijn diagram the matrix example represents, there are two nodes, with self-loops. In addition, one of the nodes feeds into the other. The eigenvector selects the first node, which always renews itself. Any attempt to involve the second node will overload the first, which would then have to feed the second without any compensating counterflow.

From the point of view of matrix theory then, to get uniform counterimages and therefore satisfy a necessary condition for a reversible rule, we have to look for matrices with eigenvalue 1, yet beware that they also manifest Jordan form. We also know that it is sufficient to look for them amongst rules where a = b, that is, where the Rule itself has balanced counterimages.

From the point of view of graph theory, it is possible to say a little more. Given that the elements of the connectivity matrix count paths, and we want to avoid that the number of non-zero elements in powers of the connectivity matrix increase, it is necessary to avoid branching paths. More precisely, paths which branch can never return to the node at which they branched in more than one way. In short, there cannot be two intersecting loops, otherwise long paths could proliferate by mixing them.

However, if there is not at least one loop, long paths would be impossible, and the power would reduce to zero. The criterion is then, that there must be loops, but that if there are more than one, they cannot be linked. For instance, the A and B matrices of Rule 23 have 1-cycles as nodes in 3-cycles. Rules 46 and 58 have A matrices with 1-cycles which are linked in 2 steps into the other 1-cycle. All of these rules disqualify for reversibility.

Among those which do qualify, there is Rule 30: the A matrix has 2 1-cycles; the B matrix has a 3-cycle. For Rule 45, A has a single 1-cycle; B has a 1-cycle and a 2-cycle. The reason that these Rules are not reversible is seen in the pair matrix: the number of counterimages is uniform, but they can be completely different.

The additional requirement is that the entire discrepancy in multiple counterimages occur "at infinity" which is to say, in a limited part of the outer boundaries of the string. This requires that there be one single image of the de Bruijn diagram - the diagonal - in the pair diagram. For the Rules mentioned, there are viable loops outside the diagonal.

To confirm that the criterion is a good one, the reversible Amoroso-Patt (2,3/2) Rules may be examined: one of them is 0000 1111 0100 1011; its A and B de Bruijn fragments have a single 1-cycle.

Another example is Kari's (3,1/2) rule 001 110 222, for which the A, B, and C fragments each have a single 1-cycle.

Sadly, this charming graphical procedure is not overly amenable to being automated. That is, the computational way to find out where the loops are is to raise the connectivity matrix to powers and examine the traces. But once one is dealing with matrices in a computer, there are more direct and effective ways to utilize them.

For example, testing the pair matrix for an eigenvalue of 2 needs evaluation of the determinant of a readily constructed matrix, and gives the answer straightaway - variance zero means uniform multiplicity. (Constructed from all the fragments of the de Bruijn matrix, the pair matrix is less likely to have Jordan form, which should not be completely forgotten about)

In any event, although the (2,1) Rules do not include any "non-trivial" reversible rules, they contain a wide enough assortment that illustrations can be found for all the precepts of reversible automata theory; even the "trivial" cases are well worth examining thoroughly enough to understand them.

There are still other approaches to reversibility. Many persons have been attracted by the fact that the de Bruijn fragments form a hypercomplex number system; for some rules the fragments resemble Dirac Matrices or other interesting algebraic entities. There may or may not be interesting algebraic structures yet to be found.

15 Corrections to Z

For the sake of continuity, this posting should be inserted in the series "Ancestors" after "Reversible Automata," yet preceding "Ancestors (14)." As such it relates to both reversible evolution rules and the parameter Z (introduced by Andrew Wuensche and Mike lesser in their "Atlas"), in the sense that there are cellular automaton rules between which restrictive relationships exist.

Actually there are many different kinds of relationships, some of them of more interest for one application than for another. For example, given a state set of composite order, it could be a cartesian product, supporting two different automata acting independently. Thus, the $4^6 4$ (4,1) automata include among their number $(2^8)^2$ automata whose rule of evolution is $F((a,b),(c,d),(e,f)) = (f_1(a,c,e), f_2(b,d,f))$. What an insignificant fraction! But if they scatter well through rule space, they could be used as starting approximations to more interesting mutants.

Cartesian products combine different rules within the same neighborhood, but other relationships apply the same rule to different neighborhoods. An important example is composition, which defines the evolution of an automaton for multiple generations: consider $f(a,b,c) = p(p(a,b), p(b,c))$ via which the single generation $(k, 1/2)$ automaton becomes a $(k, 1)$ automaton spanning two generations. Slightly more variety is possible if two different rules alternate between generations. This is the mechanism by which either the identity or the complement, which are rules for $(2, 0)$ automata, promote themselves into rules 204 and 51 amongst $(2,1)$ automata; they are just as reversible in either context.

Another relationship is the one in which the rule of evolution ignores some of the cells in the neighborhood. Allowing such a possibility is not only "mathematically correct" (that is, general); it permits the assumption that neighborhoods are solid blocks (one consecutive interval, in one dimension). In turn, the concept of an "edge sensitive" rule allows distinguishing those automata whose active area does not extend out to the full limits of the neighborhood.

The appendices to Wolfram's "Theory and Applications of Cellular Automata" take note of both possibilities; they are implicit in Erica Jen's Table 8 (she does not identify the factors as $(2, 1/2)$ automata).

Basing a theory of ancestors or of reversible rules on de Bruijn fragments supposes an awareness of their response to edge sensitivity or composition. Rule 51, whose impact on Z is described on page 41 of the Atlas, actually illustrates both concepts. First, edge sensitivity: its A and B matrices are

$$A = \begin{bmatrix} 0 & 0 & . & . \\ . & . & 1 & 1 \\ 0 & 0 & . & . \\ . & . & 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 1 & . & . \\ . & . & 0 & 0 \\ 1 & 1 & . & . \\ . & . & 0 & 0 \end{bmatrix}.$$

whose structure is slightly curious. Note that neither is stochastic, and tht the rule would be assigned $Z = 0$ as a first approximation. If we now define

$$e = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \quad f = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad g = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \quad h = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix},$$

then

$$A = \begin{bmatrix} 0 & e \\ 0 & e \end{bmatrix} = h \otimes e, \qquad B = \begin{bmatrix} f & 0 \\ f & 0 \end{bmatrix} = g \otimes f.$$

In other words A and B are each tensor products of matrices, each of which has triangular form with eigenvalues 0 and 1. The tensor products will have product eigenvalues, which will also be 0 or 1. In fact, (e, f, g, h) is part of a hypercomplex system with this fragment in its multiplication table:

	e	f	g	h
e	e	e	.	.
f	f	f	.	.
g	.	.	g	h
h	.	.	g	h

Dots indicate products (some are zero, some are new matrices) which will not occur in multiplying A and B. Although A and B do not commute, uniform multiplicity can be expected for any product of A's and B's because all the products will have a similar structure.

In fact, e and f can be recognized as a de Bruijn pair for the $(2, 1/2)$ rule $f(x, y) = \bar{y}$, while g and h are a de Bruijn pair for the rule $f(x, y) = \bar{x}$.

Although this example follows some rather clever algebra, it is really quite exceptional. The reduction of A and B to tensor products was only possible because the natural neighborhood of Rule 51 - a single cell - is detached from BOTH edges of the three-cell neighborhood in which it is immersed. However, that is the combination described in the Atlas, and it also provides a good example of a more systematic alternative.

Note that in both the A and B matrices, the bottom halves repeat the top halves as a consequence of the insensitivity of the rule to the left cell of the neighborhood. Insensitivity to the right cell is slightly more subtle, making the even columns repeat their matching odd columns, also visible in both A and B. Compare these remarks with the description of left and right templates in the Atlas, page 40 and thereafter.

With respect to the redundancy arising from ignoring the left cell, the bottom halves of A and B could be discarded without information loss. The rectangular matrices that result could be adjusted for the loss of the left cell as an index by discarding the dots, and sliding the rest of the row over to fill the hole, producing the e and f matrices, but this rearrangement gets them independently of any Kronecker product.

This sleight of hand is quite legitimate if we recall the indexing scheme for a de Bruijn matrix. If the start strings drop down from two cells to one, so should the stop strings. If we discard the first bit of the stop string, we should pick up the only part of the row containing information, which is the part whose initial bit matches the row number.

Since e and f exhibit similar structure to what we have described, but for columns rather than rows, the process can be carried a step further, to account for Rule 51's insensitivity to the right margin as well as the left. We end up with a=0 and b=1. When b=1, 0 maps into 1, and so we confirm that Rule 51 is complementation of the central cell.

If we write the transformation from A to e, B to f, in the form

$$\begin{bmatrix} 0 & 0 & . & . \\ . & . & 1 & 1 \\ 0 & 0 & . & . \\ . & . & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

that is, $AX = Xe$ (and also $BX = Xf$) (X is the rectangular matrix in the above equation), we have a formal process, which maps de Bruijn matrices of one order into those of the next lesser order. A mapping arranged in this form may seem rather mysterious, but it is standard fare in the theory of graphs. Using the formal process provides the best justification for the maneuvers described; it can also be seen underlying procedures described in the Atlas.

So at least we have a formal procedure to deflate edge-insensitive rules, which may be continued until the rule is responsive to both margins. There are also "middle insensitive" rules, such as Rule 90 (page 117), which evolves via the exclusive or of the two frontier cells, and only those two. Then there are the shift rules, clustered with Rule 240 (page 167), which need the $n1$ template. Although a shift rule can be deflated, that goes against the convention of centering the image cell with respect to its neighborhood.

But there is no problem in following the spirit of correcting Z, when such a need arises from the rule having smaller than the nominal size of neighborhood. What would correcting Z do to a classification scheme based on eigenvalues of A? If v were a row eigenvector of A and λ its eigenvalue, then $vA = \lambda v$. But then $(vA)X = v(AX) = v(Xe)$, while $(\lambda v)X = \lambda(vX)$. Hence $(vX)e = \lambda(vX)$, meaning that unless $vX = 0$, A and e have the same eigenvalues.

However, their multiplicity can (and unless X is square and nonsingular, must) differ. For the Perron eigenvalue, it is less likely that $vX = 0$. Also, one can work back from e to A employing column eigenvectors within similar reasoning, to see what might have been lost.

Unless the largest eigenvalue were lost, there would be no penalty for deflating a neighborhood.

16 Ancestors (14)

Commentary related to Andrew Wuensche and Mike Lesser's book, "The Global Dynamics of Cellular Automata." Addison-Wesley, 1992 (ISBN 0-201-55740-1) concludes. Having made an extensive analysis of the role the average number of ancestors (Langton's parameter lambda) plays in classifying and describing cellular automata, some attention needs to be given to Z, a new parameter which the authors have introduced.

We have based our own analysis on a theory of graphs, specifically, the de Bruijn diagram of the automaton, from which ancestors and their properties can be readily calculated. Langton's lambda, which is γ in the formula below, plays a prominent role in this analysis because it represents, within about ten percent, the quantities needed to derive the statistics of the automaton. First, the rate of growth, with length, of the number of ancestors of that string of cells having the greatest number of ancestors. Second, the rate of growth, with length, of the standard deviation in the number of ancestors of whatsoever string of cells.

From then on, the higher moments are simple polynomials with respect to the same parameter; to the extent that it is feasible to turn moments into frequencies, they provide the data which is required. Implicit in this point of view is the assumption that the eigenvalues of selected graphs are the real parameters which ought to be used, but that there are certain advantages to using an easily calculated approximation to them, especially if the same approximation is relevant to all the graphs.

It would seem that the concept of Z arose in the attempt to refine lambda, given that a ten percent approximation is not always sufficient, and that there were observable differences in the behavior of automata having the same value of lambda. Z itself is subject to refinement, as is explained with some care on pages 40, following, in the Atlas. In the formula below, only the first approximation to Z is used. As the layout of the Atlas demonstrates, Z does in fact refine the classification of the automata depicted.

The formulae in question define the eigenvalue of the larger de Bruijn fragment (which is our parameter of preference) in terms of the other two:

$$\begin{aligned}
\text{(Langton) } \lambda &= \gamma + n\sigma(c)\sigma(x)\cos(\theta) \\
\text{(W\&L) } \lambda &= 1 + 2n\sqrt{1-Z}\sqrt{0.5 * s^2 + 0.01561}\cos(\theta^*)
\end{aligned}$$

The Langton version simply takes the fraction of ancestors and discards the correction; the nature of the correction for many rule clusters was tabulated in the previous commentary. The W&L version could be tabulated similarly, but suffice it to say that the new quantity θ^* (always less than 90 degrees because 1 is always an underestimate) ranges over the whole quadrant.

In contrast the W&L version (which is our own invention; the Atlas does not mention such a thing) takes 1 as the basic growth factor, and augments it according to Z and a multiplier which would have to be calculated in each instance. Unfortunately it does not seem to be possible to give a single, constant, empirical estimate for the factor, and proceed from there. Let us emphasize that W&L themselves never ever intimated that such a thing should be expected; we are only describing our own attempts to relate their theory to ours.

If we agree that the growth rates mentioned, however it is that they are estimated or calculated, are the underlying parameters, we can conclude our commentary with some further general observations.

Although the following phenomenon can occur at any level, it is most noticeable for balanced or nearly balanced rules. In previous commentary, the AB = 44 Rule 29 had A and B matrices with maximum eigenvalue 1, yet this Rule is not one of zero variance. The explanation is implicit in the spectral norm of these matrices, the square root of the largest eigenvalue of their product by their transpose. The spectral norm refers to the largest growth factor for any multiplier of this matrix (which need not be a composite of A and B products); it turns out that both AB and BA have larger maximal eigenvalues than either A and B (although their norms cannot be larger than the product of norms of their factors).

The consequence, which is readily apparent from consulting the Atlas, is that neither strings of 1's nor strings of 0's have ancestors which increase in number as the length of the strings increase. However, AB generates strings in which 01 alternate, and these are observed to be strings with maximal preimaging. Can it happen that A, B, AB, and BA have maximal eigenvalue 1, and yet ABB (for example) has a larger eigenvalue? We are not prepared to say.

An earlier commentary contained a histogram for ancestor multiplicity in (2,1) Rule 22 strings of length 8. Examination of page 96 of the Atlas allows us to try out the histogram, although the detail for length 13 is easier to read than the drawings for length 8. There is indeed a variety of in-degrees, although we would not be willing to say that the sample is large enough to draw any conclusions.

It is an interesting question as to whether we have any right to apply the conclusions taken from the de Bruijn diagram for a single generation of evolution to the higher levels of the trees shown in the Atlas. Ancestors of ancestors should be no more free of correlations than descendants of descendants, and probably a great deal harder to detect if correlation is to be discovered in the branches of counterimage trees.

We could either construct multiple-generation de Bruijn diagrams, or simply assume that the correlation is not all that important, which is often a fair assumption.

A respectable portion of the Atlas is devoted to totalistic (2,2) Rules. We do not have much to say about this except to note that experience indicates that totalistic Rules are very atypical; for example, they appear to harbor an undue percentage of Class IV Rules. Nevertheless, the general precepts of graph theory which we have outlined apply to all automata Rules, and it is only the extremely large numbers of automata in categories beyond (2,1) that precludes working them all out, or including them in an Atlas such as this one.

The last few pages of the Atlas contain what may be one of its most interesting offerings, the visualization of the effect of mutations on basins of attraction.

This is also an area which is amenable to treatment by graph theory, and especially matrix theory. Unfortunately, the perturbation theory which serves so elegantly in quantum mechanics, depends very strongly on working with hermitian (or symmetric) matrices, whereas the matrices of graph theory are anything but symmetric (digraph theory, for the purists). There are still general theorems, such as those asserting that any increase or decrease in one single matrix element is immediately reflected by the corresponding modification of the maximum eigenvalue.

Since the connectivity of the de Bruijn fragment has an important bearing on all subsequent analysis, it follows that those mutations which alter connectivity will have a more drastic effect than those which merely change the number of existing connections. By the same token it is much easier to either calculate or just estimate the effect of mutations which preserve the overall connectivity.

In concluding, perhaps a comment ought to be made which is not scientific, but rather has to do with prevailing attitudes towards programs and computing. The Atlas contains a program suitable for IBM PC's and clones, by which the results contained in the Atlas may be reproduced and extended. This is indeed a valuable feature of the Atlas, and allows its owner to investigate many more combinations than ever could have been included on a reasonable amount of paper.

The Authors and/or their Publisher assert a copyright over this program disk, just as they do for the book itself. This is normal practice, and we do not know of anyone who would raise an objection, either to the disk, and especially to the book, being copyrighted.

If we understand what we have read correctly, an additional copyright is asserted over any results obtained through the use of the programs which accompany the Atlas. This is not a concept which has gained general acceptance; it is somewhat like claiming that you have copyrighted the cake after having sold the recipe book. Or more accurately, that such protection extends from the cake mixer to the batter to the cake.

Computer language compilers have been sold with the claim that the copyright inherent in the compiler program also extends to any code compiled through the use of the program. This is somewhat different from compilers which insert run-time code from a copyright subroutine library. In both cases, many users have preferred to use a product which lacks such encumbrances, rather than endure, or risk enduring, unpleasant complications.

It would be a standard element of courtesy to acknowledge the use of any program, such as the one the authors have prepared, in work of one's own; but if this restriction has been understood correctly, a person such as myself would simply write their own program, and be done with it. (We won't talk about the pretensions involved in patenting a program, as that claim has not been made).

We sincerely hope that the phrase "... and any images implicit in the software, ..." on page 61 does not carry the dire implications alluded to above.

Otherwise, in conclusion, we would like to say that we have greatly enjoyed perusing the Atlas, and that we heartily recommend it to anyone who wants to have most of the reasonable information about (2,1) and totalistic (2,2) automata readily at hand, where it can be enjoyed in an especially pleasant visual form.

The commentary is now finished.

17 Mutations

For the sake of continuity, this posting ought to be considered as an afterthought to the series "Ancestors," to be placed after "Ancestors (14)." We really intend the commentary on Wuensche and Lesser's "Atlas" to be finished; but there have been some final observations which depend too much on the background already established to be considered separate submissions.

In the process of correcting Z, it is convenient to deflate edge-insensitive neighborhoods. Deflation does not change the lambda ratio, nor the Perron eigenvalues of the de Bruijn diagrams, but it IS essential for getting a probability-based value of Z. In the opposite direction, inflation produces larger neighborhoods, which may be filled by carrying evolution through multiple generations, or which can be used for a finer degree of mutation than that of which the original neighborhood is capable.

Inflation and deflation are special cases of mappings between graphs, *special* meaning that the mappings are supposed to conserve the linkages and non-linkages within the graph. In turn, inflation and deflation are special cases of a mapping known as *duality.* At first sight, duality consists of exchanging nodes and links; the nodes of the dual are the links of the original. Dual nodes are joined when the links from the original diagram join consecutively, which is whenever they share a common node.

Duality can be expressed by matrices: invent two new matrices, which will probably be rectangular, and are related to one another by diagonal reflection. For the first matrix, the rows are indexed by nodes, the columns by links. For the second matrix, the reverse is true, rows being indexed by links, and columns by nodes. The elements are zeroes or ones, according to the Boolean predicates tail(node,link) and head(link,node) which express the relation of the node of its argument to the (directed) link which is the other argument.

The rules of matrix multiplication make sense of a sum of the form

$$\mathrm{linked}(\mathrm{node1}, \mathrm{node2}) \quad = \quad \sum_{links} \mathrm{tail}(\mathrm{node1}, \mathrm{link}) \times \mathrm{head}(\mathrm{link}, \mathrm{node2}),$$

as well as a sum of the form

$$\mathrm{joined}(\mathrm{link1}, \mathrm{link2}) \quad = \quad \sum_{nodes} \mathrm{head}(\mathrm{link1}, \mathrm{node}) \times \mathrm{tail}(\mathrm{node}, \mathrm{link2}).$$

Supposing that a link has exactly one head and one tail, there is just one node for which tail(node,link) is true, so TAIL must be column-stochastic (with just one non-zero element per column); we could call it C for short. Likewise HEAD is row-stochastic because head(link,node) allows just one non-zero element per column; call the matrix R. Therefore the connectivity matrix M for any graph can always be factored by stochastic matrices into the form M=CR (or TAIL×HEAD).

By definition, the DUAL of M is the graph whose connectivity matrix is RC. Reading this as HEAD×TAIL, the second equation, above, justifies the use of the name. Moreover, the tautology RCR=RCR, written with parentheses R(CR) = (RC)R, shows the relationship of duality to inflation and deflation, because it shows that the dual can be mapped to the original. Hardly surprising; we associate links with their endpoints. Writing CRC=CRC is also possible.

There could be a second dual $R'C'$, wherein $RC = C'R'$, and so on through an infinite tower. That a matrix is equal to its second dual is not necessarily a theorem, as it is in linear algebra; it's another kind of dual. However, all the de Bruijn diagrams for a given state set constitute exactly such a tower, graded according to the length of the neighborhood.

What is more interesting is whether or not a graph can be a dual, which means giving it the reverse factorization. The conditions are essentially those under which deflation is possible. Of course, a complete de Bruijn diagram can always be deflated; the question is whether the fragments also meet the condition, which has something to do with rows (or columns) being either identical or orthogonal. Details may be found in an article by Hemminger and Beinecke[12].

One of the applications of inflation is to assure the existence of a new graph, in which all the paths of a certain length taken from an original graph are single links in the new graph. The way to get one is recursively, which is where the dual enters in. To see this, begin with the de Bruijn matrix for (2,1) automata,

$$\begin{bmatrix} 1&1&.&. \\ .&.&1&1 \\ 1&1&.&. \\ .&.&1&1 \end{bmatrix} = \begin{bmatrix} 1&1&.&.&.&.&.&. \\ .&.&1&1&.&.&.&. \\ .&.&.&.&1&1&.&. \\ .&.&.&.&.&.&1&1 \end{bmatrix} \begin{bmatrix} 1&.&.&. \\ .&1&.&. \\ .&.&1&. \\ .&.&.&1 \\ 1&.&.&. \\ .&1&.&. \\ .&.&1&. \\ .&.&.&1 \end{bmatrix} = CR.$$

The product CR is the de Bruijn matrix for (2,3/2) automata:

$$RC = \begin{bmatrix} 1&1&.&.&.&.&.&. \\ .&.&1&1&.&.&.&. \\ .&.&.&.&1&1&.&. \\ .&.&.&.&.&.&1&1 \\ 1&1&.&.&.&.&.&. \\ .&.&1&1&.&.&.&. \\ .&.&.&.&1&1&.&. \\ .&.&.&.&.&.&1&1 \end{bmatrix} = C'R' = \begin{bmatrix} 1&1&.&.&.&.&.&.&.&.&.&.&.&.&.&. \\ .&.&1&1&.&.&.&.&.&.&.&.&.&.&.&. \\ .&.&.&.&1&1&.&.&.&.&.&.&.&.&.&. \\ .&.&.&.&.&.&1&1&.&.&.&.&.&.&.&. \\ .&.&.&.&.&.&.&.&1&1&.&.&.&.&.&. \\ .&.&.&.&.&.&.&.&.&.&1&1&.&.&.&. \\ .&.&.&.&.&.&.&.&.&.&.&.&1&1&.&. \\ .&.&.&.&.&.&.&.&.&.&.&.&.&.&1&1 \end{bmatrix} \begin{bmatrix} 1&.&.&.&.&.&.&. \\ .&1&.&.&.&.&.&. \\ .&.&1&.&.&.&.&. \\ .&.&.&1&.&.&.&. \\ .&.&.&.&1&.&.&. \\ .&.&.&.&.&1&.&. \\ .&.&.&.&.&.&1&. \\ .&.&.&.&.&.&.&1 \\ 1&.&.&.&.&.&.&. \\ .&1&.&.&.&.&.&. \\ .&.&1&.&.&.&.&. \\ .&.&.&1&.&.&.&. \\ .&.&.&.&1&.&.&. \\ .&.&.&.&.&1&.&. \\ .&.&.&.&.&.&1&. \\ .&.&.&.&.&.&.&1 \end{bmatrix}$$

wherein the neighborhoods are of length 4, which are two-step paths of neighborhoods of length 3.

In working with duality it is convenient to note that the product of two C-matrices (column stochastic) is a C-Matrix, that the product of two R-matrices is always an R-matrix, and that the factorization M=CR is always possible. Fragments can be included in a computation, if an epsilon is included amongst the matrix elements, which can take

[12] Robert L. Hemminger and Lowell W. Beinecke, "Line graphs and line digraphs", appearing in *Selected Topics in Graph Theory* edited by Lowell W. Beinecke and Robin J. Wilson, published by Academic Press in 1978.

values 0 or 1. Epsilon must always appear in at least one factor, but if desired it can always be placed in the factor on one side exclusively.

Consider the a matrix like M^3, whose elements count the number of paths of length 3. We have

$$\begin{aligned} M^3 &= CRCRCR \\ &= CC'R'C'R'R \\ &= CC'C''R''R'R. \end{aligned}$$

In other words, it is always possible to factorize a product as a single CR combination by invoking a high enough dual. Moreover, we could write

$$M^3 = (CC')(C''R'')(R'R)$$

with any epsilons in the middlemost term, thereby relating three-step paths directly to the second dual. This circumstance justifies the assertion that there is always a big enough de Bruijn diagram to inflate a given neighborhood by any desired amount, and that it then includes the paths of matching length.

In the Atlas, inflation is used to obtain finer detail when constructing mutants. Thus, a (2,1) rule is promoted to a (2,2) rule by means of the definition

$$F(a,b,c,d,e) = f(b,c,d).$$

It could equally well have been promoted to a (2,3/2) rule, a (2,3) rule, or any other; the advantage of skipping (2,3/2) is inflating symmetrically (besides the still finer gradation). As a variant on the theme, had the authors also considered the extension

$$F(a,b,c,d,e) = f(f(a,b,c), f(b,c,d), f(d,e,f)),$$

they could have compared the behavior of second generation ancestors with that of first generation ancestors.

There are two motives for describing duals; one is to show explicitly how the generation of mutants as described in the Atlas fits the matrix and graph theory approach to automata; in particular, the way that Perron eigenvalues are conserved. They are just as much governing parameters as any other which have been introduced, and we have already explored their relation to parameters such as lambda and Z at considerable length. The second motive is to evaluate the direction that hypercomplex algebra might take if the de Bruijn fragments are to be regarded in those terms. Mainly we see that there is already considerable structure present, just from graph theory alone.

The Atlas tends to confirm what has always been suspected — small changes in the rule of an automaton result in small changes in the evolution itself. It isn't all that easy to quantify such a statement, but the exceptions and anomalies which have been observed along with the regularities might be explained through reference to the diagrams, wherein the opening or closing of loops is a significant event.

There is a whole lore of bounds on the Perron eigenvalue, and also of its separation from the next largest eigenvalue, which has an effect on the rate of convergence of powers of the matrix to equilibrium. By and large, the results are probably not stringent enough to allow predictions about the basins of attraction to be made in the detail shown in the

Atlas. However, it might be interesting to do a Monte Carlo simulation of the basins on the basis of the data which we actually do have.

There is an entirely different approach which we have not mentioned at all; the algebra of regular expressions. For example, one might characterize the reversible rules as those for which the regular expression belonging to the A and B fragments must be a sum of starred expressions, none of which contains a sum within a star. That is a concise way of saying that the loops cannot be connected to one another, necessary but not sufficient for reversibility.

In any event, the series really is concluded. A report summarizing the whole commentary is now available on request, as are the two preprints on which it is based. With the help of TeX, many misspellings have been corrected, footnote references added, and formulas beautified. A supplement contains a variety of the smaller de Bruijn diagrams, pair diagrams, and subset diagrams; they may be copied and used as suggested in the commentary.