



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Instituto de Física - INFIS

Pós-Graduação em Física

CIRO ALVES JUSTINO NUNES

ALGORITMO DE LEMPEL-ZIV APLICADO À CLASSIFICAÇÃO
QUANTITATIVA DE AUTÔMATOS CELULARES

Uberlândia – MG

2014

CIRO ALVES JUSTINO NUNES

ALGORITMO DE LEMPEL-ZIV APLICADO À CLASSIFICAÇÃO
QUANTITATIVA DE AUTÔMATOS CELULARES

Dissertação apresentada ao Instituto de Física
da Universidade Federal de Uberlândia como
requisito parcial para a obtenção do título de
Mestre em Física.

Orientador: Prof. *Dr. Raimundo Lora Serrano*

Coorientador: Prof. *Dr. Ernesto Estévez Rams*

Uberlândia – MG

2014

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

N972a Nunes, Ciro Alves Justino, 1989-
2014 Algoritmo de Lempel-Ziv aplicado à classificação quantitativa de
autômatos celulares / Ciro Alves Justino Nunes. - 2014.
125 f. : il.

Orientador: Raimundo Lora Serrano.
Dissertação (mestrado) - Universidade Federal de Uberlândia,
Programa de Pós-Graduação em Física.
Inclui bibliografia.

1. Física - Teses. 2. Cristalografia - Teses. I. Serrano, Raimundo
Lora. II. Universidade Federal de Uberlândia. Programa de Pós-
Graduação em Física. III. Título.

CDU: 53



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA



CIRO ALVES JUSTINO NUNES

Dissertação apresentada ao Programa de Pós-graduação em Física, do Instituto de Física, da Universidade Federal de Uberlândia, como requisito parcial para obtenção do título de Mestre em Física.

Uberlândia, 02 de dezembro de 2014.

BANCA EXAMINADORA

Prof. Dr. Marco Aurélio Boselli
Universidade Federal de Uberlândia- UFU

Prof.^a Dra. Tânia Tomé Martins de Castro
Universidade São Paulo – USP

Prof. Dr. Raimundo Lora Serrano
Universidade Federal de Uberlândia - UFU

*À minha mãe,
ao meu pai
e à minha avó.*

Agradeço,

Ao meu orientador, Dr. Raimundo Lora Serrano... primeiramente por ter aceitado me orientar ante as complicadas condições em que aceitou, por ter tido a visão e a gentileza me colocar em uma pesquisa desta natureza, e, afinal, por exprimir significativa fé nos complicados processos mentais que verbalizaram este texto;

Ao meu Coorientador, Dr. Ernesto Estevez Rams... pelas mui preciosas instruções teóricas e por ter permitido os visionários caminhos que este trabalho almeja como futuro;

À CAPES pelo fundamental apoio financeiro e à UFU e ao INFIS pelo acréscimo de tempo necessário para a conclusão deste;

À Jéssica... pela inspiração moral, intelectual e pessoal, e sem cuja companhia dificilmente haveria sustento ou sanidade que permitissem qualquer coisa ser feita no período em que assim foi;

E, novamente, à mágica cadeia de eventos cósmicos... a qual, porém, desta vez, por sugerir que, calculando absurdos em máquinas abstratas, um dia eu possa vir a descobrir que o Universo não precisa de um tamanho para ser regular, inteligível e perfeitamente infinito...

RESUMO

O trabalho apresenta os chamados Autômatos Celulares Elementares de acordo com os princípios do algoritmo de Lempel-Ziv (LZ76) aplicado a sequências binárias. Tal descrição visa a organização quantitativa desses autômatos de acordo com a complexidade de atualização dos estados, relacionando os arranjos com a Classificação de Wolfram em função de valores numéricos sugeridos. É possível, dessa forma, discriminar tais máquinas em Classes de Complexidade. Ainda, serão percorridas as chamadas Sequências de Máxima Complexidade – MLZs – e suas propriedades; estas serão úteis na caracterização de emissões ergódicas verdadeiramente randômicas, e, portanto, na compreensão do grau de aleatoriedade das regras aplicadas aos autômatos. Esta teoria será apresentada como fundamentação para um novo tratamento de problemas de defeitos em sólidos cristalograficamente complexos (ordem atômica intermediária entre a cristalinidade ideal de longo alcance e a amorficidade), utilizando métodos da Teoria da Informação.

Palavras-chave: Autômatos Celulares, Complexidade de Lempel-Ziv, Fonte Ergódica, Teoria da Informação, Entropia Algorítmica, Sólidos Complexos, Cristalografia, Dimensão Fractal, Fatorização, Aleatoriedade, Cadeias de Dados.

ABSTRACT

This work presents the so-called elementary Cellular Automata according to the principles of the Lempel-Ziv (LZ76) algorithm applied to binary sequences. We aim at organizing them quantitatively in agreement with the complexity of updating the states, and by relating the data arrangements to the Wolfram's Classification. In this way, Complexity Classes can classify such machines. Further, sequences with maximum LZ complexity - MLZs - and their properties will be presented and discussed. The latter will be useful to characterize the truly random ergodic emissions, thus to understand the degree of randomness of the rules governing automata. The above treatment will be extended to problems of crystal defects in crystallographically challenging solids (with intermediate atomic order between the ideal long-range crystallinity and amorphicity), by using tools from information theory.

Keywords: Cellular Automata, Lempel-Ziv Complexity, Ergodic Source, Information Theory, Algorithmic Entropy, Crystallography, Crystallographically Challenging Solids, Fractal Dimension, Factorization, Randomness, Strings.

ÍNDICE

LISTA DE TABELAS.....	2
LISTA DE FIGURAS.....	2
INTRODUÇÃO	6
PRIMEIRA PARTE - TEORIA.....	14
I - CADEIAS DE DADOS	14
I.1 - CADEIAS DE DADOS	15
I.2 - TAMANHO DA CADEIA (L, N)	15
I.3 - BASE NUMÉRICA OU CARDINALIDADE DO ALFABETO ($B, \kappa, / \Sigma /$).....	16
I.4 - ENTROPIA	16
I.5 - COMPLEXIDADE.....	21
I.6 - DIMENSÃO FRACTAL	26
II - AUTÔMATOS CELULARES	38
II.1 - AUTÔMATOS CELULARES	38
II.2 - AS CLASSES DE WOLFRAM.....	41
II.3 - ADITIVIDADE EM AUTÔMATOS.....	46
SEGUNDA PARTE - RESULTADOS E DISCUSSÕES.....	50
III - RESULTADOS E DISCUSSÕES. CADEIAS DE DADOS.	50
IV - RESULTADOS E DISCUSSÕES. AUTÔMATOS CELULARES.....	64
CONCLUSÕES.....	96
REFERÊNCIAS.....	101
ANEXO.....	104
ANEXO A.....	104

LISTA DE TABELAS

Tabela 1 - Regras Aditivas	49
Tabela 2 - Diversos desvios relativos para as complexidades e comprimentos	62

LISTA DE FIGURAS

Figura 1 - Camada compacta com as três posições possíveis e os sítios intersticiais que a mesma gera.....	10
Figura 2 - Representação de uma estrutura periódica através do polítipo correspondente	10
Figura 3 – Entropia x Probabilidade: Curva de entropia para sequências geradas por lançamento de moedas	20
Figura 4 – Quadrado: construção plana ortogonal em espaço 2D.....	27
Figura 5 – Subdivisão do quadrado em partes autossimilares ao todo e idênticas entre si.....	27
Figura 6 – Ilustração simples da geometria de um Conjunto de Cantor. Cada passo da construção gera autossimilaridade ao todo, em menor escala (Fonte: Wikipédia).....	29
Figura 7 - Caixas preenchendo uma figura genérica com objetivo de destacar o espaço preenchido pela forma.....	31
Figura 8 - Autômato Genérico. Exemplificação de uma máquina que evolui. A organização dos símbolos e a regra de atualização não necessariamente existem.....	38
Figura 9 - Autômato Unidimensional. Visualização de uma máquina circular estendida sobre uma linha. As extremidades são regiões vizinhas.....	39
Figura 10 - Autômato Celular Elementar. Máquina unidimensional construída em base 2, na qual a atualização de uma célula depende unicamente dos primeiros vizinhos da mesma.....	40
Figura 11 - Autômato totalístico exterior. A atualização não depende de uma operação matemática sobre a vizinhança, apenas dos estados em si das células vizinhas, ditas exteriores: b se atualização para d quando a e c forem suas vizinhas.....	40
Figura 12 - Regra 32 (trivial). Convergência a um estado único para todas as células.....	42
Figura 13 - Regra 234 (trivial). Convergência a um estado único para todas as células.....	42
Figura 14 - Regra 4 (periódico). Convergência a um estado único.....	43
Figura 15 - Regra 202 (periódico). Convergência a um estado único.....	43
Figura 16 - Regra 139 (periódico). Convergência a um estado único com rotação.....	43
Figura 17 - Regra 88 (periódico). Convergência a um estado único com rotação.....	43
Figura 18 - Regra 22 (complexo). Formação de estruturas com distribuição aparentemente aleatória.....	44
Figura 19 - Regra 182 (complexo). Formação de estruturas com distribuição aparentemente aleatória.....	44

Figura 20 - Regra 75 (caótico). Sem formação de estruturas aparentes e distribuição aleatória.	45
Figura 21 - Regra 45 (caótico). Sem formação de estruturas aparentes e distribuição aleatória.	45
Figura 22 - Regra 57. Eventual formação de estruturas, ou repetição, mas com aspecto complexo.	46
Figura 23 - Regra 73. Regiões do arranjo com aspecto periódico e regiões com aspecto randômico.	46
Figura 24 - Regra 106. Estruturas não necessariamente autossimilares (em forma), e com distribuição aparentemente aleatória.	46
Figura 25 - Regra 150. Estruturas não necessariamente autossimilares (em cor), e com distribuição aparentemente aleatória.	46
Figura 26 - Regra 225. Estruturas não necessariamente autossimilares (em forma), e com distribuição aparentemente aleatória.	46
Figura 27 – Contagem x Comprimento. Distribuição dos comprimentos dos fatores de uma MLZs (em preto) e da média para 1000 seqüências aleatórias (em vermelho), ambas com 10^5 símbolos.	52
Figura 28 – Contagem Normalizada x Comprimento. Distribuição normalizada dos fatores pelo número de combinações disponíveis para cada comprimento de fator, das MLZs (em preto) e a média de 100 cadeias aleatórias, ambas com 10^5 símbolos.	53
Figura 29 - Complexidade x Número de Símbolos. Em azul é o valor real de LZ76 para a MLZs e em vermelho o valor gerado pela equação (III.d).	55
Figura 30 - Erro Relativo da Complexidade x Número de Símbolos. Erro relativo entre o valor real e o valor calculado pelo ansatz (III.d).	55
Figura 31 - Contagem de Subcadeias x Ordem (Base 2). A contagem do número de fatores LZ76 de mesmo comprimento relativamente a este comprimento dentro das MLZs	56
Figura 32 - Contagem de Subcadeias x Ordem (Base 3). Idem anterior.	56
Figura 33 - Contagem de Subcadeias x Ordem (Outras Bases). Idem anterior.	57
Figura 34 - Ajuste teórico da contagem (Base 2). No primeiro gráfico, a distribuição analítica (III.e) em vermelho se ajusta com adequação sobre os pontos reais, em azul. O segundo gráfico mostra o Desvio Percentual do Valor Real x Comprimento das Subcadeias, e o decréscimo deste desvio com o aumento da ordem de cálculo.	58
Figura 35 - Ajuste teórico da contagem (Base 3). Idem anterior.	59
Figura 36 - Ajuste teórico da contagem (Base 4). Idem anterior.	60
Figura 37 - Ajuste teórico da contagem (Base 5). Idem anterior.	61
Figura 38 - Ajuste teórico da contagem (Base 6). Idem anterior.	62

Figura 39 - Distribuição de Tc Médios x Complexidades Médias (RLE e LZ76) para Classe I, ambas para a Regra 32 e para a Regra 202.	66
Figura 40 - Distribuição de Tc médios para Classe II, nos moldes da figura anterior.	67
Figura 41 - Distribuição de Tc médios para Classe III. Idem anterior.	67
Figura 42 - Distribuição de Tc médios para Classe IV. Idem anterior.	68
Figura 43 - Distribuição de τ médios para Classe I. Os períodos médios contra complexidades RLE e LZ76, ambos para a Regra 32 e para a Regra 202.	69
Figura 44 - Distribuição de τ médios para Classe II, nos moldes da figura anterior.	70
Figura 45 - Distribuição de τ médios para Classe III. Idem anterior.	70
Figura 46 - Distribuição de τ médios para Classe IV. Idem anterior.	71
Figura 47 - Distribuição dos Períodos x Tempos Iniciais (Classe I). Distribuições de tempos para a Regra 32 e para a Regra 252.	72
Figura 48 - Distribuição dos Períodos x Tempos Iniciais (Classe II). Idem anterior.	73
Figura 49 - Distribuição dos Períodos x Tempos Iniciais (Classe III). Idem anterior.	73
Figura 50 - Distribuição dos Períodos x Tempos Iniciais (Classe IV). Idem anterior.	74
Figura 51 - Polígono de Sequência. Distribuição poligonal de pontos em que cada vértice representa um caractere preenchido em uma sequência binária cujas extremidades são células vizinhas.	75
Figura 52 - Distribuição de Pontos do Polígono. Pontos marcados representam 1 e pontos ausentes representam 0.76	
Figura 53 - Centro de Massa da Cadeia. Ponto representativo da localização ponderada das distribuição poligonal de pontos.	77
Figura 54 - Evolução do Centro de Massa. a) Arranjo do autômato celular, b) distribuição das coordenadas dos pontos de massa de cada estado de atualização do autômato, c) distribuição das coordenadas dos pontos de massa de cada estado de atualização do autômato segundo a evolução temporal (terceiro eixo).	78
Figura 55 - Evolução do Centro de Massa (condição inicial aleatória). Idem anterior.	79
Figura 56 - Evolução do Centro de Massa (Autômato muito grande e condição inicial aleatória). Idem anterior.	80
Figura 57 - Evolução do Centro de Massa (Regra 20 e L = 51). Idem anterior.	81
Figura 58 - Evolução do Centro de Massa (Regra 20 e L = 1000). Idem anterior.	82
Figura 59 - Evolução do Centro de Massa (Regra 169 e L = 51). Idem anterior.	83
Figura 60 - Evolução do Centro de Massa (Regra 169 e L = 51). Idem anterior.	84

Figura 61 - Distribuição dos Períodos x Tempos Iniciais (Regras Aditivas). Períodos contra tempos mostrando a pequena variedade da distribuição nessas regras.	86
Figura 62 - Centros de Massa das Regras Aditivas. Distribuição e evolução dos centros de massa nessas regras.	87
Figura 63 – Estado x Passo. Oscilações dos Autômatos. Essas oscilações representam a ocorrência de ciclos em determinado autômato, i. e., uma determinada regra aplicada em uma determinada condição inicial.	92
Figura 64 - Distribuição das complexidades para todas as regras elementares.....	94
Figura 65 – Entropia de Bloco RLE.....	108
Figura 66 – Média temporal da Entropia de Bloco RLE.....	109
Figura 67 – Entropia Média para diversas condições iniciais aleatórias	110
Figura 68 – Distribuição das entropias pelas regras (normalizadas)	111
Figura 69 – Distribuição das entropias pelas regras.....	112

INTRODUÇÃO

Quando se trata de verificação de hipóteses em ciência, a experimentação é frequentemente o método mais utilizado, haja vista a correspondência com o mundo físico, aparentemente tornando mais autêntica a previsão dos eventos com base nas tendências. Dentro da classe de correspondências buscadas pela experimentação não somente podem se encaixar os testes com o mundo real, mas as simulações baseadas na conceptualização desse mundo também podem se mostrar igualmente úteis.

Tal conceptualização dos eventos naturais tem sido traduzida (dentre outras formas) por meio de leis passíveis de simbolização matemática, capazes de se acomodar ao substrato de processamento neurológico humano. As simulações fundadas nessas leis devem expressar de forma recíproca as regularidades encontradas no mundo natural e podem ser feitas através de máquinas capazes de calcular rapidamente as soluções das diversas modelagens. Tais simulações são escritas usando regras com possibilidade de serem executadas levando uma entrada de dados qualquer (uma condição inicial) até a solução do problema.

O fato de esses instrumentos serem utilizados para auxiliar a capacidade de cálculo do cérebro humano tem como consequência que sua construção possa ser baseada no próprio modo humano de pensar. Assim, por exemplo, os circuitos de execução lógica de tarefas são semelhantes à lógica linear de causa e efeito que dá sentido ao pensamento racional desenvolvido ao longo dos milênios: *algo é verdadeiro a menos que seja falso*. Essa edificação binária da racionalização do mundo é a mais simples possível e, por este motivo, está contida nos fundamentos dos computadores. Contudo, a ideia de construir centros de processamento baseados em lógica binária, à imagem e semelhança da lógica humana, tem a importuna consequência de o código de máquina ser difícil de ler por quem deseja programar uma regra, fazendo com que fossem criadas linguagens de montagem (*assembly*) mnemônicas e facilmente legíveis para aproximar e traduzir a escrita do programador para o código da máquina.

Neste contexto se encaixa o presente trabalho. Serão apresentados a seguir vários conceitos essenciais ao corpo do trabalho, começando pela definição de **Informação**, seguido da definição de **Complexidade e desordem** do ponto de vista da **teoria da informação**. O que são chamados **sólidos complexos do ponto de vista cristalográfico** segue na sequência com o intuito de ilustrar a “semente” científica desta investigação usando ferramentas da Teoria da Informação. Assim, será comentado sobre como relacionar **o problema cristalográfico** de estudo de estruturas com defeitos e **uma abordagem diversa** (usando teoria da informação), como tem

sido feito em outros ramos da ciência contemporânea. Em particular, a abordagem será feita tendo em vista o estudo de defeitos planares em **estruturas compactas de camadas**, as quais são sucintamente explicadas de maneira a facilitar a associação do empilhamento das camadas e a **codificação binária** das mesmas em função do tipo de empilhamento compacto. Por fim, serão discorridos os conceitos de **Complexidade das cadeias de dados** e a **classificação dos autômatos celulares** na qual o trabalho será baseado.

Informação

Os valores brutos do código de máquina, verdadeiro e falso, são frequentemente colocados como passagem ou interrupção de força motriz (circuito elétrico ou mecânico), por vezes expressos em dígitos binários, e as sequências de instruções passadas à máquina podem ser imaginadas dessa forma. Entretanto, essas sequências não são providas de materialidade, não possuem carga elétrica, massa, nem extensão espacial, mas o fato de serem manipuláveis, trabalháveis, indica que possuem existência (virtual) e por este motivo são dotadas de informação.

Informação é a propriedade mais fundamental de qualquer coisa que existe, real ou virtualmente. O dicionário define informação como “tudo aquilo que, por ter alguma característica distinta, pode ser ou é apreendido, assimilado ou armazenado pela percepção e pela mente humanas” (FEREIRA, Aurélio Buarque de Holanda, 2004). É, naturalmente, um efeito da observação humana do mundo – por “humano” entenda-se que interage inteligentemente –, não se podendo dizer com certeza que exista quando não há quem observe. Assim, quantificar a informação de modo a ser estudada pela ciência pode ser muito conveniente.

Por outro lado, costuma-se pensar que um sistema que exige mais termos (de qualquer natureza) para ser classificado possui mais informação que outro com menos termos ou que necessite de mais informação de natureza descritiva. Um sistema assim tem seus constituintes dispostos em uma organização mais dificilmente qualificada – e desde que se relacione uma quantidade àquela qualidade – e quantificada; por exemplo, a quantidade de termos que expressam toda a natureza culinário-físico-química de um torrão de rapadura é maior do que a quantidade necessária para expressar uma molécula de sacarose. Matematicamente, o número de combinações disponíveis para esses elementos deverá ser uma expressão de quão complexa pode ou não ser o seu arranjo particular. Dessa forma, definir uma grandeza cujo argumento é o

número de combinações disponíveis para as sequências de instruções dadas à uma máquina será também definir quanta informação está contida naquele processo.

A partir do momento em que se sabe quanta informação um processo contém, pode-se saber quanta informação pode ser transferida, quanta pode ser perdida, quanta pode ser retida e assim por diante. Assim, toda uma variedade de características relacionadas às máquinas que simulam os processos naturais podem ser elencadas a fim de se saber qual é a real qualidade daquele processamento (ou virtualização de experimento).

Complexidade, desordem e teoria da informação

O estudo de sistemas aleatórios e a emergência de complexidades é um dos assuntos que maior atenção cria na comunidade científica na atualidade (para uma revisão do desenvolvimento atual: [26]). São comuns na literatura termos tais como complexidade, caos determinístico ou dinâmica não linear, todos aplicados em áreas como a física, a química, a biologia, dentre outras. Muito apesar da generalidade dos termos acima, são poucos os estudos quantitativos sobre desordem e o aparecimento de complexidade. A própria definição do termo “complexo”, e como se quantifica, é tema de debate [27].

Nesse cenário, os estudos da relação desordem-complexidade tem ido cada vez mais de aproximações gerais ao estudo em sistemas específicos [28, 29, 30], em concordância com a evolução da ciência nesses aspectos: amadurecimento de ideias novas a partir de seu uso intensivo e extensivo em casos específicos.

Após a II Guerra Mundial, Claude Elwood Shannon estabeleceu as bases de como medir a informação de uma fonte a partir de considerações probabilísticas [31]. Ele demonstrou que a incerteza média (taxa de entropia) é a propriedade fundamental para descrever uma fonte estocástica. Interessado na transmissão de informação entre fonte e receptor através de canais de diferentes naturezas e sujeita a ruído, definiu a informação mútua para medir a exatidão na transmissão. Ambos os conceitos, entropia da informação e informação mútua, ultrapassaram o objetivo para o qual foram definidas inicialmente e hoje são utilizadas em uma grande variedade de sistemas produtores de informação. É comum que tais sistemas ultrapassem o milhão ou até a dezena de milhões de dados e seu estudo é feito através de cálculo intensivo utilizando grandezas entrópicas.

Sólidos complexos do ponto de vista cristalográfico

Um sólido cristalino é um arranjo de átomos caracterizados por uma ordem perfeita de longo alcance. O conceito de ordem atômica de longo alcance entende-se pela capacidade de se reproduzir a estrutura completa do sólido a partir do conhecimento da posição dos átomos em um volume finito e um conjunto finito de regras. O descobrimento do ordenamento perfeito não periódico abriu novos caminhos na pesquisa sobre cristalinidade [32]. Por outro lado, os sólidos amorfos apresentam ordem (imperfeito) de curto alcance e carecem de correlação além das distâncias aos primeiros vizinhos de qualquer átomo. Entre a desordem amorfa e a ordem cristalina há um intervalo de desordem e ordem estrutural que tem sido pouco explorado pela comunidade da física do estado sólido. Muitos materiais de interesse prático e teórico se encaixam precisamente nesta área pouco explorada. Talvez o exemplo mais conhecido de tais sistemas são os materiais nanocristalinos, onde o termo “cristal infinito” já não pode ser assumido, mas ainda existem correlações entre diferentes regiões nanocristalinas do material enquanto a ordem de longo alcance é frustrada de diferentes maneiras [33].

Os sólidos que exibem uma complexa arquitetura de átomos são chamados de sólidos complexos do ponto de vista cristalográfico. Os métodos computacionais menos intensivos de estudo de sólidos cristalinos, bem como os critérios estatísticos usuais para o estudo de materiais amorfos não são suficientes para o estudo dos materiais complexos. O surgimento de regularidades locais, correlações de longo alcance e outros fenômenos em diferentes escalas, convertem as simulações computacionais de sistemas de muitas variáveis em ferramenta essencial para a análise destes sistemas. O número de variáveis pode ir de algumas centenas a milhões delas.

O problema cristalográfico e uma abordagem diferente

O estudo computacional de sólidos complexos, como os mencionados acima, pode ser feito a partir do uso de conceitos vindos da teoria da informação tais como sistemas dinâmicos, teoria ergódica, dentre outros. Existem descrições baseadas na teoria computacional de cadeias de dados para seu uso em bioinformática e outras aplicações de “mineração” de dados. Em outros casos, o caminho seguido baseia-se na experiência adquirida de alguns procedimentos utilizados para descobrir o surgimento de complexidade em sistemas dinâmicos em campos tão variados como a neurociência [34] ou biossinais [35].

Estruturas compactas de camadas

Dentre os materiais complexos do ponto de vista cristalográfico se encontram aqueles cuja estrutura cristalina pode ser descrita por camadas de átomos empilhadas ao longo de uma determinada direção cristalina e, especificamente, por empilhamentos compactos. No caso ideal, o empilhamento se apresenta de forma ordenada, isto é, composto por repetições periódicas de uma de suas partes. Em outros casos, existe desordem ligada aos defeitos do empilhamento, como no caso das transições de fases cristalográficas (durante as quais o sólido pode passar por uma série de fases ordenadas e desordenadas).

As estruturas cristalinas de camadas consistem de planos de átomos agrupados de forma compacta, situados uns sobre os outros, não necessariamente de maneira periódica ao longo da direção perpendicular à camada. A forma mais compacta de juntar esferas num plano é esquematizada na figura seguinte (codificada com a letra A)

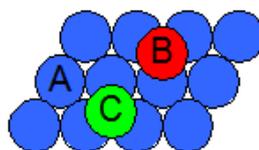


Figura 1 - Camada compacta com as três posições possíveis e os sítios intersticiais que a mesma gera

A camada “A” consiste de átomos dispostos em um arranjo hexagonal. As camadas “B” e “C” situam-se na parte superior da camada “A”, mas os átomos das respectivas camadas “B” e “C” ocupam locais diferentes com relação aos átomos na camada “A”. Acima de “C” volta a aparecer uma camada “A”, depois uma camada “B”, e assim sucessivamente. Ao substituir a camada “C” pela “A” obtém-se um arranjo atômico diferente.

Com a codificação de letras, o estado de compactação se reduz a que não podem aparecer juntas duas letras idênticas. No caso em que a estrutura compacta é periódica, é suficiente representa-la como a repetição de um bloco mínimo de camadas que chamamos de politipos, que são casos especiais de polimorfismos [36]. A Fig. II mostra as duas estruturas compactas mais simples às quais podem ser associados os códigos ABC e AB. São elas, a estrutura cúbica de face centrada (FCC) e hexagonal compacta (HCP), respectivamente.

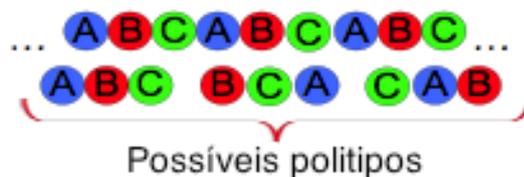


Figura 2 - Representação de uma estrutura periódica através do politipo correspondente

Codificação binária de estruturas de camadas

Como em uma estrutura de camadas compacta têm-se apenas três posições possíveis para as camadas (ver acima), é possível codificar uma determinada sequência considerando o deslocamento relativo de duas camadas sucessivas e não suas posições absolutas. Essa codificação é conhecida como código ou notação de Hägg [37], que atribui um sinal de “+” ao ir “para frente” na sequência (AB, BC, CA) e um sinal “-” quando vai “para trás” (BA, CB, AC). Uma notação parecida pode ser adotada se associamos um “1” para substituir o sinal “+” e um “0” para substituir o sinal “-”. Dessa forma, obtém-se um código binário de “zeros” e “uns”. Assim, por exemplo, a notação de Hägg das sequências abaixo são indicadas à direita:

- +++--- (politipo 6H, ABCACB) → 111000;
- ++-+- (politipo 6H, ABCBCB) → 110100;
- +++++ (politipo 5H, ABCBC) → 11110;
- 2H (ABABABAB...) e 3C (ABCABCABC...) → 2H (1010101...) e 3C (11111111...)

Aproximação entrópica à análise de politipos

Com o objetivo de utilizar uma abordagem diferente à análise de politipos, é possível usar conceitos entrópicos de informação para classificar e estudar a ordem de empilhamento nas estruturas de camadas e as transformações de fase estruturais. Um de tais modelos é conhecido como mecânica computacional [38], que se fundamenta no conceito de entropia de Shannon. Vários quantificadores de entropia têm sido introduzidos para descrever o surgimento de ordem. Conceitos tais como excesso de entropia, densidade de entropia, informação transiente, foram utilizados para tais propósitos e aplicados a sequências periódicas unidimensionais [39]. Tais quantificadores fornecem informação interessante que não está contida em outras grandezas como a correlação, susceptibilidade ou o fator de estrutura. Nesse sentido, os mesmos parecem complementar outras grandezas conhecidas da física estatística.

Complexidade das cadeias de dados

O presente trabalho está relacionado com a aproximação entrópica chamada de complexidade de Lempel-Ziv (LZ76) [6]. A mesma tem sido utilizada para analisar o DNA [40] e fontes neuronais [34]. Em 1978, Ziv demonstrou a relação entre LZ76 e a densidade de entropia definida por Shannon. Em particular, a complexidade LZ76 normalizada irá tender, no limite de

uma sequência infinita, à taxa de entropia¹ de uma fonte ergódica [41]. Dessa forma, a LZ76 é uma alternativa, prática, para estimar a densidade de um sistema e (também) fornece informação sobre o limiar de caos como ferramenta de comparação com outros métodos.

A capacidade que uma cadeia de dados tem de ser comprimida pode ser analisada através do método LZ76. A fatoração ou compressão da cadeia pode representar a complexidade da disposição dos constituintes. A complexidade LZ76 é considerada como uma medida de aleatoriedade não probabilística. É sempre possível estabelecer um paralelo da complexidade LZ76 à chamada Complexidade de Kolmogorov-Chaitin – Complexidade KC [5]. A última, mede o comprimento do programa mais curto, executado em uma Máquina Universal de Turing, que permite reproduzir a sequência original. Está relacionada com a aleatoriedade de uma determinada sequência de dados infinita e maiores valores de complexidade estão associados à maior aleatoriedade. A complexidade de KC é o método mais abrangente para este tipo de análise.

Por via da construção de LZ76 será tratado um algoritmo que gera sequências de máxima complexidade de LZ76, chamadas de MLZs (Maximum Lempel-Ziv Complexity sequences [2, 2.1], construídas a partir de um algoritmo muito menor que a sequência em si. As MLZs têm, para um comprimento finito N, maior complexidade LZ76 do que as cadeias aleatórias típicas com o mesmo comprimento. A natureza algorítmica das MLZs as tornam opostas ao conceito de aleatoriedade no sentido da complexidade KC. Neste âmbito, as MLZs servem para normalizar a complexidade LZ76 obtidas por diferentes formulações. Serão discutidas algumas características das MLZs, tais como a distribuição de fatores, tamanhos; dois ansatz obtidos para modelar a previsão de sua complexidade serão apresentados devido à sua utilidade para a determinação das taxas de entropia do teorema de Ziv.

Classificação dos autômatos celulares

As sequências numéricas são muito usadas para expressar o estado de máquinas de estado finito (ou autômatos de estado finito), as quais são modelos matemáticos de computação utilizados para representar circuitos lógicos e programas de computador, e se desenvolvem de maneira autônoma a partir de uma condição (condição desencadeadora). São consideradas como

¹ *Taxa de entropia é uma medida, de comprimento invariante, da quantidade de nova informação obtida por unidade de tempo em um processo dinâmico [T. M. Cover and J. A. Thomas, **Elements of information theory. Second edition**, Wiley Interscience, New Jersey, 2006].*

máquinas abstratas que podem estar apenas em um estado, em um determinado instante de tempo, dentre um número finito de estados. Dentro dessa classe de máquinas, encontram-se os chamados Autômatos Celulares Elementares, os quais serão estudados com o objetivo de classificá-los quantitativamente e estabelecer a relação com as sequências que os formam.

Stephen Wolfram trabalhou extensamente esses autômatos, criados por von Neumann a partir das sugestões de Ulam quando em Los Alamos, e estabeleceu uma classificação qualitativa fundada nos aspectos dos diagramas gerados pela sua evolução, as chamadas Classes de Wolfram. De acordo com Wolfram, os autômatos podem ser classificados em triviais, periódicos, complexos e caóticos. Será sugerida uma tendência matemática que quantifica a distinção entre essas classes. A análise da complexidade de Lempel-Ziv aplicada a essas máquinas será apresentada e sua evolução discutida em função do tipo de autômato celular. Sugerir-se-á, a partir da dinâmica de ação das regras (algoritmos) sobre os estados, da rapidez de convergência de um autômato e da duração de sua condição estacionária, como a complexidade LZ76 poderá ser adequadamente aplicada e como seu valor numérico tem relação com os diagramas de classes descritos por Wolfram.

Nesse contexto se encaixa o presente trabalho, que objetiva o estudo das características dos elementos de tais constructos, dessas máquinas virtuais que podem ser levadas à realidade física. Serão discutidas propriedades das cadeias de dados como tamanho, base numérica e entropia, e qual é sua relação com a informação contida nessas sequências. É discorrida uma faceta geométrica dessas estruturas, mostrando-se como pode ser calculada sua dimensão fractal, e sua utilidade na quantificação do conteúdo de informação das cadeias de dados, com a intenção de se comparar a outros métodos de cálculo como o descrito [12, 21, 22].

O trabalho faz parte de um projeto geral de estudos computacionais em sólidos cristalinos complexos (do ponto de vista cristalográfico) para contribuir a transformar o paradigma atual de estudo e classificação da desordem estrutural usando teoria da informação. Pretende-se estabelecer algoritmos de identificação de regularidades e surgimento de padrões de comportamento. O objetivo futuro é classificar os politipos (acima) em função de sua complexidade e usar LZ76 como quantificador de processos de ordem-desordem nas transformações de fase de um politipo a outro a partir do cálculo de magnitudes entrópicas de sequências binárias.

PRIMEIRA PARTE – TEORIA

I – CADEIAS DE DADOS

I.1 - Cadeias de Dados

Cadeias são seqüências de objetos, caracteres, figuras ou dados de natureza diversa. Frequentemente utilizadas para descrever o ordenamento ou distribuição dos elementos de um conjunto, são úteis na tradução da informação externa para um computador, geralmente transliterando essa informação para uma base alfanumérica familiar ao núcleo de processamento, isto é, à linguagem utilizada pela máquina.

Linguagem, neste caso, não é o conjunto de comandos atribuídos no momento em que se escreve um programa, algoritmo, rotina, etc., mas o número de instruções permitidas ao circuito lógico do sistema, por exemplo, circuito "fechado" e "aberto", "ligado" e "desligado", o qual define a base numérica para a qual deve ser traduzida um conjunto qualquer de instruções (neste caso, base 2).

De modo geral, cadeias de dados podem ser justapostas sem um símbolo separador - ponto ("."), vírgula (","), ponto e vírgula (";"), etc., e podem ser representadas como:

A1B1A987CKP333ÇXT

10011101010001010110

Podem-se definir as mais diversas propriedades para as seqüências a fim de se estudá-las quantitativamente: tamanho, complexidade, entropia, base numérica, ordem lexicográfica, etc. Algumas delas são descritas a seguir.

I.2 - Tamanho da Cadeia (L, N)

O tamanho de uma seqüência é uma propriedade relacionada ao conceito de extensão, e poderia, com certa facilidade, indicar o comprimento espacial de determinado objeto, porém, em muitos aspectos, isso resultaria ineficiente se tratando de cadeias de dados, já que frequentemente são expressões virtuais de entidades abstratas.

Uma maneira mais útil seria definir o tamanho de uma cadeia (L, N) como o número total de constituintes (símbolos ou caracteres) desse objeto. Assim:

A1B1A987CKP333ÇXT

$L = 17$

10011101010001010110

$$N = 20$$



$$L = 8$$

L, N serão chamados de Tamanho da Cadeia de Dados.

I.3 - Base Numérica ou Cardinalidade do Alfabeto ($b, k, |\Sigma|$)

Define-se *base numérica* b , em um sistema de numeração, como o número que exprime a relação entre as diversas unidades sucessivas deste sistema. Em outras palavras, é a quantidade de símbolos distintos que compõem um sistema de numeração. Desta forma, 10, é a base numérica do sistema decimal, o qual é constituído por 10 símbolos distintos $\{0,1,2,3,4,5,6,7,8,9\}$; 2 é a base numérica do sistema binário $\{0,1\}$ e assim por diante.

Um *alfabeto* é o conjunto de símbolos diferentes da cadeia. Sua *cardinalidade* ($|\Sigma|$) é o número de elementos que o compõem. Dessa forma:

$$\Sigma = \{i, j\} \therefore |\Sigma| \cong b \tag{I.3.a}$$

Desde que o $\{i, j\}$ seja o conjunto dos elementos de Σ .

Para a maioria dos fins deste trabalho, o sistema binário é estabelecido como padrão e quase tudo que há de se fazer está escrito na base 2. Quando não for o caso, uma referência explícita será realizada.

I.4 - Entropia

Existe uma razão, uma função de estado termodinâmica, denominada Entropia S :

$$S = \int \frac{dQ}{T} \tag{I.4.a}$$

Cujo máximo local representa o estado de equilíbrio de um sistema. Esta é uma propriedade macroscópica e função extensiva dependente apenas da energia trocada no processo e da temperatura. É um tipo de princípio variacional equivalente a dizer que a condição de equilíbrio

será quando a energia interna U for mínima após as trocas de calor e trabalhos devidamente realizados ou não.

Quando se tenta explicar as propriedades termodinâmicas, que são macroscópicas, por métodos de mecânica clássica, aplica-se inevitavelmente procedimentos estatísticos no formalismo. Isto, em boa parte devido a incapacidade humana ou computacional de descrever o conjuntos de massas, velocidades e sentidos de todos os quase infinitos constituintes de um sistema termodinâmico - um fluido, por exemplo. Também, em parte, porque mesmo que fosse possível tal feito, a quantidade de dados seria também quase infinita, e muito pouco se poderia fazer com eles.

A vantagem do método estatístico é que trabalha com valores esperados, médias e desvios, que são de toda forma grandezas que representam o sistema como um todo, o que recai, novamente, nas grandezas termodinâmicas pretendidas.

Fundamentando-se nessa ideia, Ludwig Boltzmann aplicou os métodos estatísticos à mecânica de partículas e observou que aquela propriedade descrita por Gibbs e por Clausius, a entropia, poderia ser compreendida em termos da probabilidade de uma partícula (molécula, átomo) estar em determinado estado, e escreveu que:

$$S = k \cdot \ln \Omega, \quad (\text{I.4.b})$$

para:

$$\Omega = \frac{1}{p} \quad (\text{I.4.c})$$

ser o número de microestados acessíveis no sistema, ou seja, o inverso da probabilidade p de ocupação de um microestado por uma partícula, desde que, pela infinidade de partículas, todos os microestados são igualmente prováveis.

Isto é válido para um sistema que não troque energia ou constituintes com um reservatório, mas já representa microscopicamente aquela razão entre a energia em trânsito e a temperatura relacionada. Implica também que, maior o valor dessa entropia quer dizer maior número de estados acessíveis, portanto menor a probabilidade de uma partícula ser encontrada em determinado estado, portanto mais difícil de classificar esse sistema será.

A entropia estatística seria, então, uma medida do grau de dificuldade de classificar um sistema, de modo que quanto mais *complexamente* arranjando for, maior será sua entropia. Ainda, se o princípio diz que a condição de equilíbrio será quando essa grandeza for máxima localmente, então, com o passar do tempo, o sistema estudado passará de estágios mais facilmente classificáveis a estados menos facilmente classificáveis, e não retornará a menos que um trabalho externo o force a isso. O equilíbrio estará quando um tempo *muito longo* se passar, e nisto se instala a noção de ergodicidade.

Quando se diz que um sistema é composto de partículas, diz-se tacitamente que ele é discretizado. Logo, para estruturas na escala atômica também deve haver um equivalente estatístico à entropia estatística clássica. De fato, von Neumann conclui que:

$$S = -(k) \cdot \text{tr}(\rho \ln \rho) = -(k) \sum_i w_i \ln w_i, \quad (\text{I.4.d})$$

em que a constante de proporcionalidade fica em evidência pois não é regra a equação ser dimensional. Os termos w_i denotam as probabilidades segundo os postulados quânticos, visto que ρ é a matriz densidade do sistema.

Se estas probabilidades forem substituídas pelos pesos estatísticos adequados, a *forma* dessa entropia retornará à forma de Boltzmann. Por exemplo, se fizer a equivalência:

$$w_i \equiv \frac{1}{e^{\beta E_i}}, \quad (\text{I.4.e})$$

o peso de Boltzmann, então a entropia tornará ao *ensemble* canônico:

$$S = k \ln Z + \frac{\langle E \rangle}{T}, \quad (\text{I.4.f})$$

para Z a função de partição:

$$Z = \sum_i e^{-\beta E_i} \quad (\text{I.4.g})$$

Isto torna razoável a lei de formação do tipo (I.4.d). Com efeito, se a entropia expressa o quão difícil é de se classificar um sistema, discreto neste caso, então ela quer dizer que um

sistema com mais *informação* é menos facilmente classificável, *i. e.*, o zoológico de estados (conjuntos de posições, momento, energias, instantes de tempo...) é mais variado. Assim, a entropia, que já foi definida como uma função variacional para o equilíbrio, depois uma medição da complexidade de arranjo de um sistema, agora passa a ser uma *medida da informação* de um conjunto discreto.

Com isto em mente, Claude Shannon conclui que uma cadeia de dados de natureza diversa, ainda que seja virtual, deverá ser dotada de informação, já que, dentro do conjunto definido de elementos, cada um deles terá uma probabilidade de ocorrência [1]. Assim, para um conjunto discreto de dados na forma:

$$U = \{u_i, u_{i+1}, \dots, u_{N-1}, u_N\} \quad (\text{I.4.h})$$

Sua entropia será:

$$S, H(U) = -\sum_{i=1}^N p_i \log_b p_i \quad (\text{I.4.i})$$

em que p_i é probabilidade de se encontrar o elemento u_i na sequência. Se u_i for uma variável aleatória, então H será a incerteza média sobre esta variável, e, também, se a base b do logaritmo for 2, será a quantidade média de *bits* necessários para descrever a variável. Isto é fácil de se visualizar quando se imagina que a probabilidade de se encontrar $\log_b p_i$ é p_i , tal que:

$$S, H(U) = -\sum_{i=1}^N p_i \log_b p_i = -\langle \log_b p_i \rangle \quad (\text{I.4.j})$$

Como já dito, se a base logarítmica for binária, então a entropia (e a informação) será medida em *bits*, se for decimal será medida em *dits*, se for natural, em *nats*.

Para notar como isso funciona, imagine-se que uma variável aleatória pode ter como saída 32 valores distintos uniformemente. É natural concluir que uma cadeia de *5-bits* deverá ser necessária para descrever todas essas saídas:

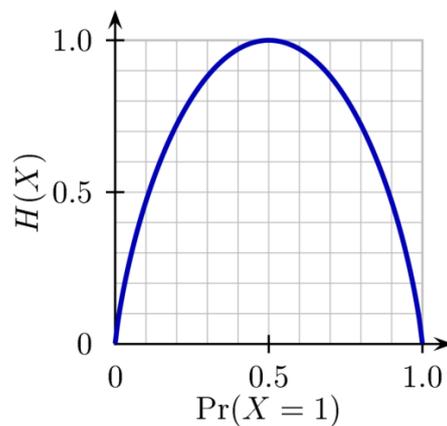
$$\{x_1, x_2, x_3, x_4, x_5\} : 5 - \text{bits}, \quad (\text{I.4.j})$$

pois 2^5 possibilidades varrem 32 valores distintos de saída. Para provar isto, considere-se a entropia:

$$H(x) = -\sum_{i=1}^{32} \frac{1}{32} \log_2 \frac{1}{32} = 5 \text{ bits}, \quad (\text{I.4.k})$$

de 32 valores distintos igualmente prováveis uniformemente distribuídos. O que exprime a quantidade média de informação para descrever a variável aleatória x .

Em um lançamento de moedas (cara ou coroa), cada lançamento não guarda memória do lançamento anterior, de forma que a probabilidade ocorrer cada uma das faces será sempre 0,5. Assim:



Fonte:
Wikipédia

Figura 3 – Entropia x Probabilidade: Curva de entropia para sequências geradas por lançamento de moedas

a entropia é máxima para o valor esperado de probabilidade.

É interessante notar que, apesar de a entropia de Shannon, da forma como foi descrita, expressar o conteúdo de informação de uma distribuição de dados qualquer, em teoria ela não exprime a complexidade do arranjo desses conjuntos, visto que depende apenas da probabilidade de ocorrência de cada elemento dentro da cadeia, ou seja, não expressa a informação contida no ordenamento do arranjo.

Como a entropia representa também a dificuldade de se classificar um sistema, precipitadamente chamada de *grau de desordem*, então um sistema visualmente mais simples e

cheio de padrões deveria exibir uma entropia menor que um arranjo de organização mais complexa.

De fato, se uma cadeia for dividida em N partes de tamanho l cada uma, tal que:

$$H^l(x) = -\sum_{i=1}^N p_i^l \log p_i^l, \quad (\text{I.4.1})$$

de acordo com a probabilidade p_i^l de se encontrar cada bloco dentro da fatorização, então essa entropia do tipo Shannon será sensível ao ordenamento dos constituintes ao longo da cadeia, e exprimirá a informação contida na organização em si da sequência. Esse tipo de entropia é chamada de *Entropia de Bloco (Block Entropy)* [1, 2]. Outra entropia de bloco será descrita na discussão dos resultados.

Para o estudo de sequências de objetos, faz-se útil a aplicação da Entropia de Bloco, sobretudo em qualquer descrição dinâmica de cadeias ou arranjos. Isto será feito para os Autômatos Celulares, cuja evolução será entendida como um sistema dinâmico (ou termodinâmico), e para a aproximação de seu comportamento ao mundo descrito pela física natural.

I.5 - Complexidade

Em termos linguísticos, a complexidade representa o quão difícil ou fácil é a apreensão de uma entidade, e por apreensão entenda-se compreensão, integração, incorporação, assimilação, fundamentalmente da essência de tal objeto, mas também de seu aspecto superficial.

Em Teoria da Informação essa dificuldade ou facilidade de apreensão é exprimida em termos do tamanho da descrição daquela entidade, ou seja, o quanto de recursos descritivos é utilizado para expressar aquilo que se deseja, seja uma cadeia ou conjunto de dados. Neste campo insere-se uma chamada Complexidade de Kolmogorov.

Seja uma máquina abstrata provida de blocos de memória (alocações de estado) e capaz de realizar leitura, processamento e gravação de um número finito de estados baseado em um alfabeto de tamanho finito; esta máquina levará um estado inicial em blocos de memória a um estado final segundo um conjunto de regras pré-escritas (programa). Esta máquina é chamada Máquina Universal de Turing, por conta de seu inventor Alan Turing [3].

Uma Máquina de Turing é uma modelagem abstrata de computador, que é ergódico neste contexto, e no qual o tamanho do menor programa rodando nela que, dada uma entrada, descreve uma saída, é chamado de Complexidade de Kolmogorov (de Kolmogorov-Chaitin, entropia algorítmica, etc.) [1,4,5,6]. Assim, a Complexidade de Kolmogorov fornece a quantidade de recursos de computabilidade abstrata necessários para especificar um objeto.

A teoria estabelece que uma fonte ergódica de sequências aleatórias será uma fonte com máxima complexidades algorítmica, portanto, com máxima taxa de entropia. Entretanto, um algoritmo descrito no riquíssimo trabalho de 1976 de Lempel e Ziv fornece o caminho para mostrar que há certas sutilezas nessa análise [6].

Dada uma sequência u na forma:

$$u = u_1 u_2 u_3 \dots u_N, \quad (\text{I.5.a})$$

existe uma operação sobre esta sequência tal que:

$$\begin{aligned} \forall u(i, j) \in u(1, N) \exists u(l, k) \neq u(i, j) \ni u(l, k) \in u(1, N) \\ \wedge \{i, j\} \cup \{l, k\} \subseteq \{1, N\} \end{aligned} \quad (\text{I.5.b})$$

Esta operação define uma fatoração, que nada mais é do que subdividir a grande sequência u em partes menores segundo uma regra pré-definida. Seja agora um operador "queda", identificado com a letra grega Π , da seguinte forma:

$$u(i, j)\Pi = u(i, j-1)$$

e:

$$u(i, j)\Pi^k = u(i, j-k)$$

A fatoração LZ76² $E(u)$ da sequência u :

² Trata-se do algoritmo descrito em 1976 por Lempel e Ziv. Há diferentes fatorizações de Lempel-Ziv [7]

$$E(u) = u(1, h_1)u(h_1 + 1, h_2) \dots u(h_{m-1} + 1, N) \quad (\text{I.5.c})$$

é feita em m fatores $u(h_{k-1}+1, h_k)$ tais que:

- (1) $u(h_{k-1} + 1, h_k) \pi \subset u(1, h_k) \Pi^2$
- (2) $u(h_{k-1} + 1, h_k) \not\subset u(1, h_k)$ exceto para o último fator $u(h_{m-1} + 1, N)$

A primeira condição define $E(u)$ como uma *história* de u , e a segunda define cada história como sendo *exaustiva*. Isto quer dizer que cada partição é única, e sempre que se for extrair a próxima deve-se verificar se aquele fator não ocorreu anteriormente, e isso é feito até que todos os elementos da cadeia se esgotem. Ou seja, cada subcadeia de u nunca é igual à próxima, mas a próxima deve conter qualquer elemento dentre as cadeias anteriores, e é interrompida quando não contém mais. Exemplificando, a cadeia será fatorizada da seguinte maneira:

$$1000110100011101 \rightarrow 1.0.001.101.000111.01$$

Na qual pode-se verificar que os novos elementos são extraídos na medida em vão fazendo uma subcadeia nova na sequência original (até aquela posição):

$$ABCBABCBC CABABBC \rightarrow (A)(B)(C)(BA)(BCBC)(CA)(BABB)(C)$$

Dada, então, uma fatoração dessa natureza, pode-se definir a *complexidade C de uma sequência como sendo o número desses fatores em cada cadeia*:

$$1000110100011101 \rightarrow 1.0.00.11.010.00111.01 \quad C = 7$$

$$ABCBABCBC CABABBC \rightarrow (A)(B)(C)(BA)(BCBC)(CA)(BABB)(C) \quad C = 8$$

Como LZ76 é uma descrição algorítmica de complexidade de sequência, quer dizer que uma fonte ergódica de cadeias aleatórias emitirá, em média, sequências com máximas complexidades, já que LZ76 fornece uma descrição algorítmica da aleatoriedade em sequências. Contudo, isso não quer dizer que sequências com máximas complexidades LZ76 terão necessariamente máximas complexidade de Kolmogorov (por serem aleatórias).

De fato, Estevez-Rams e colaboradores [2] construíram um algoritmo que descreve sequências de máximas complexidades de Lempel-Ziv. Trata-se de um algoritmo LZ76 reverso, o qual adiciona subcadeias ainda inexistentes a uma grande cadeia, tornando-a de máxima complexidade.

Para exemplificar o mecanismo, considere um alfabeto $\Sigma = \{a,b,c\}$. Primeiramente se constroi uma pequena sequência $E(MLZs)$ esgotando-se as combinações de um único elemento desse alfabeto:

$$E(MLZs) = (a)(b)(c),$$

depois, faz-se o mesmo com as combinações de dois elementos daquele espaço da base:

$$\begin{array}{l} aa \quad ab \quad ac \\ ba \quad bb \quad bc, \\ ca \quad cb \quad cc \end{array}$$

então:

$$E(MLZs) = (a)(b)(c)(aa)(ac)(ba)(bb)(cc)$$

Note-se que os elementos riscados:

$$\begin{array}{l} aa \quad \text{\textit{ab}} \quad ac \\ ba \quad bb \quad \text{\textit{bc}} \\ \text{\textit{ca}} \quad \text{\textit{cb}} \quad cc \end{array}$$

foram excluídos, pois já teriam aparecido em alguma posição anterior na sequência formada, mantendo a característica de *história exaustiva* do algoritmo. E assim será feito sucessivamente com as combinações de três, quatro, cinco, seis... elementos até a ordem que se deseja para a construção da sequência de máxima complexidade LZ76.

Essa regra é formalizada da seguinte maneira:

1. Considere $l = 1$;
2. Construa o conjunto Λ^l de todas as cadeias de comprimento l ;
3. (Passo teste) Considere em ordem lexicográfica um elemento $\lambda \in \Lambda^l$, se $\lambda \not\subseteq u\lambda\pi$, anexe λ a u , ignore outro caso;
4. Repita o passo 3 até que todos os elementos de Λ^l tenham sido considerados ou, $|u| \geq N$;
5. Se $|u| < N$, faça $l = l + 1$ e retorne ao passo 2;
6. Se $|u| > N$, interrompa u no comprimento N do final;
7. Pare.

Ergo, devido à maneira determinística e reduzida com a qual é possível construir seqüências de máxima complexidade de Lempel-Ziv, mostra-se que máximas complexidades de Lempel-Ziv nem sempre estão para máximas complexidades de Kolmogorov. Entretanto, tais seqüências não são as únicas de máxima complexidade dentre as de mesmo tamanho L . Há um conjunto de cadeias não descritas pelo algoritmo reverso LZ76 que também exibem características de máxima complexidade gerada pelo algoritmo e que podem, eventualmente, estar para máximas complexidades de Kolmogorov.

A fatoração de Lempel-Ziv fornece uma medida algorítmica de complexidade de seqüências de dados, sendo útil na descrição de compressão de dados, em termos práticos. Mas há outras maneiras de se fatorizar cadeias de modo a retirar outros tipos de informação das mesmas.

Um algoritmo conhecido com *Run Length Encoding* (RLE) é utilizado para se fatorizar seqüências segundo a semelhança dos elementos vizinhos [8]. Ele separa os constituintes idênticos e próximos em uma cadeia em blocos distintos. Uma seqüência como:

11101001100111010

seria fatorizada como:

$$11101001100111010 \rightarrow (111)(0)(1)(00)(11)(00)(111)(0)(1)(0)$$

$$\{1,1,1,0,1,0,0,1,1,0,0,1,1,0,1,0\} \rightarrow \{111,0,1,00,11,00,111,0,1,0\}$$

E poder-se-ia dizer, também, que há uma complexidade associada a essa fatoração, e que ela seria, diga-se, o comprimento da nova sequência para qual levou a regra (*o número de fatores*) – no exemplo acima seria 10.

Por meio de RLE, uma cadeia de caracteres construída com um alfabeto qualquer é levada a outra cadeia escrita em outra base. Esta base seria o número de regiões distintas entre si que podem expressar essa nova cadeia. Equivalente fazer:

$$\{1,1,1,0,1,0,0,1,1,0,0,1,1,1,0,1,0\} \rightarrow \{111,0,1,00,11,00,11,0,1,0\} \equiv \{A, B, C, D, E, D, E, B, C, B\}$$

$$b = 5; \Sigma = \{A, B, C, D, E\}$$

Uma nova cadeia em uma nova base com o mesmo conjunto de dados.

Pode-se, ainda, construir uma entropia de bloco específica, de acordo com a probabilidade de se encontrar cada conjunto de caracteres semelhantes. A esses conjunto de caracteres semelhantes alguém poderia chamar, por exemplo, *domínios de periodicidade*, devido ao fato de sua repetição ou não ao longo das sequências determinar o caráter periódico da mesma.

Uma entropia escrita dessa maneira, nessa nova base, como se verá na discussão dos resultados do trabalho, é sensível ao ordenamento dos constituintes, carregando, então, informação a respeito da organização da sequência. Os definidos *domínios* dessa *periodicidade* não definida - não se desejou aqui definir uma propriedade como Periodicidade, apesar de seu sentido e uso ficarem implícitos doravante - através da codificação RLE são úteis também na determinação da estrutura geométrica das cadeias de dados, mais propriamente de sua *dimensionalidade*, e isto será discutido agora.

I.6 - Dimensão Fractal

O conceito de dimensionalidade está intimamente ligado à geometria de uma entidade qualquer, seja ela objetivamente topográfica ou ligada à forma em si, ou também abstrata algébrica exprimindo o grau de homogeneidade de relações matemáticas.

Tipicamente, quando se pensa na dimensão de algo o que quer que seja, intuitivamente se pensa em um valor inteiro (uma, duas, três dimensões) visto que as coisas no mundo ou tem comprimento, ou comprimento e altura, ou comprimento, altura e profundidade, etc.; não fazendo parte da abstração qualquer propriedade que não seja discreta. O número de vezes que se realiza

uma operação (a derivada, por exemplo) é uma, duas ou três vezes, perdendo o sentido sentenças como "2,31 derivadas de uma função".

Neste sentido, o matemático Felix Hausdorff pensou em uma quantificação dessas grandezas relacionadas à dimensão, sejam geométricas ou algébricas, e fez isto imaginando o preenchimento do espaço que contém uma estrutura qualquer. Seja a figura:

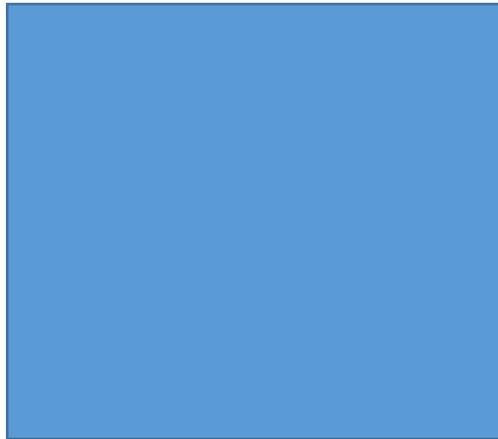


Figura 4 – Quadrado: construção plana ortogonal em espaço 2D.

É um quadrado e, sabidamente, é uma figura plana, portanto de 2 dimensões. Agora divide-se esse quadrado em partes idênticas entre e si e na mesma *proporção* da grande figura.

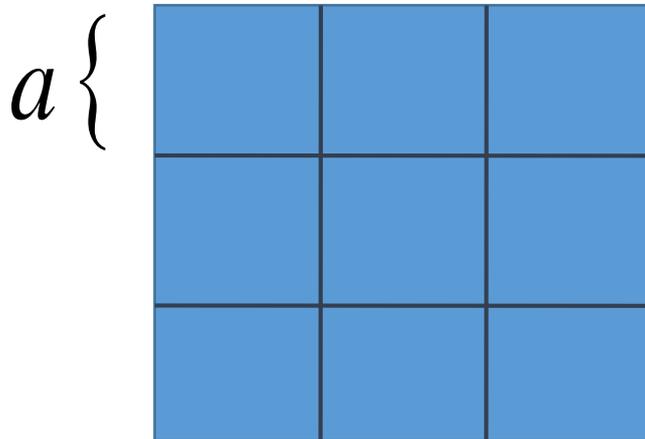


Figura 5 – Subdivisão do quadrado em partes autossimilares ao todo e idênticas entre si.

cada parte com lado a de mesmo tamanho.

O número $N(a)$ de subdivisões feitas na figura é contado da seguinte maneira:

$$N(a) = \left(\frac{3a}{a}\right)^2 = 9 \quad (\text{I.6.a})$$

Em que o numerador entre parênteses exprime o lado da grande figura nas medidas das pequenas, estas no denominador. Caso fosse um cubo ou hexaedro regular repartido da mesma maneira, a operação ficaria:

$$N(a) = \left(\frac{3a}{a}\right)^3 = 27 \quad (\text{I.6.b})$$

Totalizando 27 entidades *autossimilares* à original.

Generalizando-se a ideia, pode-se concluir facilmente que o número de subdivisões quaisquer feitas em *escala de autossimilaridade* de uma forma é:

$$N(a) = \left(\frac{l}{a}\right)^D, \quad (\text{I.6.c})$$

em que D representação a dimensão da forma, l é o tamanho. Invertendo a equação:

$$\log[N(a)] = \log\left[\left(\frac{l}{a}\right)^D\right]$$

∴

$$D = \frac{\log[N(a)]}{\log\left(\frac{l}{a}\right)} \quad (\text{I.6.d})$$

D é chamada de Dimensão de Hausdorff [9].

A Dimensão de Hausdorff mede a dimensionalidade de uma figura com base no preenchimento do espaço no qual ela é contida. Tradicionalmente figuras conhecidas como formas regulares ou irregulares, tetraedro, octógono, tesseracto têm suas dimensões em números

inteiros. Contudo, existe uma classe de estruturas - difícil dizer que são formas, ainda que sejam formações - nas quais, quando se faz o cálculo de D , obtem-se valores não inteiros, *i.e.*, fracionais, dízimas não periódicas, etc. A essas estruturas chama-se de *Fractais*, estruturas com dimensões *fracionadas* ou não inteiras.

Um exemplo clássico de fractal e por meio do qual se fazem diversas demonstrações em geometria fractal é o Conjunto de Cantor. Imagine-se a construção:



Figura 6 – Ilustração simples da geometria de um Conjunto de Cantor. Cada passo da construção gera autossimilaridade ao todo, em menor escala (Fonte: Wikipédia)

A parte inteira (topo) é subdividida em terços, então se elimina o terço médio. Repete-se a operação com os terços restantes, e assim por diante até o infinito. A cada passo, as figuras restantes são idênticas às anteriores, mas em escala menor. Isto define uma *escala de auto-similaridade*. A operação de particionamento deve preservar a geometria da estrutura, sua dimensão de Hausdorff, tornando-se, assim, um invariante por dimensionalidade.

Se no segundo passo os terços ficam com tamanho 1 ($a = 1$) cada, e existem 2 subdivisões ($N(a) = 2$) de um total de 3 regiões ($l = 3$), então D fica:

$$D = \frac{\log[N(a)]}{\log\left(\frac{l}{a}\right)}$$

$$D = \frac{\log[2]}{\log\left(\frac{3}{1}\right)}$$

$$\therefore$$

$$D = \frac{\log 2}{\log 3} \tag{I.6.e}$$

Se no terceiro passo as partes ficam com tamanho 1 cada ($a = 1$), e existem 4 subdivisões ($N(a) = 4$) de um total de 9 regiões ($l = 9$), então:

$$\frac{\log [N(a)]}{\log \left(\frac{l}{a}\right)}$$

$$\therefore$$

$$D = \frac{\log 4}{\log 9} = \frac{\log 2}{\log 3} \tag{I.6.f}$$

E assim por diante, preservando, de fato, a dimensionalidade, a qual não é um número inteiro neste caso. Esse é um Conjunto de Cantor simétrico, porque as subdivisões são feitas de maneira idêntica. Mas existem conjuntos assimétricos, e estes apresentarão outros valores de D .

Com esse exemplo fica mais claro o entendimento de que a dimensão tem algo a ver com: *como a figura (ou estrutura) preenche o espaço no qual ela é contida.*

A dimensão fractal de uma formação pode ser uma grandeza calculável caso os parâmetros que a constituem tenham um padrão matemático de correspondência em si, como os lados dos pequenos quadrados relativamente ao lado do grande (figura 4 acima); dessa forma, a subdivisão é feita em uma escala coerente. Mas no caso de uma figura genérica, como a faixa litorânea de um país, tem de se utilizar parâmetros de medição diferentes caso se queira saber sua dimensionalidade.

Isto quer dizer que o cálculo precisa levar em conta um preenchimento estatístico do espaço que contém a formação, já que se se estabelecer uma grade de escala sobre uma figura qualquer (Figura 6), verifica-se que o espaço que a contém pode não estar preenchido em toda parte do mesmo modo. Neste caso, pode-se utilizar o método de *Box Counting* [10, 11, 12], o qual, na situação da Figura 6, conta o número $N(s)$ de *caixas* preenchidas pela figura na escala s , isto é, o número de vezes que o lado da imagem será dividido, fazendo com que:

$$D = \frac{\log N(s)}{\log \left(\frac{1}{s}\right)} \tag{I.6.g}$$

E quanto mais caixas houverem, e quanto menores elas forem, mais precisa será a medida e mais próximo do valor real será D .

Box Counting é um método famoso e muito utilizado, e também deixa evidente que não somente figuras com auto-similaridade bem definida podem ter dimensão fractal, mesmo que muitas delas não sejam de fato fractais. É uma maneira implicitamente estatística de se medir D para uma formação genérica, porém, leva em consideração que tais formações são constituídas da mesma *matéria*, ou da mesma essência, ou do mesmo tipo de conteúdo, como figuras com bordas sempre pretas e de mesma densidade, sempre com peso estatístico idêntico à vizinha etc.



Fonte:
<<http://teclustelecom.com.br/>>

Figura 7 - Caixas preenchendo uma figura genérica com objetivo de destacar o espaço preenchido pela forma.

Pensando neste tipo de situação foi que surgiu o conceito de *Multifractal*, ou *Multifractal Geométrico*, o que, a partir da interpretação do trabalho de Tél, Fülöp e Vicsek [13], trata-se de um fractal constituído por matérias de diferentes naturezas, ou estruturas que têm seu crescimento distinto ao longo de sua composição.

De todo modo, uma maneira estatística de interpretar essa formação é necessária. Para tanto imagine-se que aquela caixa de *clusters* sobre a Figura 6 tenha sua grade composta, cada unidade, com uma *matéria* de *massa* $M(a)$, e que esta massa queira dizer, por exemplo, o número de partículas dentro do *cluster* de tamanho a . Se M_0 é a *massa* (escrita em itálico para que se subentenda a conotação não objetiva) total da estrutura de lado L , a conclusão dos autores foi que sua dimensão é da forma:

$$D = \frac{\log \left\langle \left(\frac{M(a)}{M_0} \right) \right\rangle}{\log \left(\frac{a}{L} \right)} \quad (\text{I.6.h})$$

A média no argumento do numerador é um valor esperado segundo a probabilidade de encontrar aquela razão na estrutura observada. De forma generalizada:

$$D^q = \frac{\log \left\langle \left(\frac{M(a)}{M_0} \right)^{q-1} \right\rangle}{\log \left(\frac{a}{L} \right)} \frac{1}{q-1} \quad (\text{I.6.i})$$

Para q ordens de cálculo. Aqui se utilizará $q = 2$ para concordar com a usual dimensão de Hausdorff. Note-se que existe algum grau de semelhança com a q -entropia de Tsallis. Entretanto, não é objetivo deste trabalho discorrer sobre o assunto.

Essa é uma maneira muito útil de obtenção da dimensão fractal de uma estrutura constituída de várias entidades distintas. E funciona. Porém, a premissa dos autores da ref. [13] é que o tamanho a do *cluster* é fixo, o que não deixa de ser uma verdade, contudo a experiência mostrou que, para que fique totalmente funcional, a dedução deve levar em conta que o *tamanho médio* dos *clusters* deve ser constante. E isto completa a maneira estatística da descrição, de modo que:

$$D = \frac{\log \left\langle \left(\frac{M(a)}{M_0} \right) \right\rangle}{\log \left(\frac{\langle a \rangle}{L} \right)} \quad (\text{I.6.j})$$

Quando todos os *clusters* têm o mesmo tamanho, então (I.6.j) retorna a (I.6.h). Apesar de ser uma constatação empírica, não deve ser difícil a quem deseja deduzir a solução com a premissa do tamanho médio dos *clusters*.

Essa consideração é fundamental para o que pretende ser feito com as cadeias de dados. Por essa avaliação, cadeias de dados podem ser consideradas estruturas multifractais, desde que

haja uma análise estatística a respeito da distribuição dos seus constituintes, já que possuem geometria, pois unidades são dispostas com organização em algum nível espacial (abstrato ou não) e pelo fato desses elementos poderem ter natureza diversificada.

Destarte, deixe-se representar um Conjunto de Cantor como uma sequência numérica:

11111111111111111111111111111111,

e comece-se o processo de eliminação dos terços médios:

11111111111111111111111111111111

111111111000000000011111111111

1110001110000000000111000111

1010001010000000000101000101

Em que a sequência original tem tamanho $L = 27$, e cada caractere representa uma unidade de tamanho 1. Assumindo que "0" representa "espaços vazios" neste caso, cada conjunto de "1" justapostos representa um cluster de tamanho igual a quantidade deles. Assim, no primeiro passo de eliminação, além de L , tem-se $M(a) = 18$ (cada bloco com 9 unidades "1"), $M_0 = 18$ (tamanho total dos blocos "1"), (2 blocos com nove elementos "1" dividido pela quantidade deles quantidade deles, 2):

$$D = \frac{\log \left\langle \left(\frac{M(a)}{M_0} \right) \right\rangle}{\log \left(\frac{\langle a \rangle}{L} \right)}$$

$$D = \frac{\log \left[\left(\frac{9}{18} + \frac{9}{18} \right) / 2 \right]}{\log \left[\left(\frac{(9+9)/2}{27} \right) \right]}$$

$$D = \frac{\log \left[\frac{1}{2} \right]}{\log \left[\frac{1}{3} \right]} = \frac{\log 2}{\log 3} \quad \text{(I.6.k)}$$

De acordo com (I.6.e). Para o segundo passo de eliminação, tem-se $M(a) = 3$ (quatro blocos com 3 unidades “1” cada), $M_0 = 12$ (tamanho total dos blocos “1”), (4 blocos com três elementos “1” dividido pela quantidade deles, 4):

$$D = \frac{\log \left\langle \left(\frac{M(a)}{M_0} \right) \right\rangle}{\log \left(\frac{\langle a \rangle}{L} \right)}$$

$$D = \frac{\log \left[\left(\frac{3}{12} + \frac{3}{12} + \frac{3}{12} + \frac{3}{12} \right) / 4 \right]}{\log \left[\left(\frac{(3+3+3+3)/4}{27} \right) \right]}$$

$$D = \frac{\log \left[\frac{1}{4} \right]}{\log \left[\frac{1}{9} \right]} = \frac{\log 2}{\log 3}$$

(I.6.l)

E, para o terceiro passo, $M(a) = 1$ (oito blocos com 1 unidade “1” cada), $M_0 = 8$ (quantidade total dos blocos “1”), (8 blocos com 1 elemento “1” dividido pela quantidade deles, 8):

$$D = \frac{\log \left\langle \left(\frac{M(a)}{M_0} \right) \right\rangle}{\log \left(\frac{\langle a \rangle}{L} \right)}$$

$$D = \frac{\log \left[\left(\frac{1}{8} + \frac{1}{8} \right) / 8 \right]}{\log \left[\left(\frac{(1+1+1+1+1+1+1+1)/8}{27} \right) \right]}$$

$$D = \frac{\log \left[\frac{1}{8} \right]}{\log \left[\frac{1}{27} \right]} = \frac{\log 2}{\log 3}$$

(I.6.m)

Todas concordando com (I.6.e) e sustentado a dimensão fractal. É evidente que não se pode fazer a subdivisão em infinitos passos, visto que a unidade fundamental tem tamanho finito. Contudo, isso fornece uma boa descrição do comportamento fractal médio de uma sequência de dados. É importante ressaltar que "0" neste caso representa um "espaço vazio", portanto com peso estatístico igual a 0. Porém, ao gosto e à necessidade de quem faz o cálculo, cada caractere distinto pode assumir uma importância diferente, e isto é o que entra na conta dos valores médios.

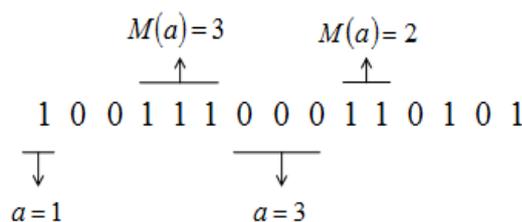
Verifica-se a importância de se considerar uma cadeia de dados como sendo um multifractal, pois os elementos que a constituem podem ter as mais distintas naturezas, assumindo importâncias diversas ao longo da cadeia. E caso deseje-se fazer uma consideração geométrica da mesma, é preciso levar em conta essa importância.

No caso dos "0" terem o mesmo peso dos "1", a dimensão fractal de todas aquelas sequências seria igual à unidade, porque estatisticamente representariam linhas, ou seja, estruturas com uma dimensão.

Neste exemplo, foi utilizado um conjunto conhecido com uma Dimensão de Hausdorff também conhecida, e nele, todos os *clusters* (os "1" de uma mesma vizinhança) têm o mesmo tamanho dentro da sequência. Agora, em uma situação em que os constituintes não estão distribuídos com uma organização tão bem definida:

10011000110101

A análise não é tão diferente:



Apenas levando-se em conta o tamanho e quantidade existente dentro de uma mesma vizinhança, através do particionamento RLE. $M(a) = (1 + 3 + 2 + 1 + 1)$ – cinco blocos com 1, (1, 3, 2, 1 e 1 elementos “1”, respectivamente), $M_0 = 8$ (quantidade de elementos “1”), (5 blocos com 1, 3, 2, 1, 1 elementos “1”, respectivamente, dividido pela quantidade deles, 5). Neste caso:

$$D = \frac{\log \left\langle \left(\frac{M(a)}{M_0} \right) \right\rangle}{\log \left(\frac{\langle a \rangle}{L} \right)}$$

$$D = \frac{\log \left[\left(\frac{1}{8} + \frac{3}{8} + \frac{2}{8} + \frac{1}{8} + \frac{1}{8} \right) / 5 \right]}{\log \left[\left(\frac{(1+2+1+1)/5}{15} \right) \right]}$$

$$D = \frac{\log \left[\frac{1}{5} \right]}{\log \left[\frac{1}{15} \right]} = \frac{\log 5}{\log 15} \quad (\text{I.6.n})$$

Os mesmos caracteres, mas com outra organização:

111111110000000

A dimensão:

$$D = \frac{\log \left\langle \left(\frac{M(a)}{M_0} \right) \right\rangle}{\log \left(\frac{\langle a \rangle}{L} \right)}$$

$$D = \frac{\log \left[\left(\frac{8}{8} \right) / 1 \right]}{\log \left[\left(\frac{(8)/1}{15} \right) \right]}$$

$$D = \frac{\log [1]}{\log \left[\frac{8}{15} \right]} = 0 \quad (\text{I.6.o})$$

Um resultado curioso, mas que deixa clara a interpretação do que vem a ser a dimensão fractal. É evidente que o conjunto de "1" todos lado a lado naquela sequência, e com espaço vazio em redor ("0"), deveria configurar uma linha, portanto com $D = 1$, contudo, deixar explicitamente o

conjunto vazio na fronteira da estrutura faz com que ela não preencha toda região ao longo do espaço que a contém, tornando-a um ponto em relação a ele. Isto sugere o seguinte, só faz sentido se falar em preenchimento do espaço dentro da região limítrofe da estrutura, precisamente com fronteiras preenchidas de matéria, visto que em relação à uma região de inexistência exterior, qualquer formação interna não é dotada de extensão.

É lógico que esta é uma análise demasiado rigorosa da situação, e eventualmente será necessário considerar as bordas nulas no cálculo de D . Entretanto, isso aponta para a evidência de que *clusters* isolados pela codificação RLE, independente da matéria, são regiões pontuais, portanto com dimensão não finita.

Outros exemplos sequências:

1111111111

$$D = \frac{\log \langle [M(a)/M_0] \rangle}{\log \langle (a)/L \rangle} = \frac{\log(10/10)}{\log(10/L)} = 1$$

AAA000AAA

$$D = \frac{\log \langle [M(a)/M_0] \rangle}{\log \langle (a)/L \rangle} = \frac{\log(3/6)}{\log(3/9)} = \frac{\log 2}{\log 3} \approx 0,63$$

101000110011

$$D = \frac{\log \langle [M(a)/M_0] \rangle}{\log \langle (a)/L \rangle} = \frac{\log(1/4)}{\log((3/2)/12)} = \frac{\log 4}{\log 8} = 0,666\dots$$

Todas "0" com peso nulo... caso o peso fosse idêntico aos demais caracteres, todas as cadeias apresentariam $D = 1$.

II – AUTÔMATOS CELULARES

II.1 - Autômatos Celulares

Autômatos celulares, autômatos de estado finito, arranjos iterativos, espaços celulares etc, são máquinas de estado finito. Por *máquinas* entenda-se que são composições abstratas capazes de levar um estado de memória a outro através de um programa ou regra; *finito* quer dizer que cada alocação de memória deve receber uma quantidade de valores diferente de zero. Os estados que a máquina pode apresentar são, por vezes, delimitados por células, daí o adjetivo *celular*.

Seja uma regra ou conjunto de regras que transformam as células de uma cadeia em células de outra, a cada passo (ou instante de tempo), a cadeia anterior será atualizada segundo a orientação do programa, e assim sucessivamente até o instante (ou passo) em que se deseje interromper a evolução. Por exemplo, uma composição genérica poderia fazer:

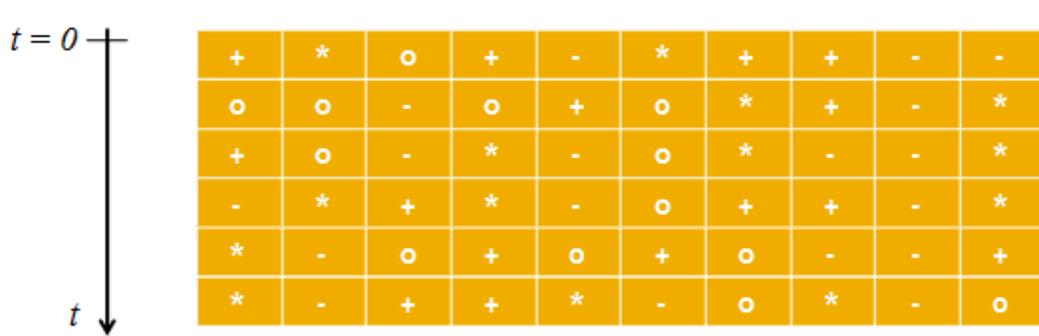


Figura 8 - Autômato Genérico. Exemplificação de uma máquina que evolui. A organização dos símbolos e a regra de atualização não necessariamente existem.

Levando uma cadeia em um estado inicial no instante $t = 0$, a outro estado em $t = 1$, tomando este estado em $t = 1$ e o levando a outro estado em $t = 2$, e assim por diante. O autômato celular em si pode ser definido como a regra propriamente dita (o conjunto de instruções para a atualização das células) bem como a estrutura formada pela evolução temporal daquela regra aplicada à condição inicial.

Essas máquinas foram primeiramente descobertas em 1940 por von Neumann [14], e são meramente uma grade de células que confinam estados, estes em uma quantidade finita de possibilidades.

A quantidade b de estados que uma célula pode apresentar depende do espaço de estados da base Σ :

$$\Sigma = \{i, j\} \ni |\Sigma| = b \quad (\text{II.1.a})$$

sobre o qual o sistema é fundamentado. Essa célula pode pertencer a uma vizinha de tamanho l , ou seja, o conjunto de células próximas que interagirão com ela. Neste comprimento l poderão existir b^l vizinhanças distintas. O número ξ de configuração possível dessa vizinhanças será, portanto:

$$\xi = b^{(b^l)} \quad (\text{II.1.b})$$

Cada uma dessas formas em que uma vizinhança pode se organizar define uma regra de atualização diferente. Em outras palavras, podem haver tantas regras quanto vizinhanças possíveis, embora nem sempre a forma como agirão sobre as cadeias difiram entre si; *atualização* é a mudança, ou não, do estado de uma célula para um instante posterior.

Se a grade de células se dispor ao longo de uma linha:



Figura 9 - Autômato Unidimensional. Visualização de uma máquina circular estendida sobre uma linha. As extremidades são regiões vizinhas.

O autômato será unidimensional, e as extremidades da cadeia são tratadas como sendo vizinhas, ou seja, a máquina se fecha em um anel. Neste caso, a passagem do tempo é representada como na Figura 10, com regiões inferiores representando instantes posteriores no tempo, e a condição inicial fixa em $t = 0$.

Na ocasião em que a grade se dispor ao longo de um plano, diz-se que o autômato é bidimensional. Aqui a passagem do tempo pode ser representada como um terceiro eixo coordenado, ou simplesmente como uma animação de evolução temporal de fato. Exemplos são os chamados Planadores e os Jogos da Vida de Conway [15]. E as dimensões seguem o padrão até quando se deseje.

A evolução de um autômato depende de uma regra, e essa regra é determinística, ou seja, existe um algoritmo que a descreve. Para as regras, em geral, o estado de uma célula no instante $t + 1$ depende do estado da vizinhança no instante t . Quer dizer, sendo $f(t)$ o estado de uma célula no instante t :

$$Uf(t) = f(t+1), \quad (\text{II.1.c})$$

e U será o operador de evolução de um passo.

Quando um autômato é submetido a uma regra que depende apenas dos valores totais ou médios das células da vizinhança no instante anterior, diz-se que esse autômato (e essa regra) são *totalísticos*. Ainda, se essa regra depende apenas dos primeiros vizinhos de uma célula, e é aplicada sobre um autômato unidimensional no espaço de estados da base $\Sigma = \{0,1\}$, diz-se que este é um *Autômato Celular Elementar* (Wolfram, 1983) [16].

Para um autômato celular elementar, é usual representar as células como:

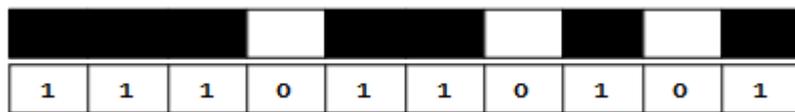


Figura 10 - Autômato Celular Elementar. Máquina unidimensional construída em base 2, na qual a atualização de uma célula depende unicamente dos primeiros vizinhos da mesma.

Em que estados "1" estão para células pretas e estados "0" estão para células brancas, ou vice-versa. Se o valor de uma célula depende do estado, e não dos valores totais ou médios das células da vizinhança em uma atualização, então esse autômato será chamado de *totalístico exterior*:



Figura 11 - Autômato totalístico exterior. A atualização não depende de uma operação matemática sobre a vizinhança, apenas dos estados em si das células vizinhas, ditas exteriores: b se atualização para d quando a e c forem suas vizinhas.

Neste caso, a célula com b foi atualizada para d dependendo apenas dos valores a e c na ordem especificada.

Para um autômato celular elementar, ocorrendo que:

$$\Sigma = \{0,1\} \ni |\Sigma| = 2, \quad (\text{II.1.d})$$

o número possível de configurações para dois vizinhos é, de acordo com (II.1.d):

$$\xi = b^{(b^l)}$$

∴

$$\xi = 2^{(2^3)} = 256 \quad (\text{II.1.e})$$

Isto fornece, portanto, um total de 256 regras totalísticas disponíveis para um autômato celular elementar.

As regras totalísticas são aplicadas de uma maneira tal que não variam no tempo e são as mesmas para todas as células da rede: a cada passo de tempo a mesma regra invariante é aplicada a todas as células da cadeia. No caso de isso não ser obedecido, o autômato é chamado de assíncrono ou estocástico, devido a característica não homogênea de como e evolução é realizada.

Pelo fato de para os autômatos elementares a base numérica utilizada é 2, costuma-se representar cada uma das 256 regras possíveis como um número binário:

$$x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 \quad x_i \in [0,1]$$

De modo a cada uma das possíveis combinações de algoritmos estar relacionada a cada uma das possíveis combinações de dígitos binários. Contudo, é mais usual representar a regra binária, ainda, em base decimal. Por exemplo, a regra:

$$11111110_2 \equiv 254_{10}$$

Seria essa equivalência na base 10. É importante lembrar que cada regra totalística de autômatos é uma operação de médias, densidades e somas com os elementos que constituem uma vizinhança. Um interessante trabalho que discorre a respeito dos algoritmos que formam o conjunto de regras totalísticas está em [17], que escreveu as regras para o software MATLAB.

II.2 - As Classes de Wolfram

Stephen Wolfram foi um dos grandes estudiosos dos autômatos de estado finito, e encontrou uma maneira de classificá-lo de acordo com os padrões exibidos no arranjo temporal de células. De acordo com a trivialidade ou complexidade dos padrões das figuras, os autômatos, segundo Wolfram, podem ser separados nas seguintes classes:

- Classe 1: Triviais ou Nulos;
- Classe 2: Periódicos;
- Classe 3: Complexos;
- Classe 4: Caóticos.

E segue agora uma descrição a respeito dessas classes de autômatos:

i) Classe 1: Triviais ou Nulos

Autômatos Triviais ou nulos são aqueles os quais, dada uma condição inicial qualquer, a regra a converge sempre para padrões uniformes, transformando a cadeia inicial em uma cadeia com todos os caracteres idênticos:



Figura 12 - Regra 32 (trivial).
Convergência a um estado único
para todas as células.



Figura 13 - Regra 234 (trivial).
Convergência a um estado único
para todas as células.

Lembrando que pontos pretos representam "1" e pontos brancos representam "0". Ambos os autômatos para condições iniciais aleatórias, com 100 caracteres de comprimento e o passo de tempo variando de $t_0 = 0$ até $t_f = 100$. A regra sempre os força a homogeneizar os elementos da

cadeia que os constituem. E de acordo com o aspecto visual do diagrama, Wolfram classificou as figuras triviais como Classe 1. Autômatos Triviais são casos especiais de Autômatos Periódicos.

ii) Classe 2: Autômatos Periódicos

Autômatos Periódicos sempre convergem para estados não triviais, de maneira estacionária. Eventualmente podem aparecer como diagramas fixos:

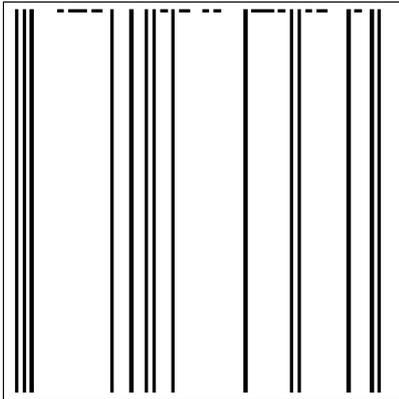


Figura 14 - Regra 4 (periódico).
Convergência a um estado único.

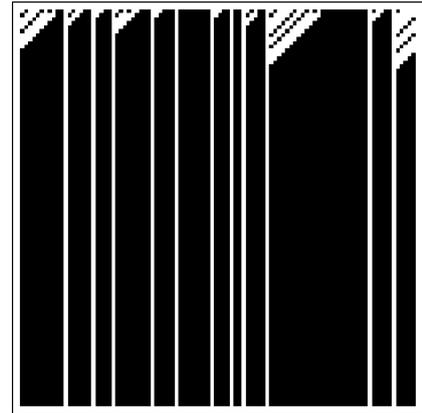


Figura 15 - Regra 202 (periódico).
Convergência a um estado único.

Ou como arranjos que são fixos por uma operação de rotação:

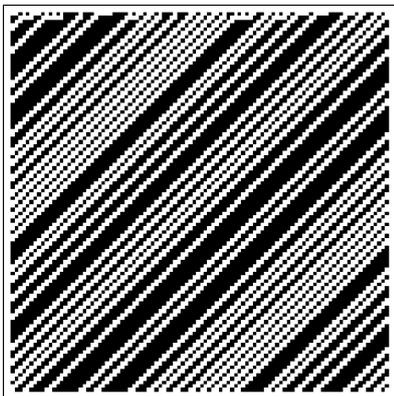


Figura 16 - Regra 139 (periódico).
Convergência a um estado único
com rotação.

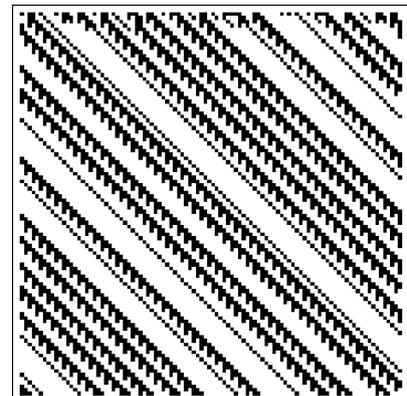


Figura 17 - Regra 88 (periódico).
Convergência a um estado único
com rotação.

Também são periódicos porque a regra atua sobre a cadeia apenas realizando uma operação de rotação sobre elas.

Discorrer a respeito da periodicidade de um autômato exige cuidado, pois mesmo com figuras mais complexas como as que se verá adiante, uma regra atua modificando uma cadeia em um estado e , eventualmente quando todas as possibilidades de combinação forem esgotadas, aquela mesma cadeia em que ela atuou retornará e seguirá o processo novamente. Isto porque as regras são ações determinísticas sobre as cadeias, e sempre agirão da mesma maneira se a mesma regra for aplicada sobre a mesma sequência. Assim, mesmo que leve mais passos, outros tipos de autômatos são, a rigor, periódicos, já que apresentam um período de ocorrência. Contudo, Wolfram classificou esses diagramas fixos como sendo periódicos.

iii) Classe 3: Autômatos Complexos

Autômatos Complexos exibem nos arranjos padrões e estruturas, as quais, ainda que sejam padrões e estruturas, se distribuem de maneira cujo ordenamento é difícil de classificar. São padrões entranhados e frequentemente manifestam ocorrência de auto-similaridade. Alguns:

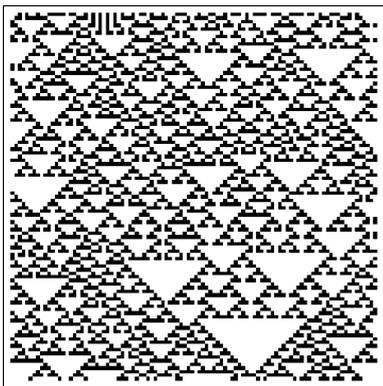


Figura 18 - Regra 22 (complexo).
Formação de estruturas com
distribuição aparentemente
aleatória.

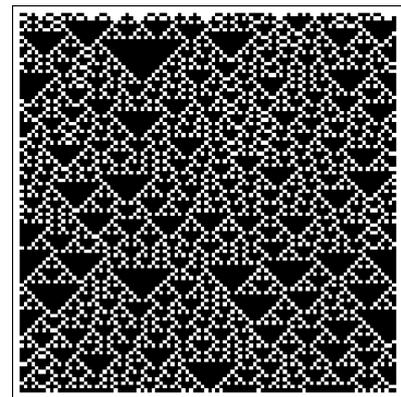


Figura 19 - Regra 182 (complexo).
Formação de estruturas com
distribuição aparentemente
aleatória.

Como dito anteriormente, uma regra sempre atua de um mesmo jeito sobre uma mesma condição, de forma que, inclusive nessas figuras complexamente classificáveis, eventualmente todas as possibilidades de sequências se esgotarão, e retornará a um estado progressivo, fazendo com que a regra atue e torne os eventos acontecerem novamente, caracterizando a manifestação de periodicidade. Mais adiante no texto, far-se-á a pergunta de por que uma regra leva mais ou menos tempo para encontrar uma região de ciclo.

iv) Classe 4: Autômatos Caóticos

De acordo com os padrões exibidos, a classe 4 de autômatos pode ser descrita como a classe de figuras que são aparentemente aleatórias, não exibindo padrões ou estruturas facilmente identificáveis, imprimindo a noção de que divergem completamente das condições iniciais com o passar do tempo:

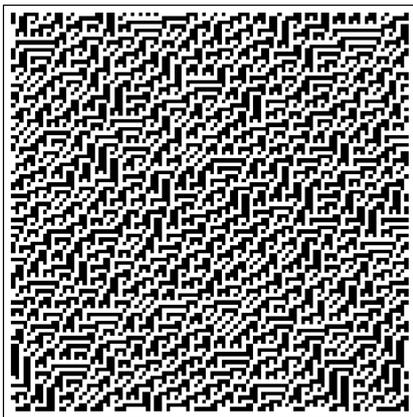


Figura 20 - Regra 75 (caótico). Sem formação de estruturas aparentes e distribuição aleatória.

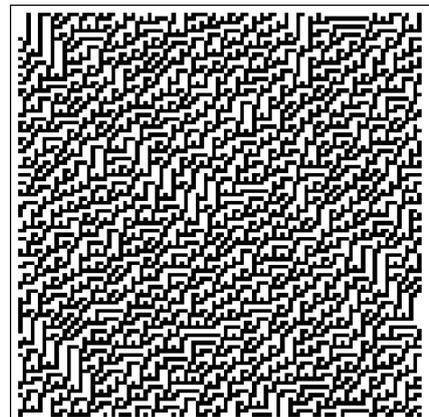


Figura 21 - Regra 45 (caótico). Sem formação de estruturas aparentes e distribuição aleatória.

Os quais, como se vê, têm suas cadeias distribuídas de maneira praticamente aleatória, dificultando o trabalho de encontrar padrões e estruturas. Mas é necessário ressaltar que essas cadeias não aparecem aleatoriamente ou tornam-se outras sequência de forma aleatória. Existe uma relação de causalidade entre elas, pois, como já foi dito, uma regra sempre atua da mesma maneira sobre uma mesma condição. É um programa, um algoritmo determinístico.

Ainda que existam autômatos que nitidamente podem ser classificadas desta ou daquela maneira, de acordo com esta ou aquela classe, há máquinas que simplesmente exibem comportamento misto, sendo difícil a discriminação entre um conjunto ou outro de aparências. Existem autômatos com comportamento entre o complexo e o periódico:

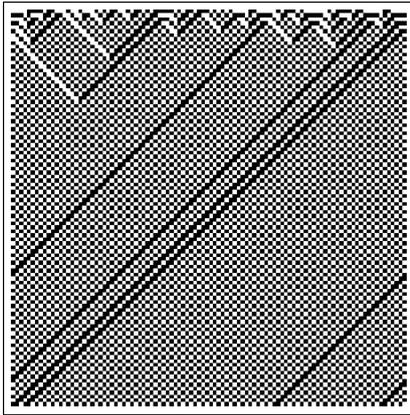


Figura 22 - Regra 57. Eventual formação de estruturas, ou repetição, mas com aspecto complexo.

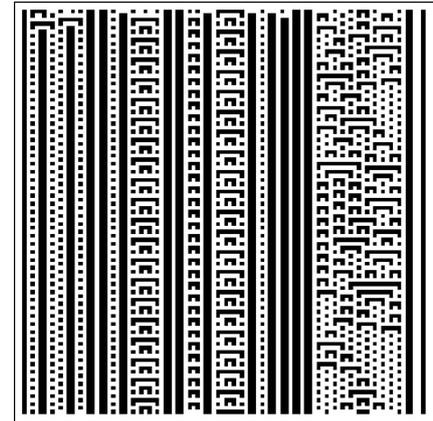


Figura 23 - Regra 73. Regiões do arranjo com aspecto periódico e regiões com aspecto randômico.

Existem outros, ainda, que se apresentam entre caóticos e complexos:

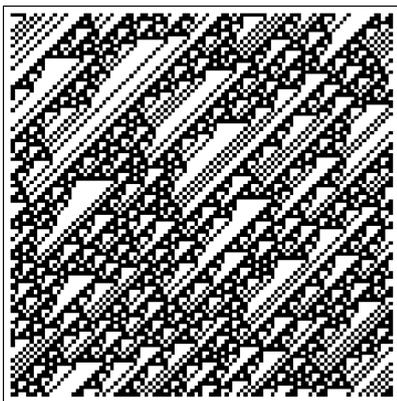


Figura 24 - Regra 106. Estruturas não necessariamente autossimilares (em forma), e com distribuição aparentemente aleatória.

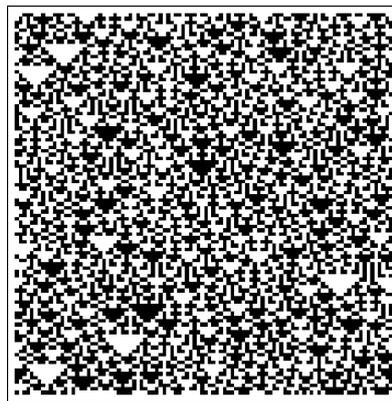


Figura 25 - Regra 150. Estruturas não necessariamente autossimilares (em cor), e com distribuição aparentemente aleatória.

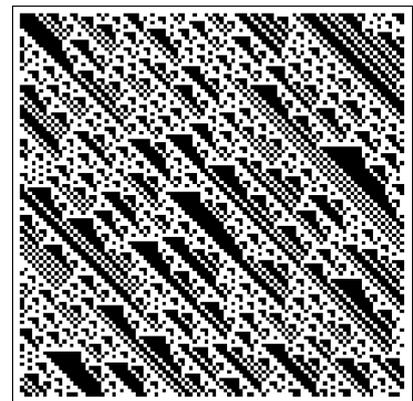


Figura 26 - Regra 225. Estruturas não necessariamente autossimilares (em forma), e com distribuição aparentemente aleatória.

Wolfram classificou os autômatos elementares de maneira heurística e qualitativa, baseando-se nos padrões das figuras. Isto têm implicações na generalizações do método, implica dificuldades por sua própria natureza qualitativa, posto que visual. De forma que, caso se queira levar em conta apenas o comportamento estrutural dos arranjos, é uma classificação muito útil, contudo enfraquece a discriminação objetiva entre as regras de atuação.

II.3 – Aditividade em Autômatos

Um assunto relevante para se tratar é a respeito da composição das regras dos autômatos. Como já mostrado que existem 256 regras elementares, e que a notação criada por Wolfram relaciona cada uma ao seu valor correspondente no sistema decimal, pode-se construir facilmente o modo de atuação dos algoritmos sobre as configurações celulares apenas escrevendo sua forma binária. Por exemplo, a regra 22 é expressa em base 2 da seguinte forma:

$$\begin{array}{c} 22_{10} \\ 10110_2 \end{array}$$

Mas para cobrir as 256 regras são necessários 8 bits, portanto escreva-se a com 8 caracteres:

$$\begin{array}{c} 22 \\ 0-0-0-1-0-1-1-0 \end{array}$$

Assim, dispondo todas as 8 combinações de 2 caracteres na vizinhança de 3 elementos ao longo das 8 posições:

$$22$$

Valor retornado quando	0	0	0	1	0	1	1	0
esta vizinhança é encontrada	111	110	101	100	011	010	001	000

da direita para a esquerda, aqueles valores são retornados no passo t quando a vizinhança relacionada é encontrada no passo $t-1$. Outro exemplo, a regra 160:

$$160$$

Valor retornado quando	1	0	1	0	0	0	0	0
------------------------	---	---	---	---	---	---	---	---

esta vizinhança é encontrada	111	110	101	100	011	010	001	000
------------------------------------	-----	-----	-----	-----	-----	-----	-----	-----

Simple assim para todas as 256 regras elementares, e todas as regras superiores e em outras bases numéricas.

Dentro desse conjunto de regras elementares, existe um grupo que é compatível com a adição em aritmética módulo 2 de estados, e sua principal característica é que resultados da combinação de estados independentes, gerados por exemplo por diferentes condições iniciais, são meramente calculados pela adição módulo 2 desses estados. Vale lembrar que a adição modular (aritmética modular como um todo) é aquela que antes de propagar um dígito para a próxima casa retorna ao valor inicial, como as somas das horas em um relógio.

Assim dada uma vizinha de dígitos:

$$[c_-, c_0, c_+]$$

A célula gerada no próximo estado é simplesmente uma combinação modular dessas vizinhanças.

Como a regra 90:

$$90 - c_- + c_+$$

Se esta vizinhança for encontrada	<u>111</u>	<u>110</u>	<u>101</u>	<u>100</u>	<u>011</u>	<u>010</u>	<u>001</u>	<u>000</u>
Este valor retornado quando	0	1	0	1	1	0	1	0

Apenas fazer a soma: $1 + 1 \pmod{2} = 0$, $0 + 0 \pmod{2} = 0$, $1 + 0 \pmod{2} = 1$.

Todo o conjunto de regras elementares aditivas pode ser encontrado com:

Tabela 1 - Regras Aditivas

Regra	Adição módulo 2
0	0
60	$c_- + c_0$
90	$c_- + c_+$
102	$c_- + c_1$
150	$c_- + c_0 + c_1$
170	c_1
204	c_0
240	c_-

Dentre as quais a 0 é nula; 170, 204 e 240 são periódicas; e 60 e 102 são essencialmente equivalentes à 90 ou 150, estas complexas [18, 19]. Essas duas últimas regras aditivas possuem propriedades dinâmicas interessantemente contra intuitivas, as quais serão discutidas posteriormente.

SEGUNDA PARTE – RESULTADOS E DISCUSSÕES

III – RESULTADOS E DISCUSSÕES. CADEIAS DE DADOS.

Cadeias de dados são sequências de objetos; para este trabalho, são ordenamentos específicos de caracteres ou dígitos em um alfabeto ou base numérica preestabelecidos. De todas as propriedades abordadas na teoria, a Complexidade de Lempel-Ziv LZ76 foi a mais extensamente estudada. E, a partir do conceito dessa complexidade C , um algoritmo capaz de gerar sequências de máxima LZ76 foi analisado [2].

Dado um comprimento L de cadeia qualquer, para a base numérica binária, existe um conjunto de combinações possíveis entre os elementos que geram os maiores valores de C para aquele comprimento. Essas sequências são de máxima complexidade de Lempel-Ziv. Ainda, dentro do conjunto de todas as sequências de máxima LZ76, existe um subconjunto que respeita o algoritmo reverso descrito em I.5.

Tal algoritmo, por esta forma, é capaz de gerar cadeias $E(MLZs)$ de máxima complexidade LZ76, que não são únicas por construção, e que exibem um conjunto específico de propriedades observadas. Um exemplo que recorde o caráter do processo pode ser feito para o espaço da base $\Sigma = \{0,1\}$:

$$\begin{aligned}E(MLZs) &= 1.0 \\E(MLZs) &= 1.0.11.00 \\E(MLZs) &= 1.0.11.00.111.010\end{aligned}$$

Até a terceira ordem, isto é, até subcadeias de tamanho 3, ou de maneira simétrica trocando-se 0 por 1 e vice-versa, sempre acrescentando fatores não repetidos de menor ordem lexicográfica possível de cada vez, até a ordem que se deseje.

É evidente que o último fator da sequência poderia conter uma quantidade menor de caracteres e sustentar a propriedade de máxima complexidade, fornecendo um alcance muito maior na quantidade de cadeias $E(MLZs)$. Entretanto, é razoável que se mantenha o valor da ordem nesse fator, tanto a fim de facilitar a escrita do algoritmo gerador, quanto para manter a homogeneidade e sistematização da construção.

Entre as propriedades observadas desse conjunto de sequências – submetidas em [2.1] –, é possível citar o cunho não aleatório das MLZs. A própria natureza algorítmica em si fornece essa qualidade, junto ao fato de sequências verdadeiramente aleatórias exibirem um conjunto específico de características.

Uma sequência verdadeiramente aleatória tem a propriedade de parecer aleatória para qualquer algoritmo que a tente descrever, tendo grandes complexidades de Kolmogorov-Chaitin. Também, deve passar no teste de aleatoriedade de Martin-Löf [20], o qual, entre outras, tem como consequência a incompressibilidade de todos os prefixos de uma sequência quando esta é infinita, isto significando, segundo interpretações, que a própria cadeia em si é incompressível.

Contudo, em virtude da existência de um algoritmo gerador das MLZs, as mesmas apresentam pequenas complexidades de Kolmogorov-Chaitin quando comparadas às sequências aleatórias, mesmo exibindo máximas complexidades LZ76. Outra distinção importante entre as duas classes de cadeias é que, dada uma fatorização do tipo LZ76, a distribuição dos comprimentos l dos fatores é notadamente diferente para cada uma:

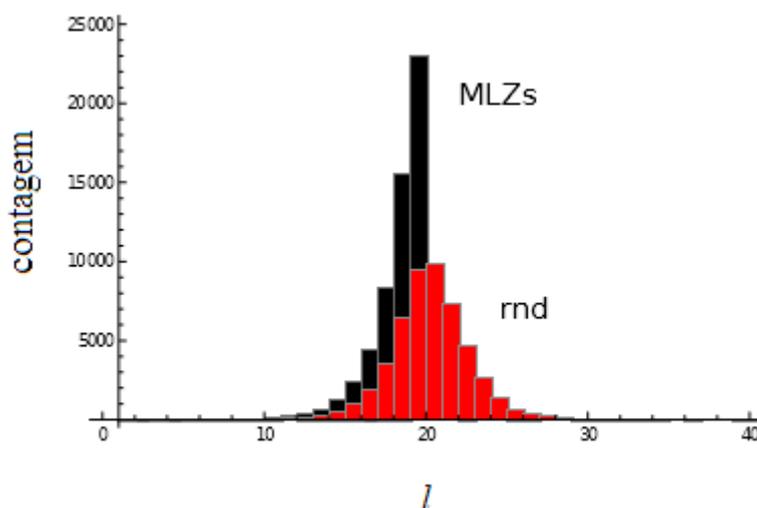


Figura 27 – Contagem x Comprimento. Distribuição dos comprimentos dos fatores de uma MLZs (em preto) e da média para 1000 sequências aleatórias (em vermelho), ambas com 10^5 símbolos.

Em que a distribuição *rnd* (vermelha) foi gerada para 1000 cadeias de 10^5 caracteres, aleatórias, e a MLZs é sequência obtida pelo algoritmo com o mesmo comprimento. A distribuição das sequências aleatórias obedecem um padrão gaussiano, enquanto a de máxima LZ76 pelo algoritmo reverso é explicitamente dessemelhante.

Isto é reflexo direto da normalidade de Émile Borel para os fatores de comprimento g , os quais, dada uma base numérica b , se distribuem com probabilidade b^{-g} . As sequências verdadeiramente aleatórias seguem essa distribuição normal, enquanto as MLZs não a respeitam necessariamente, ou seja, não são normais no sentido de Borel. É importante destacar que o conceito de normalidade é original para os números reais, porém, devido ao fato de sequências

serem identificáveis com números reais por uma transformação de base, uma cadeia aleatória também pode ser chamada de número real aleatório.

Esse comportamento é o que se pode relatar como sendo o mais notável a respeito da não aleatoriedade das cadeias MLZs. Entretanto, pode ser interessante verificar a distribuição de fatores relativamente às combinações disponíveis dos elementos nas sequências. Considere-se, então, que para um fator de comprimento l haja 2^l combinações disponíveis de dígitos binários. Assim, o número de ocorrências n_l de fatores de tamanho l define uma probabilidade:

$$p_l = \frac{n_l}{2^l} \quad (\text{III.a})$$

sobre todas as possibilidades. Esta probabilidade, ou normalização com respeito às combinações, define uma fração que se distribui com as contagens da seguinte maneira:

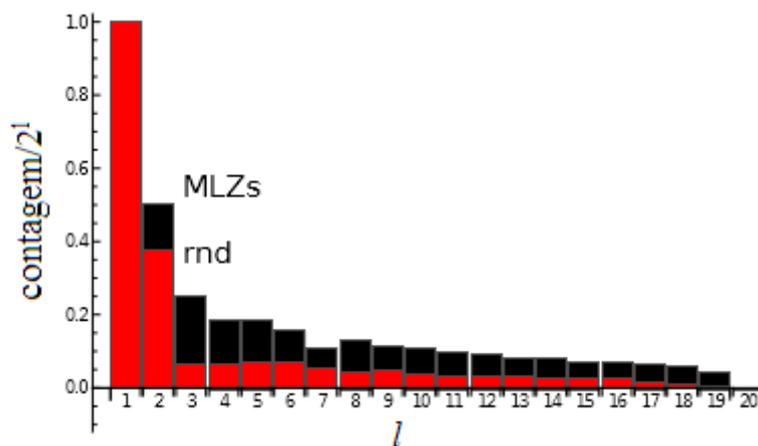


Figura 28 – Contagem Normalizada x Comprimento. Distribuição normalizada dos fatores pelo número de combinações disponíveis para cada comprimento de fator, das MLZs (em preto) e a média de 100 cadeias aleatórias, ambas com 10^5 símbolos.

Em que novamente as barras vermelhas representam as contagens aleatórias, a média para 100 neste caso, de cada comprimento de fator. Uma probabilidade de ocorrência pode definir uma entropia do tipo Shannon para cada l dentro das sequências:

$$H = -\sum_{l=1}^{\infty} p_l \log_2 p_l \quad (\text{III.b})$$

Que não é rigorosamente uma entropia de Shannon, mas descreve o conteúdo de informação relativo à fração de fatores. Para as 100 cadeias aleatórias geradas a entropia apresentou um valor médio $H(\text{rnd}) = 3.31$, e para a MLZs – ambas com $L = 10^5$ –, $H(\text{MLZs}) = 2.25$.

Embora as distribuições aleatórias apresentem comportamento mais organizado, o conteúdo de informação em si da contagem de fatores mostra o caráter menos randômico das MLZs, ou, ainda, que sua aleatoriedade não é bem distribuída, entendido que a máxima entropia está para quando todas as frações são idênticas na soma.

Outra característica digna de destaque com respeito às MLZs é o crescimento da complexidade LZ com o aumento do número de caracteres. Lempel e Ziv demonstraram que o crescimento da complexidade C é delimitado por uma borda como [6]:

$$C < \frac{N}{[1 - \varepsilon(N)] \log N} \quad (\text{III.c})$$

Em que $\varepsilon(N)$ é uma função de muito pequena expressão, *i. e.*, decai muito lentamente, e N é o número de símbolos, ou o tamanho da cadeia. $\varepsilon(N)$ está na ordem de 0.1 quando a base é 2 e N na região de 10^{50} símbolos.

Baseando-se nessa construção, um ansatz empírico com ajuste:

$$C = b \frac{N^d}{[1 - \varepsilon(N)] \log N} \quad (\text{III.d})$$

tal que $d = 1,06$ e $b = 0,153$ tenham gerado um encaixe como o seguinte:

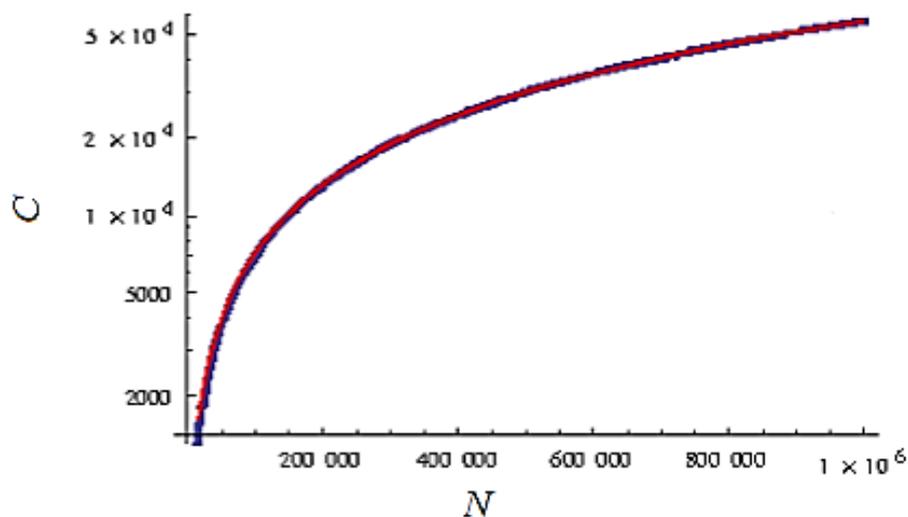


Figura 29 - Complexidade x Número de Símbolos. Em azul é o valor real de LZ76 para a MLZs e em vermelho o valor gerado pela equação (III.d).

Em que a distribuição em azul é da complexidade LZ76 e a curva vermelha é gerada pela equação (III.d). O erro relativo não excede 1,1% acima da região de 150.000 caracteres:

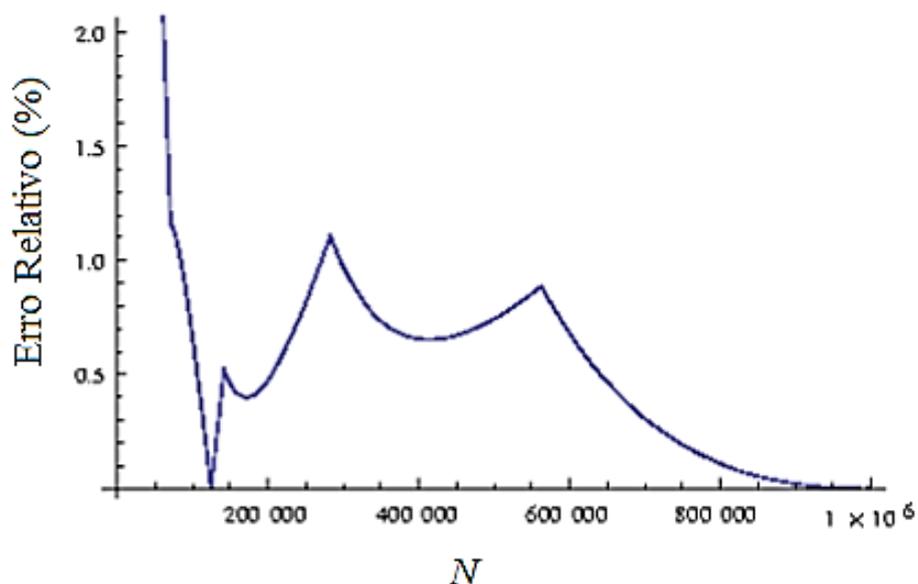


Figura 30 - Erro Relativo da Complexidade x Número de Símbolos. Erro relativo entre o valor real e o valor calculado pelo ansatz (III.d).

E decrescendo.

Outro modo de se verificar o comportamento da complexidade de uma MLZs é com a contagem dos fatores de Lempel-Ziv ao longo da cadeia. A cada ordem adicionada pelo algoritmo gerador, uma quantidade diferente de subcadeias de mesma ordem é colocada, e a

ordem, neste caso, coincide com o tamanho desses fatores. Assim, por exemplo, para a base $k = 2$, o seguinte crescimento é observado:

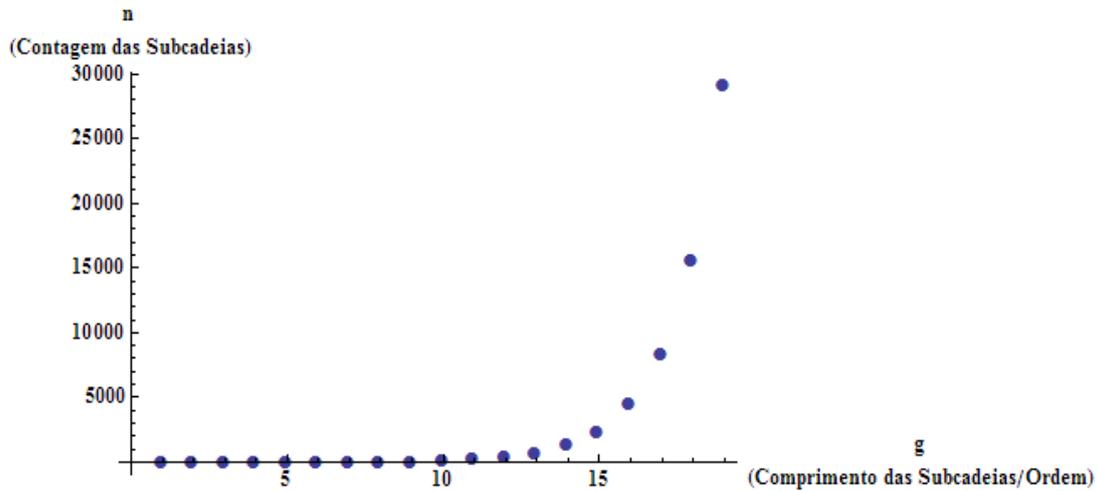


Figura 31 - Contagem de Subcadeias x Ordem (Base 2). A contagem do número de fatores LZ76 de mesmo comprimento relativamente a este comprimento dentro das MLZs

Até a ordem 16, no caso, ou até subcadeias na fatorização de tamanho 16; foi assim gerada uma cadeia de $L, N = 142.512$ de complexidade $C = 9.615$. Esse crescimento é visualmente bem comportado, o que é reforçado quando se muda a base numérica. Para $k = 3$:

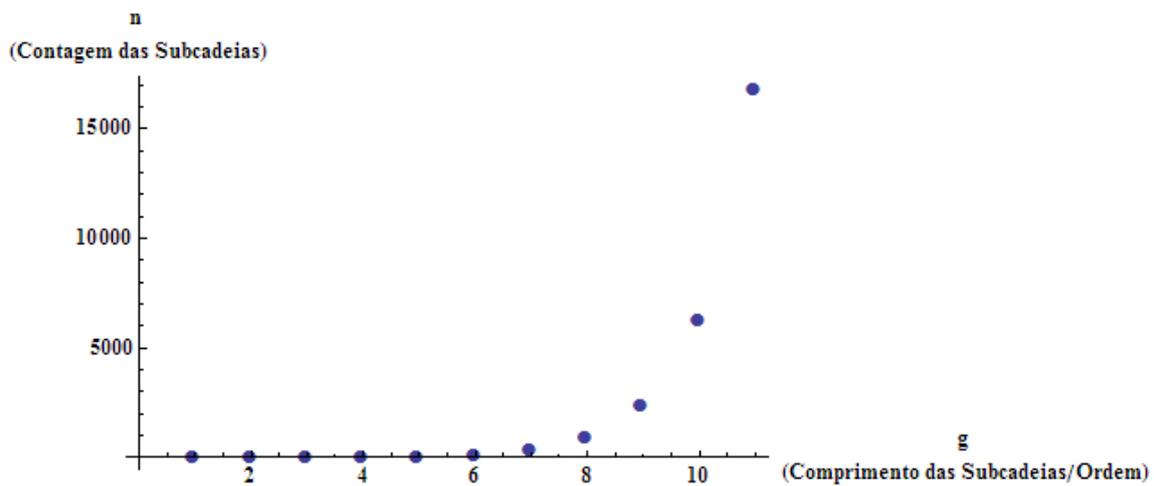
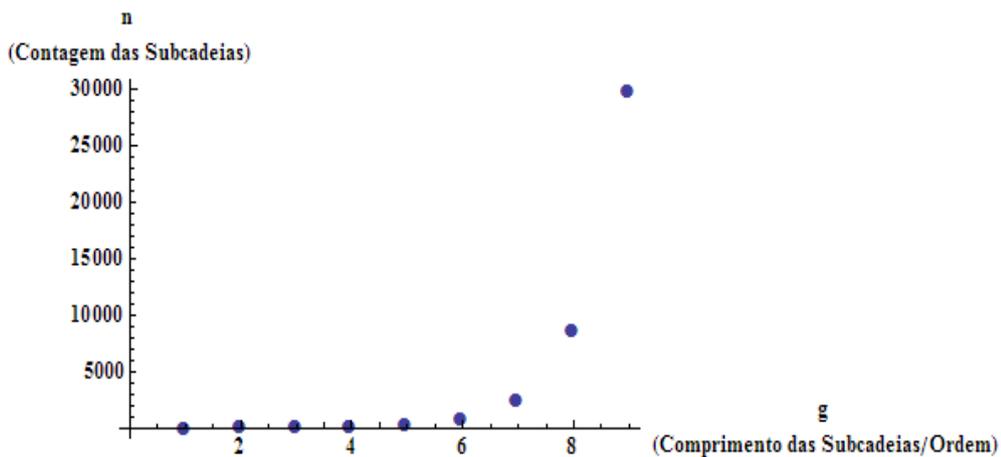


Figura 32 - Contagem de Subcadeias x Ordem (Base 3). Idem anterior.

Com $L = 280.142$ e $C = 26.972$, seguindo o mesmo comportamento, até bases numéricas superiores:

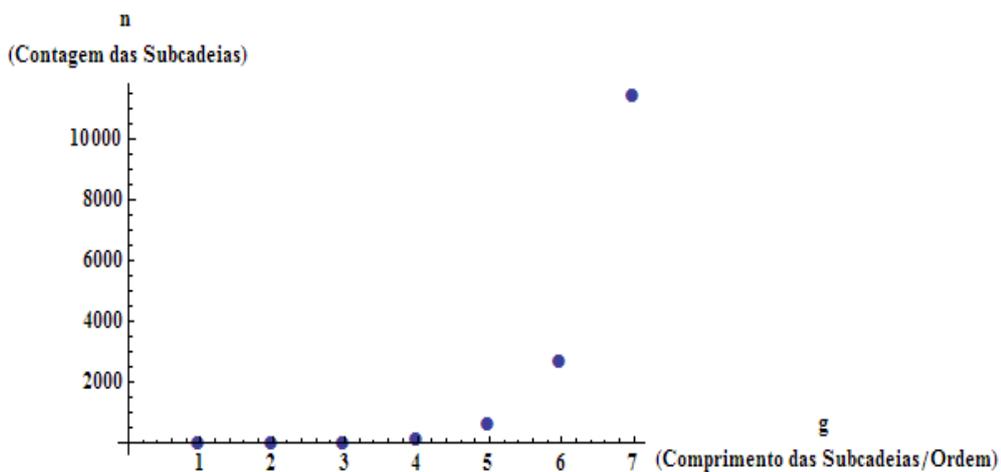


a)

$$k = 4$$

$$L = 360.220$$

$$C = 41.944$$

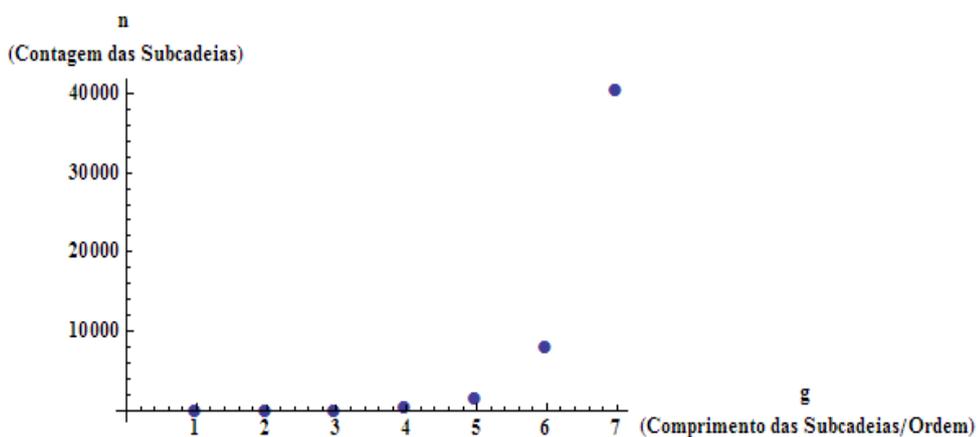


b)

$$k = 5$$

$$L = 100.584$$

$$C = 15.064$$



c)

$$k = 6$$

$$L = 340.330$$

$$C = 50.430$$

Figura 33 - Contagem de Subcadeias x Ordem (Outras Bases). Idem anterior.

Esse comportamento semelhante para todas as bases numéricas fez surgir o questionamento sobre a lei de formação de uma curva que pudesse se encaixar pelos pontos. Sendo o número de subcadeias n dependente da ordem g de um modo bem comportado, é possível imaginar, após várias suposições, que essa dependência é relativa a uma combinatória

dos elementos disponíveis para cada tamanho de fator. Assim, poderiam existir k^g fatores disponíveis em cada ordem g ocorrida, exceto aqueles que já teriam acontecido nas ordens inferiores. De modo que foi conjecturada a seguinte forma:

$$n = \frac{k^g}{(g-1)} \tag{III.e}$$

Uma combinatória intuitivamente adequada considerando as repetições em ordens inferiores, mas de difícil demonstração, embora verdadeiramente funcional:

$$k = 2$$

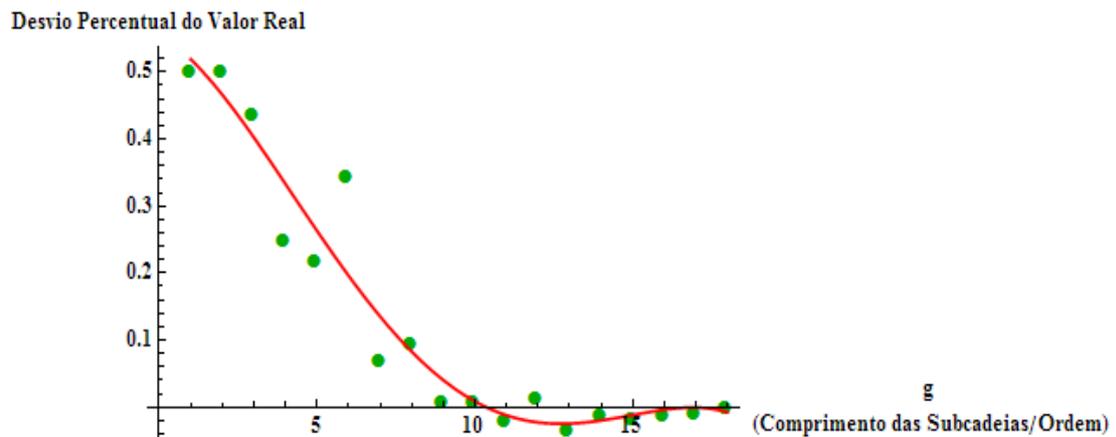
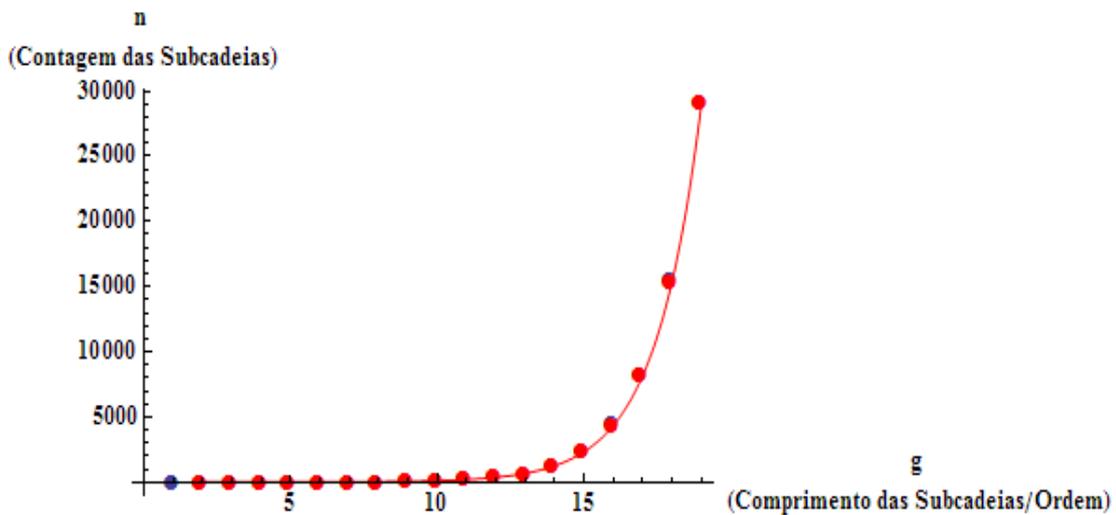


Figura 34 - Ajuste teórico da contagem (Base 2). No primeiro gráfico, a distribuição analítica (III.e) em vermelho se ajusta com adequação sobre os pontos reais, em azul. O segundo gráfico mostra o Desvio

Percentual do Valor Real x Comprimento das Subcadeias, e o decrescimento deste desvio com o aumento da ordem de cálculo.

No primeiro gráfico, o ajuste da curva e pontos teóricos em vermelho sobre os pontos reais em azul, e no gráfico inferior o desvio percentual entre esses pontos menos um ponto inicial que diverge, com a curva vermelha representando o ajuste computacional por polinômios. Igualmente para outras bases:

$$k = 3$$

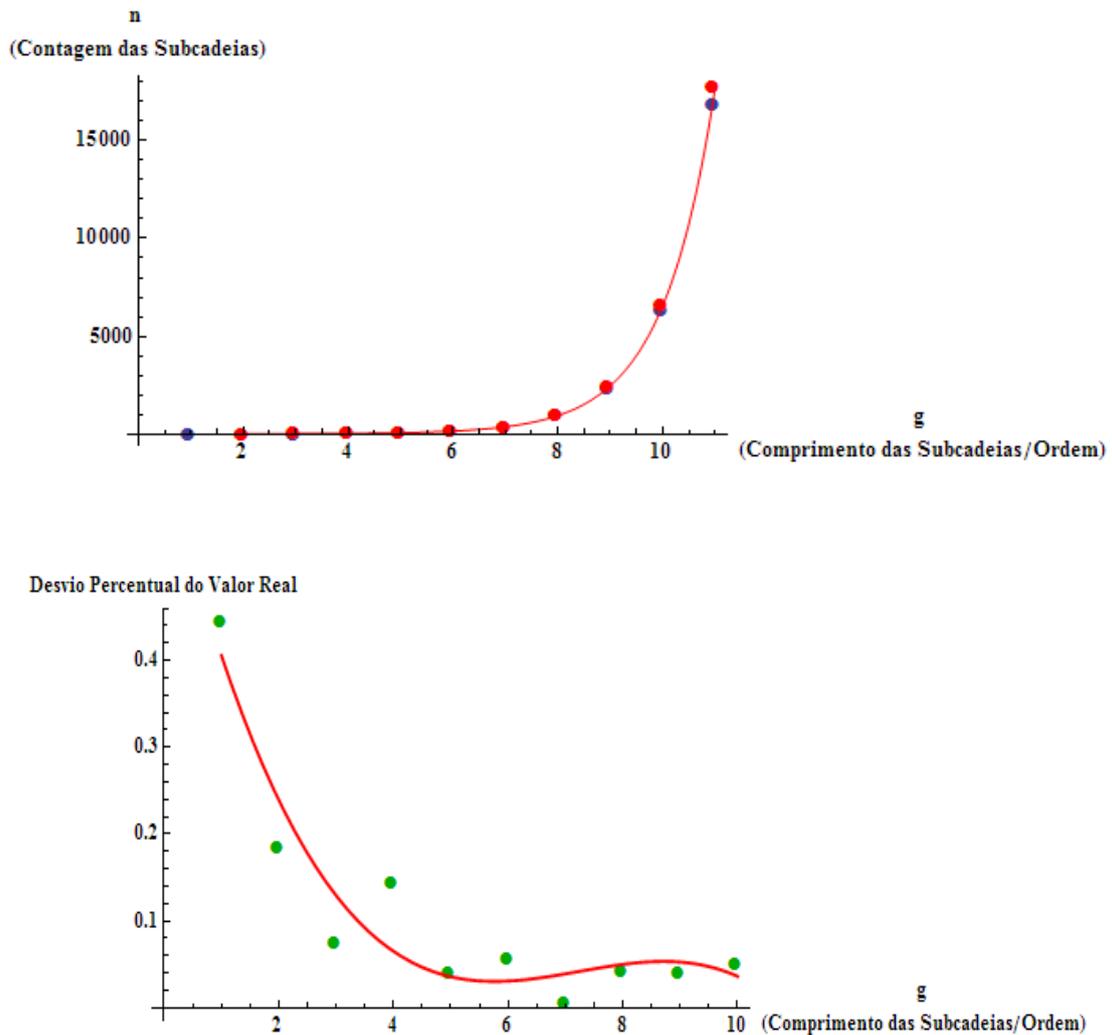


Figura 35 - Ajuste teórico da contagem (Base 3). Idem anterior.

Para a base 4:

$$k = 4$$

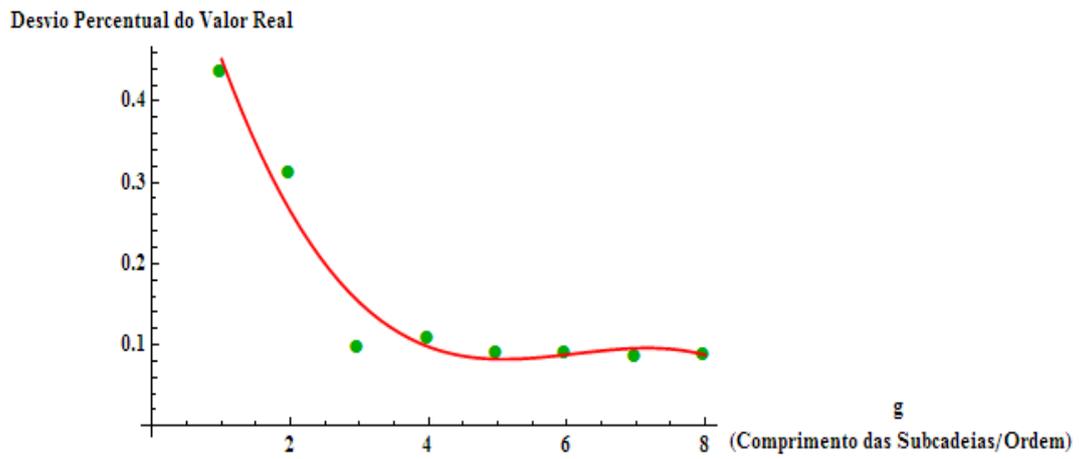
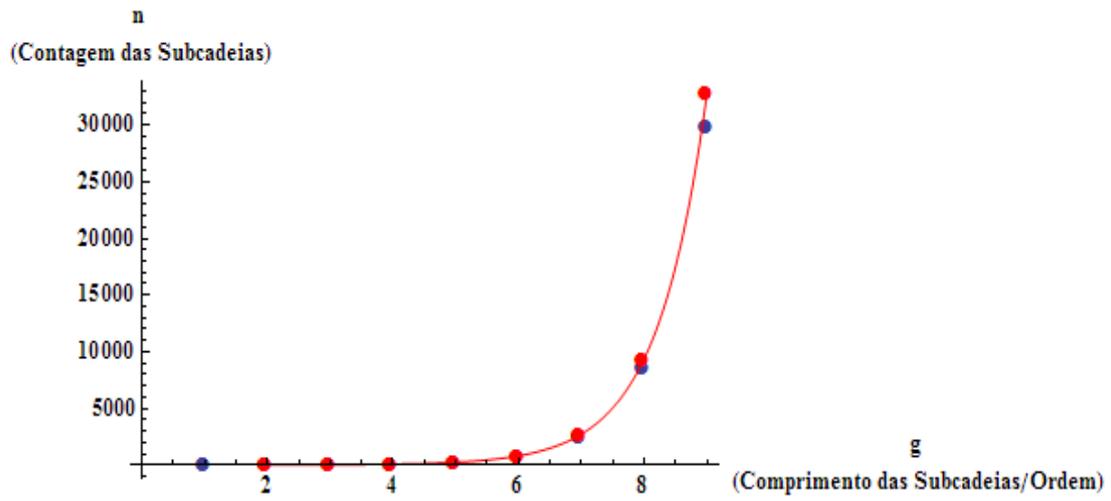
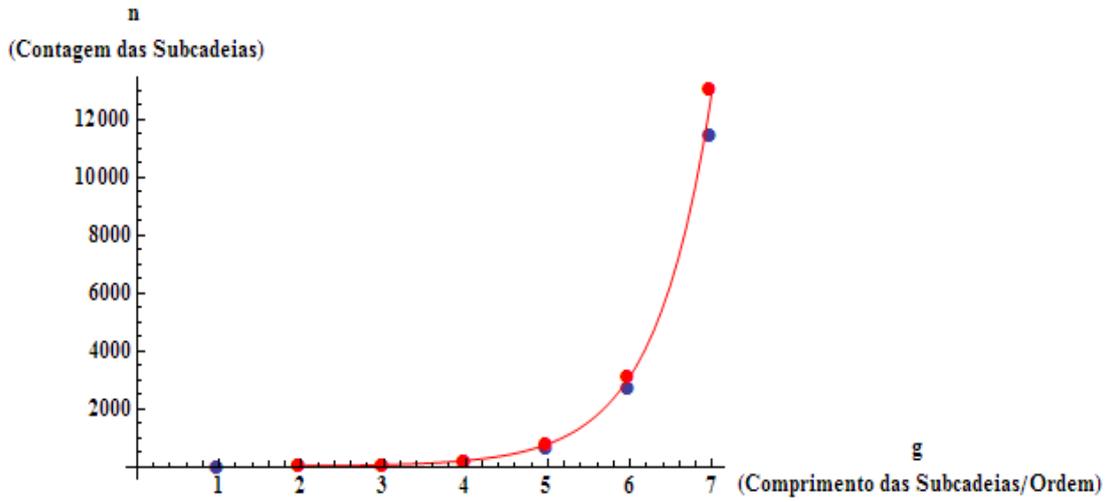


Figura 36 - Ajuste teórico da contagem (Base 4). Idem anterior.

Na base 5:

$$k = 5$$



Desvio Percentual do Valor Real

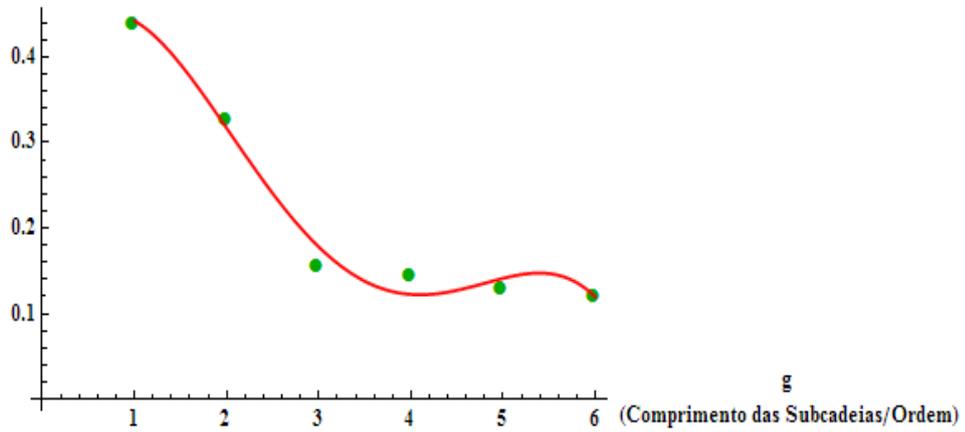
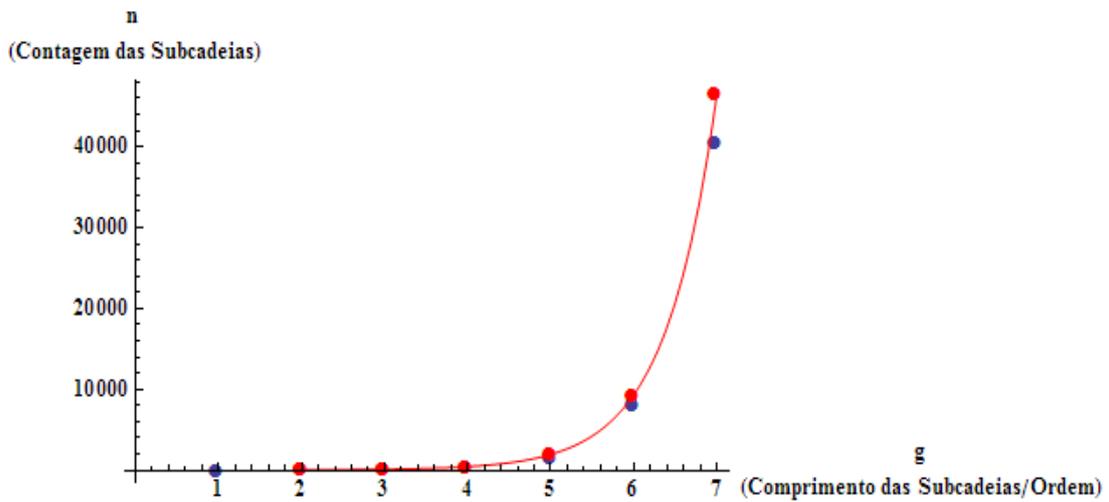


Figura 37 - Ajuste teórico da contagem (Base 5). Idem anterior.

$$k = 6$$



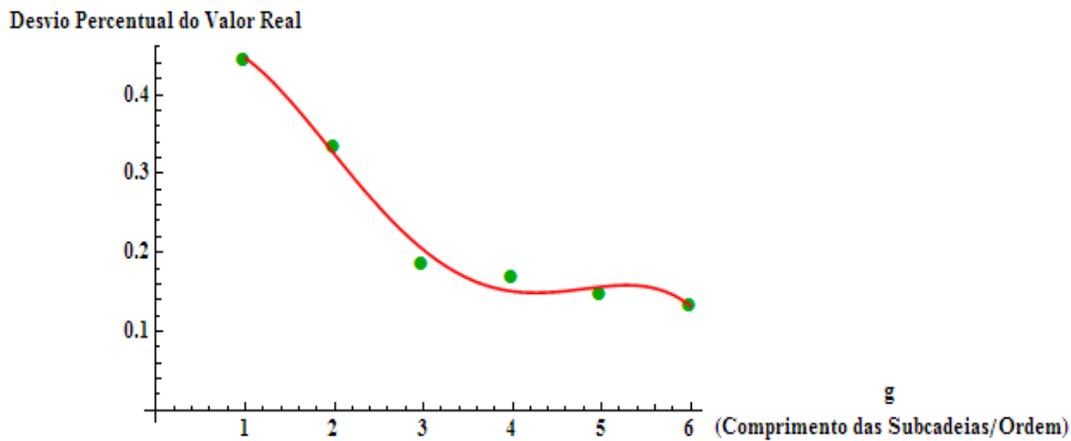


Figura 38 - Ajuste teórico da contagem (Base 6). Idem anterior.

É uma pequena quantidade de pontos, mas para sequências muito grandes e indica uma tendência, tornando, assim, pela falta de demonstração e pela funcionalidade outro ansatz experimental. De forma que, supondo a validade para o caso geral da equação (III.e), é possível se fazer extrapolações precisas para a complexidade LZ76 das MLZs.

A complexidade C de uma MLZs não é nada além do número de subcadeias ou fatores gerados pela fatorização; se em cada ordem g são gerados n fatores de tamanho g , então a complexidade total da sequência será:

$$C \approx \sum_{j=2}^g \frac{k^j}{(k-1)} \quad \text{(III.f)}$$

Até a ordem que se deseje. E o comprimento L total dessa sequência será a soma dos comprimentos de todos os fatores contidos em cada ordem. Então:

$$L \approx \sum_{j=2}^g j \cdot \frac{k^j}{(k-1)} \quad \text{(III.g)}$$

Visto que o tamanho das subcadeias em cada ordem é o próprio valor da ordem. A tabela:

Tabela 2 - Diversos desvios relativos para as complexidades e comprimentos

k	g	L_{real}	$L_{previsto}$	C_{real}	$C_{previsto}$	δ_L (%)	δ_C (%)
2	19	1.116.507	1.108.109	62.568	62.239	0,75	0,53

3	11	280.142	294.010	26.972	28.294	-4,95	-4,90
4	9	360.220	395.589	41.944	46.070	-9,82	-9,84
5	7	100.584	114.873	15.064	17.223	-14,21	-14,33
6	7	340.330	394.423	50.430	58.507	-15,89	-16,02

mostra que o módulo da diferença percentual entre o valor real e o valor da função é tão menor quanto maior for a ordem do cálculo. Entretanto, o cálculo é tão mais rápido quanto menor o valor da base numérica. Isso é indício de que, para cadeias de tamanho cada vez maior, mais precisa será a previsão do ansatz.

Essa suposição, de que a contagem de fatores cresce com a ordem de acordo com (III.e), deverá ser útil em diversas situações futuras, sobretudo para o caso binário, pois assumindo ordens de cálculo superiores será possível inferir o valor da complexidade de LZ e o tamanho da MLZs gerada com muitíssimo mais rapidez, fornecendo um teste mais visualizável do teorema de Ziv, o qual explica que:

$$\limsup_{N \rightarrow \infty} \frac{C}{N / \log N} = h = \lim_{N \rightarrow \infty} \frac{H(1, N)}{N} \quad (\text{III.h})$$

Para a sequência de tamanho entre 1 e N , sendo as MLZs (de máxima complexidade) fontes de normalização dessas taxas de entropia h ; entre outros usos possíveis do ansatz.

IV – RESULTADOS E DISCUSSÕES. AUTÔMATOS CELULARES.

A respeito da classificação do Wolfram para os autômatos celulares elementares, a Classe I se trata dos autômatos para os quais o diagrama de evolução temporal é trivial ou converge para a trivialidade, a Classe II o mesmo converge para o estado estacionário ou periódico, contendo, deste modo, a Classe I de acordo com o conjunto de propriedades. Na Classe III de autômatos estão os diagramas os quais, embora menos facilmente organizáveis, exibem estruturas com autossimilaridade distribuídas de maneira quase randômica, e a Classe IV é aquela cujos diagramas de evolução não exibem estruturas definidas espacialmente nem periodicidade aparente.

Por essas considerações as classes são respectivamente denominadas Nula, Periódica, Complexa e Caótica, e exibem, em geral, o mesmo tipo de comportamento para qualquer condição inicial.

Pode-se entender, por conta deste comportamento característico de cada classe, que as regras discriminadas em cada conjunto atuam sobre as condições iniciais, cada qual a seu modo, e as levam para determinados estados de ocorrência. De acordo com o que foi discutido na teoria do texto, uma regra, após atuar um determinado número T_c de vezes sobre uma condição inicial, leva o autômato a um regime estacionário, visto que, depois de esgotar todas as possibilidades de transformação das células, devido à sua natureza algorítmica e determinística, o último estado de um ciclo tornará o primeiro, e este sob a ação da regra iniciará novamente todas as sequências daquele intervalo cíclico, pois uma regra, enquanto algoritmo, só pode atuar de uma única maneira em uma cadeia.

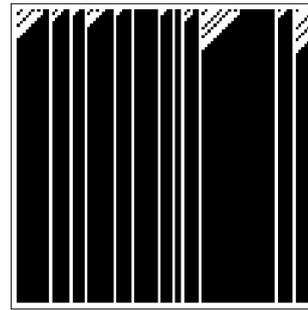
Quando um autômato entra em estado de ciclo, este terá um intervalo τ , um período, o qual é característico da regra utilizada e da condição inicial. E o que se pode verificar é que, tanto o tempo T_c que uma regra leva para transformar uma condição inicial qualquer no estado inicial do ciclo, como o próprio tamanho τ do ciclo, o período, dependem do tipo de classe à qual o autômato pertence.

Caso se note a forma de convergência ao intervalo estacionário das diversas classes:

Classe I

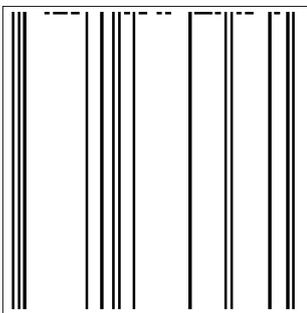


Regra 32

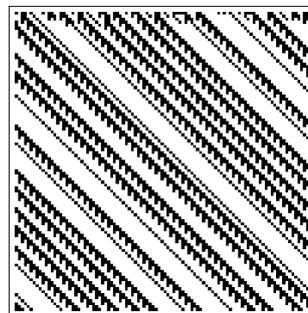


Regra 202

Classe II

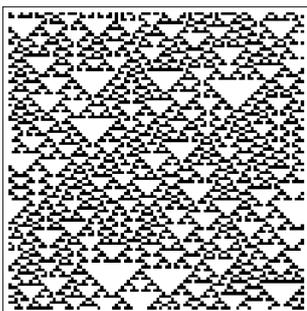


Regra 4

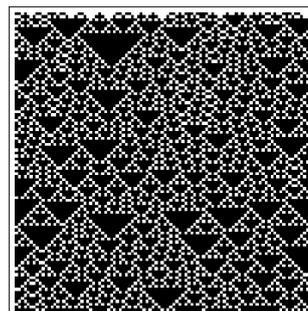


Regra 88

Classe III

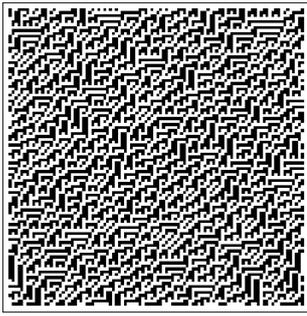


Regra 22

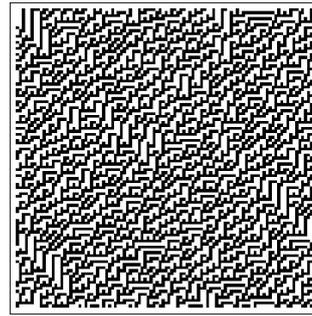


Regra 182

Classe IV



Regra 75

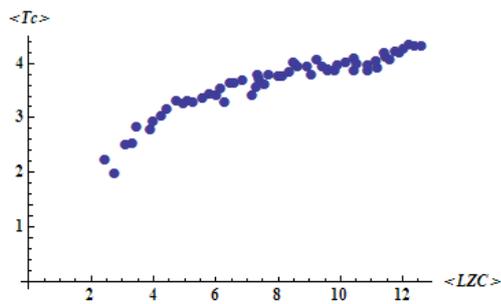


Regra 45

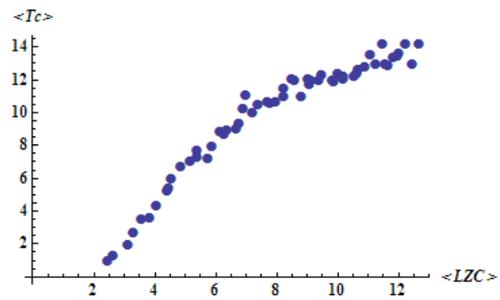
Conclui-se que regras triviais e periódicas convergem ao intervalo estacionário muito mais rapidamente que as demais, agindo mais fortemente sobre as condições iniciais, isto é, gastam muito menos tempo para levar a condição inicial aleatória até o ciclo do sistema.

Tomando a relação entre os T_c e os τ para diferentes classes, é possível verificar como esses valores variam em média com relação à complexidade das condições iniciais. Os seguintes gráficos exemplificam o comportamento geral dos tempos, períodos e complexidades médias:

Classe I



Regra 32



Regra 202

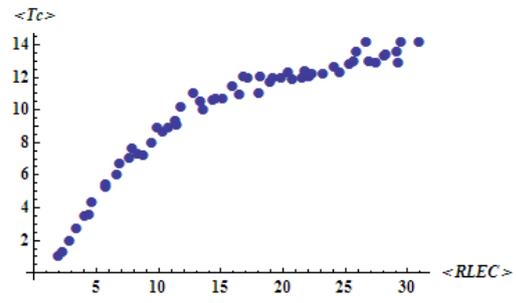
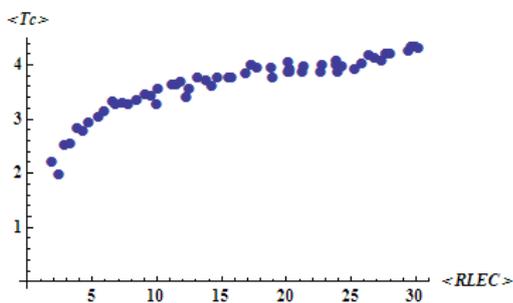
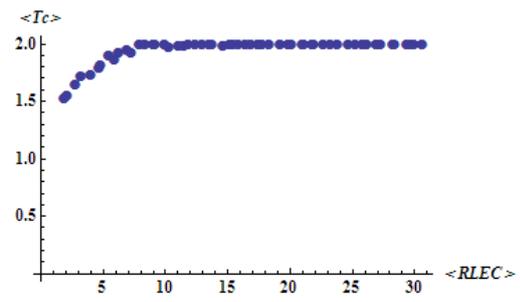
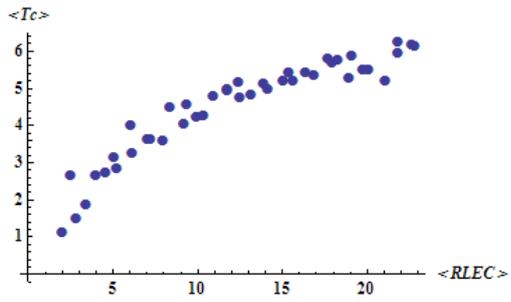
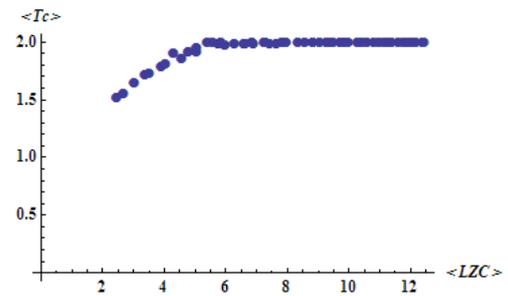
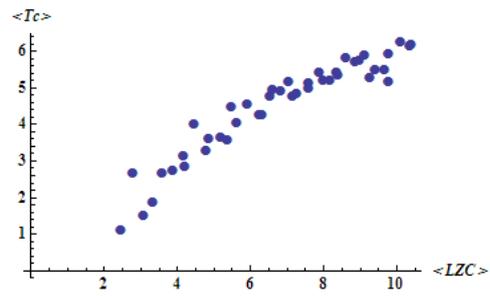


Figura 39 - Distribuição de T_c Médios x Complexidades Médias (RLE e LZ76) para Classe I, ambas para a Regra 32 e para a Regra 202.

Classe II

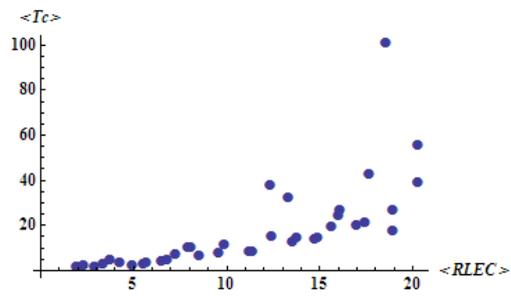
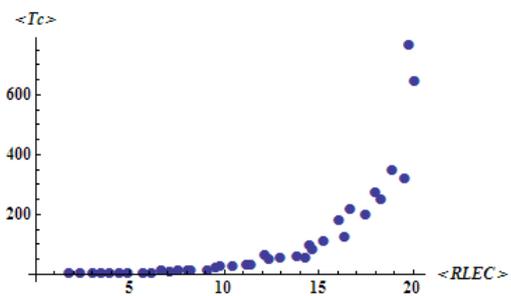
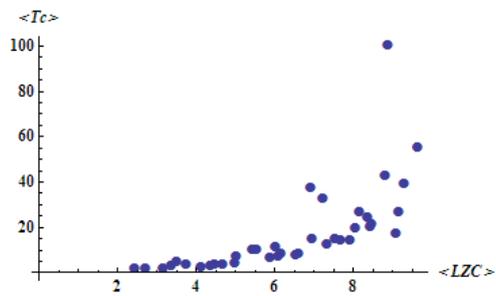
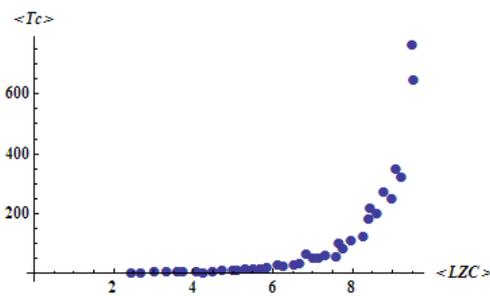


Regra 88

Regra 4

Figura 40 - Distribuição de T_c médios para Classe II, nos moldes da figura anterior.

Classe III



Regra 22

Regra 182

Figura 41 - Distribuição de T_c médios para Classe III. Idem anterior.

Classe IV

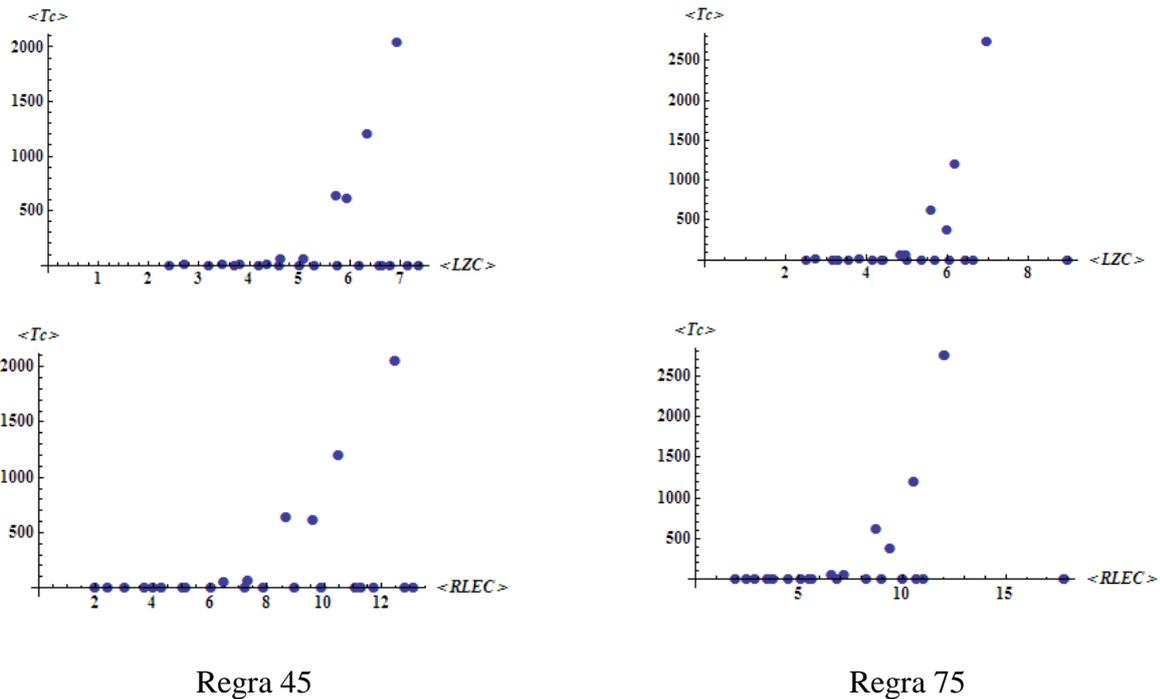


Figura 42 - Distribuição de T_c médios para Classe IV. Idem anterior.

Em que $\langle LZC \rangle$ e $\langle RLEC \rangle$ representam, respectivamente, as complexidades médias de Lempel-Ziv 76 e RLE, de acordo com as definições em I.5, para o estado desencadeador; para um conjunto de 100 condições iniciais pseudoaleatórias para cada comprimento de cadeias, estes variando entre 3 e 40, 50 ou 60, tamanhos suficientemente reduzidos para haver a manifestação de ciclos. O valor médio dessas complexidades iniciais, para cada comprimento, é, então, colocado em relação aos tempos médios, também para cada respectivo comprimento do estado.

A semelhança entre a forma das distribuições para cada regra, mesmo para complexidades de naturezas distintas, sugere que os instantes T_c não dependem das complexidades das condições iniciais por si, mas sim do seu comprimento ou número de caracteres L , visto que para uma quantidade estatisticamente adequada de cadeias aleatórias, as sequências maiores tenderão a exibir valores esperados de complexidade maiores do que as cadeia com menos elementos.

Embora o tempo que uma regra gasta para levar uma condição inicial até o estado estacionário dependa do tamanho dela, a forma dessa dependência pode variar entre as regras dentro de uma mesma classe, como se nota nas distribuições. Contudo, é explicitamente identificável que as regras complexas e caóticas exibem crescimento consideravelmente mais

rápido que as demais. Por este motivo que se deixou claro que os autômatos deveriam ter comprimento suficientemente pequeno para que haja manifestação de regiões estacionárias.

Neste sentido, a média dos T_c representa uma espécie de força inversa de uma regra que evolui um autômato, mas não deve ser a única propriedade relacionada à dinâmica da máquina. O tempo que um autômato permanece em regime estacionário também deve expressar a complexidade dinâmica se sua evolução.

Por complexidade dinâmica entenda-se também a complexidade medida por um algoritmo, pois um número de passos necessários até o intervalo de ciclo exprime quão complexamente uma regra atua sobre as cadeias, portanto diagramas de autômatos mais complexos ou difíceis de classificar devem ser matematicamente mais complexos, visto que a regra é que atua mais ou menos complexamente. Isto será mostrado adiante no texto, mas antes considere o mesmo conjunto distribuições anteriormente colocadas agora para o período τ dos autômatos. Seja o mesmo conjunto de autômatos:

Classe I

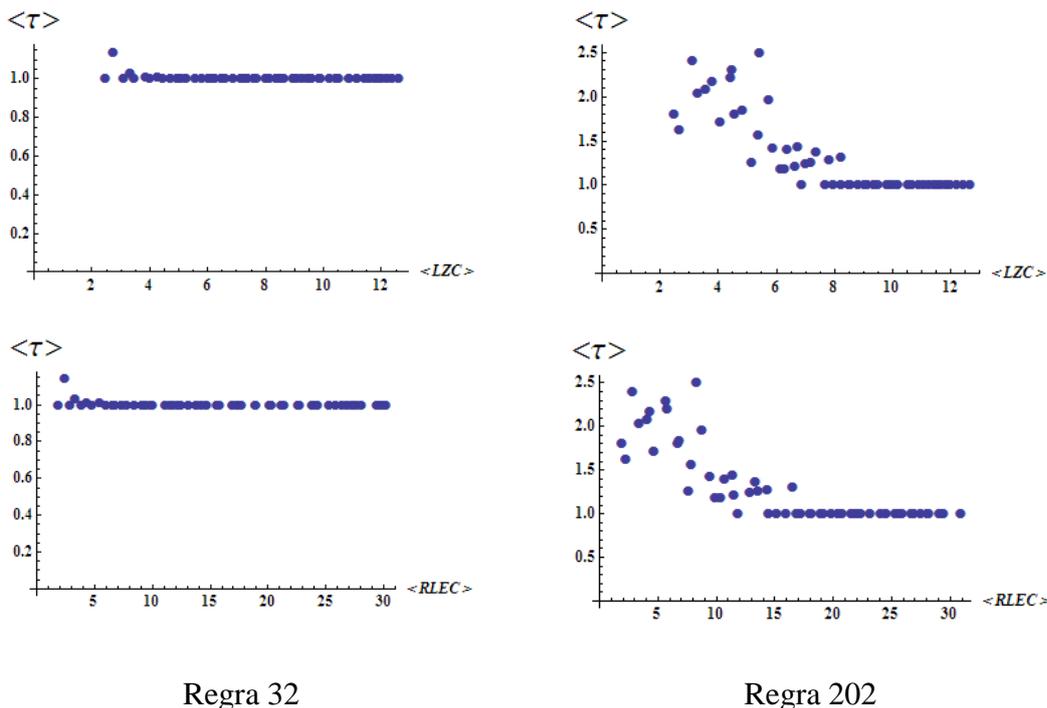
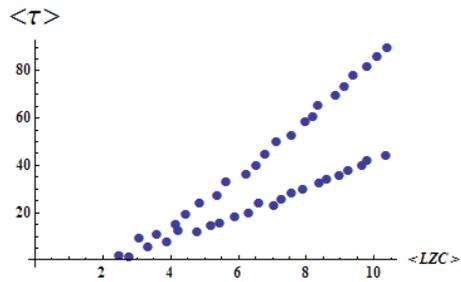
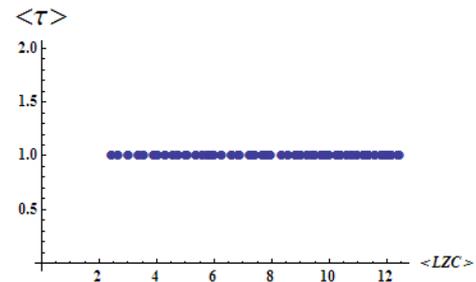


Figura 43 - Distribuição de τ médios para Classe I. Os períodos médios contra complexidades RLE e LZ76, ambos para a Regra 32 e para a Regra 202.

Classe II



Regra 88



Regra 4

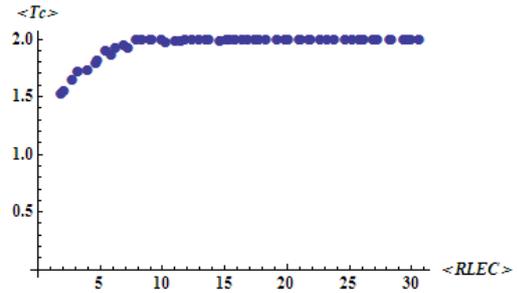
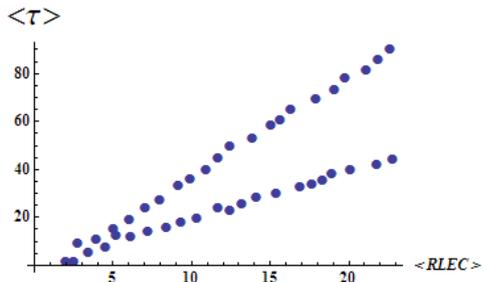
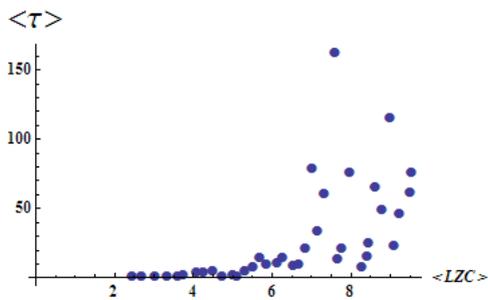
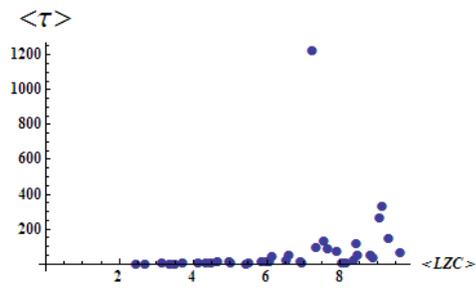


Figura 44 - Distribuição de τ médios para Classe II, nos moldes da figura anterior.

Classe III



Regra 22



Regra 182

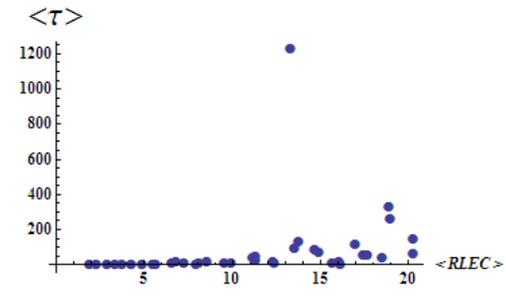
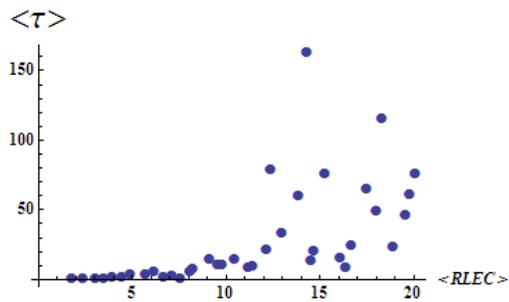


Figura 45 - Distribuição de τ médios para Classe III. Idem anterior.

Classe IV

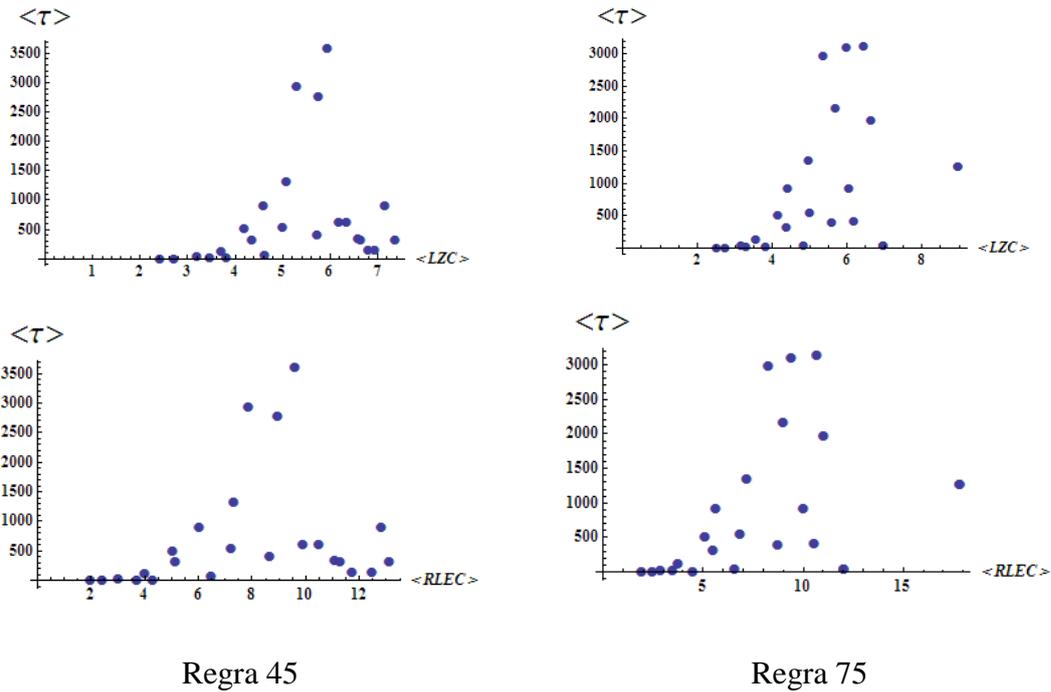


Figura 46 - Distribuição de τ médios para Classe IV. Idem anterior.

Para os quais, $\langle \tau \rangle$ representa os valores médios dos períodos dos autômatos, para o comprimento variando da mesma maneira ao caso anterior. O período τ é, como já foi dito, o tempo ou número de passos que um autômato exhibe *após* entrar em regime estacionário, isto é, quando as possibilidades de transições irreversíveis se esgotam, e isto ocorrerá a partir de um instante T_c , em que uma cadeia evoluirá até retornar a si mesma.

É importante destacar que a reversibilidade do autômato após T_c é genuína, pois se uma inversão do algoritmo acontecer, estados que anteriormente seriam causas, tornar-se-iam efeitos, tal como a recíproca, sendo o próprio algoritmo a origem da mudança de estado. Isto não é válido para configurações anteriores a T_c , pois não necessariamente um *estado-efeito* implicaria seu *estado-causa* pela inversão do código, do que vem sua irreversibilidade. Em outras palavras, a *informação* contida até o regime de ciclo será perdida.

Igualmente ao caso anterior, a forma das curvas independe do tipo de complexidade adotada, mudando de acordo com o tipo de condição inicial colocada, sugerindo que também depende do tamanho ou da forma dessa condição.

O tempo em que um autômato evolui de forma estacionária deve expressar também quantificação para sua complexidade, bem como o número de passos necessários até a condição de ciclos. Isto se deve à maior ou menor complexidade da evolução, ditada pela regra, manifestada pelo número de passos em questão.

É evidente que regras menos complexas como as periódicas e as nulas gastam um tempo menor para formar a região de ciclo, bem como a variedade de períodos encontrados. Regras caóticas permitem o surgimento de ciclos com os mais diversos períodos e instantes T_c igualmente diversificados, impactando diretamente na forma dos diagramas. Assim uma *variância σ (ou σ^2) dessas grandezas deve ser uma quantidade classificativa dos autômatos*. Contudo, antes deixe-se ficar mais claro a variedade de períodos exibidos pelos mais diversos autômatos.

Considere-se o conjunto de 1000 condições iniciais pseudoaleatórias de comprimento $L = 50$, os seguintes autômatos foram gerados sobre essas mesmas condições:

Classe I

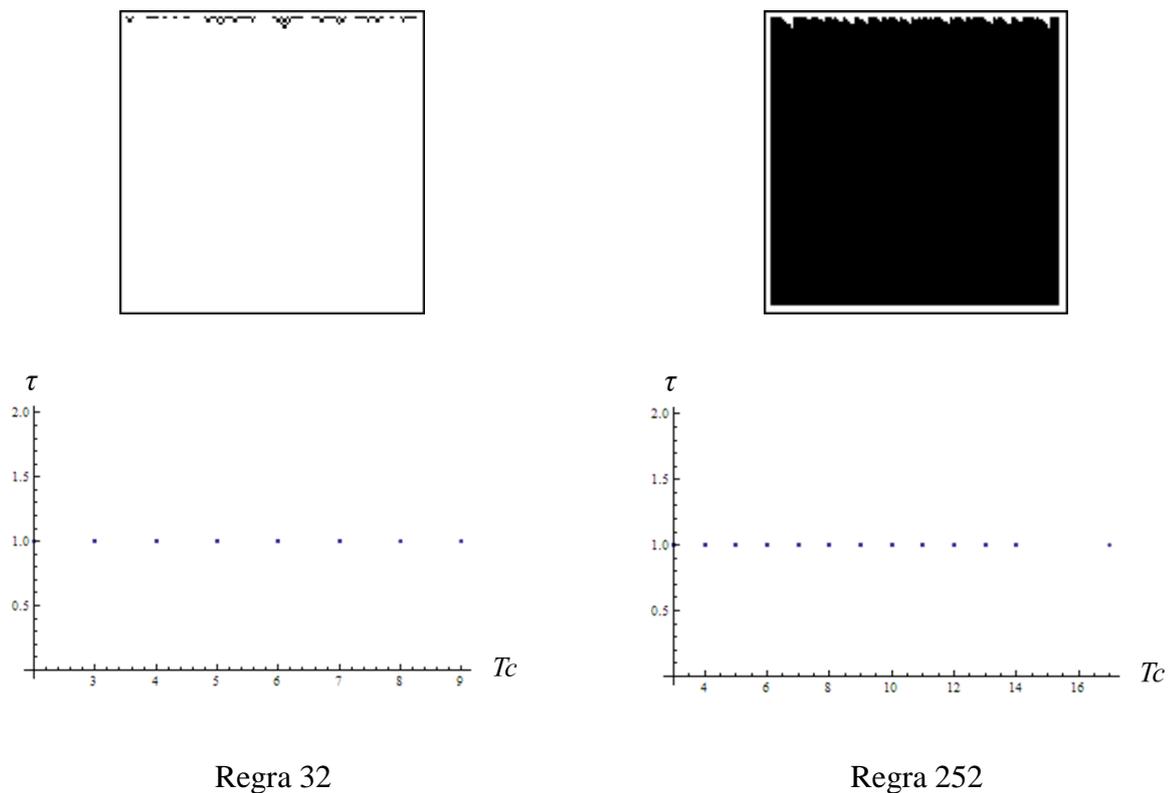
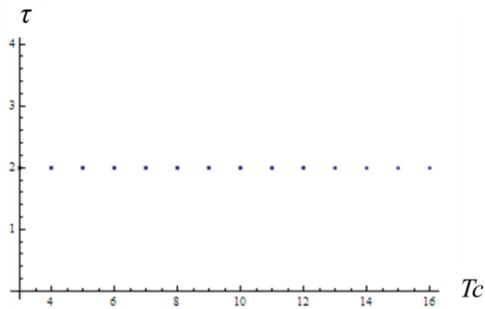
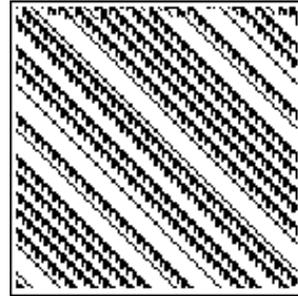
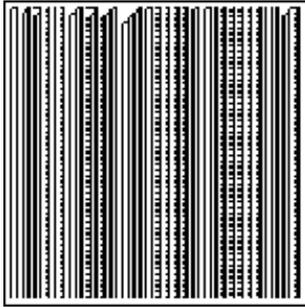
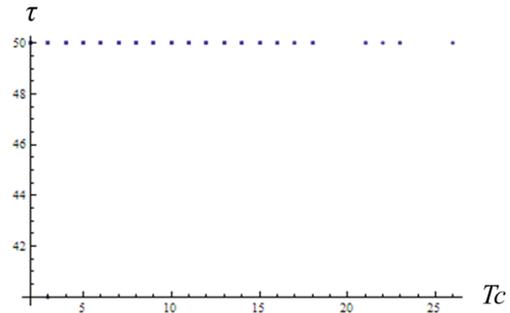


Figura 47 - Distribuição dos Períodos x Tempos Iniciais (Classe I). Distribuições de tempos para a Regra 32 e para a Regra 252.

Classe II



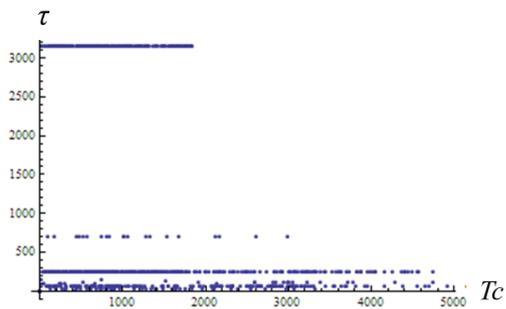
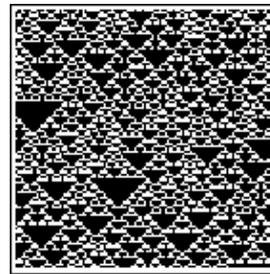
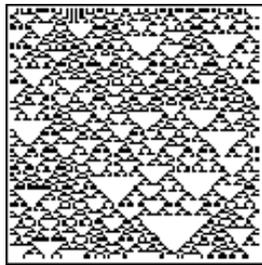
Regra 70



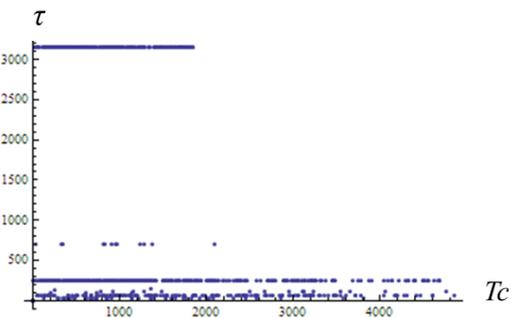
Regra 88

Figura 48 - Distribuição dos Períodos x Tempos Iniciais (Classe II). Idem anterior.

Classe III



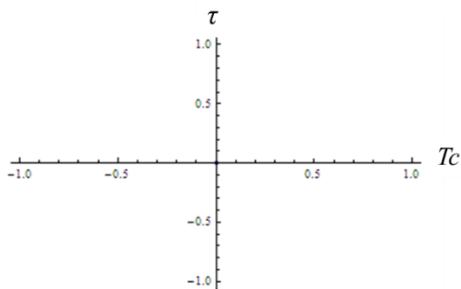
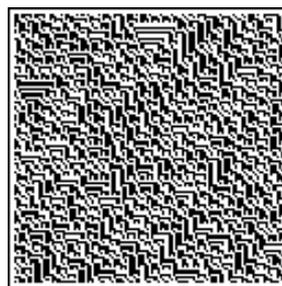
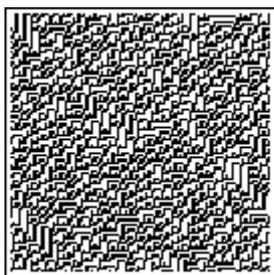
Regra 22



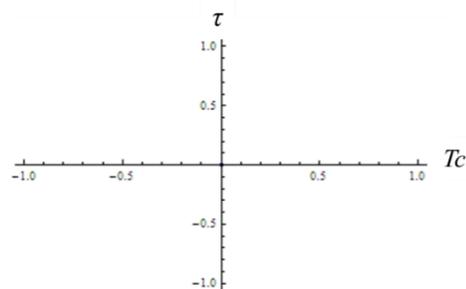
Regra 151

Figura 49 - Distribuição dos Períodos x Tempos Iniciais (Classe III). Idem anterior.

Classe IV



Regra 45



Regra 89

Figura 50 - Distribuição dos Períodos x Tempos Iniciais (Classe IV). Idem anterior.

Estas são as distribuições de cada T_c com o seu τ relacionado, para o mesmo autômato, ou seja, a mesma regra aplicada à mesma condição inicial. Nelas, observa-se que as classes mais simples – nula e periódica – apresenta uma pequena variedade de períodos τ distintos para dos diversos T_c , e, também, que aqueles estão na ordem de grandeza ou menores que $L = 50$. A Classe III exhibe grande variedade de períodos τ e instantes T_c , ambos muito grandes comparadas ao tamanho da sequência. A Classe IV não expressou ocorrência de ciclos, motivo pelo qual não há distribuição de pontos. Isto quer dizer que o tempo do autômato aplicado, $t = 5000$, não foi suficiente para alcançar o estado estacionário, diferentemente das outras classes.

Tais distribuições indicam que seu *desvio padrão* (ou *variância*) σ deve ser uma grandeza classificativa dos autômatos, haja vista a nítida relação com a complexidade dos arranjos. Autômatos mais simples devem apresentar pequenas variâncias (ou desvios padrões, a depender da normalização adotada) com relação à distribuições de períodos e instantes críticos de ciclo, e autômatos mais complexos, maiores dessas grandezas.

Uma maneira mais elegante de se representar a existência de ciclos nessas máquinas, e verificar a complexidade da atuação de uma regra sobre uma condição inicial, é imaginar o autômato de acordo com a sua natureza.

Autômatos Celulares são máquinas circulares, o que diz que suas bordas são comunicáveis e interagem entre na evolução dos mesmos. Assim, uma condição inicial com L fixo e igual a 10, por exemplo, evoluirá como um anel de 10 células até quando se deseje. Os arranjos, como são feitos, podem passar a ideia ou acostumar o leitor com o equívoco de que são estruturas bidimensionais, quando na verdade são anéis (unidimensionais) que se alteram com o passar do tempo, aqueles colocados sobre uma linha reta e este como uma segunda coordenada. Mas isto já foi devidamente colocado em teoria.

O caso é que a representação circular dos autômatos pode expor uma característica interessante de como uma regra atua sobre as sequências consideradas.

Imagine-se, então, que uma dessas máquinas com L caracteres seja representada como um polígono regular com L vértices (e L lados), cada lado com tamanho igual à unidade. Assim, um autômato com $N = L = 5$ caracteres estaria para:

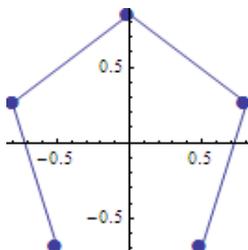


Figura 51 - Polígono de Sequência. Distribuição poligonal de pontos em que cada vértice representa um caractere preenchido em uma sequência binária cujas extremidades são células vizinhas.

Cada vértice representando um caractere.

Para o caso dos autômatos elementares, defina-se que se o elemento celular for 1 (um), então este será um ponto, se for 0 (zero), ele será um espaço vazio. Então o anel de células $\{0,0,0,1,1,1,1,1,0,0,1,1,0,1,0\}$, por exemplo, será expresso como:

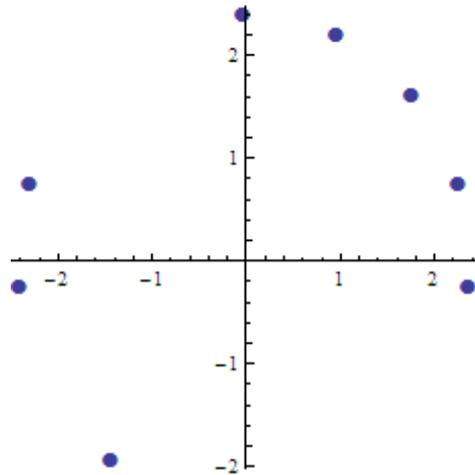


Figura 52 - Distribuição de Pontos do Polígono. Pontos marcados representam 1 e pontos ausentes representam 0.

Iniciando do quarto quadrante em sentido anti-horário, e terminando simetricamente ao eixo ordenado. O autômato é, então, representado como um anel, como realmente é, sendo a sequência uma poeira dispersa sobre o polígono regular.

Se a célula preenchida com 1 tem, diga-se, um peso estatístico igual a 1, é possível calcular um centro estatístico ou centro de massa para cada distribuição. Isto é útil na verificação de como a regra atua sobre as cadeias, visualizando-se como ela atua sobre o centro de massa e construindo a dinâmica da máquina. O centro estatístico (doravante centro de massa) será um ponto no interior do polígono, com coordenadas calculadas segundo:

$$C_M = \frac{\sum_i m_i r_i}{\sum_i m_i} \quad \text{IV.a}$$

No exemplo da distribuição acima, o centro de massa estará em {x,y}:

{0.168971,0.656119}

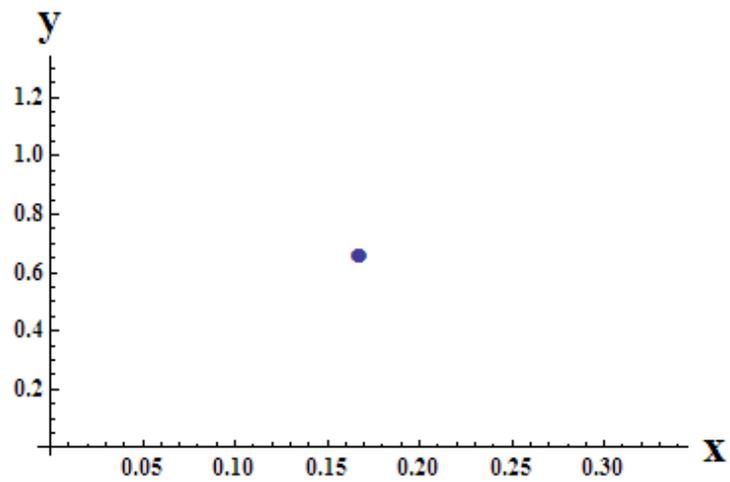
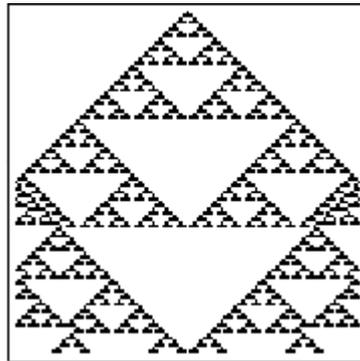


Figura 53 - Centro de Massa da Cadeia. Ponto representativo da localização ponderada das distribuição poligonal de pontos.

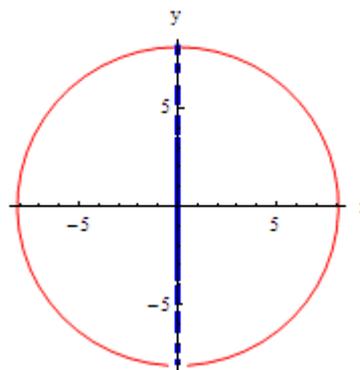
Fazendo com que cada sequência ao longo da evolução do autômato celulare esteja para um centro de massa, e essa evolução sendo representada como um terceiro eixo coordenado. Assim, as seguintes representações são obtidas:

Regra 22 com condição inicial 1 ponto no meio.



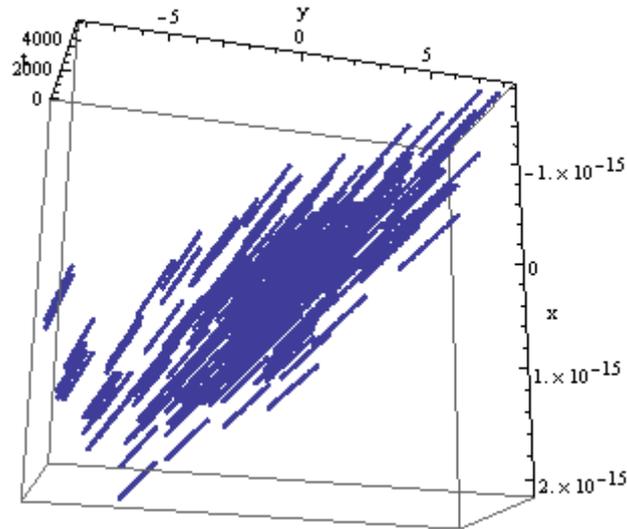
a)

Pontos azuis são a distribuição 2D dos centros de massa. A linha vermelha é o polígono de N lados circunscritor dos centros de massa.



b)

Evolução temporal dos centros de massa. A precisão da máquina encontrou largura mínima na ordem de 10^{-15} , mas é virtualmente nula.

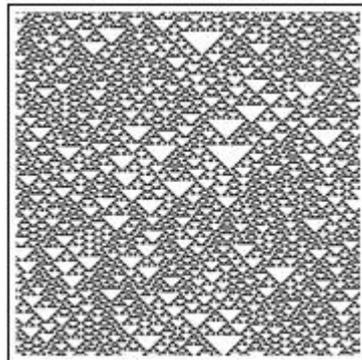


c)

Figura 54 - Evolução do Centro de Massa. a) Arranjo do autômato celular, b) distribuição das coordenadas dos pontos de massa de cada estado de atualização do autômato, c) distribuição das coordenadas dos pontos de massa de cada estado de atualização do autômato segundo a evolução temporal (terceiro eixo).

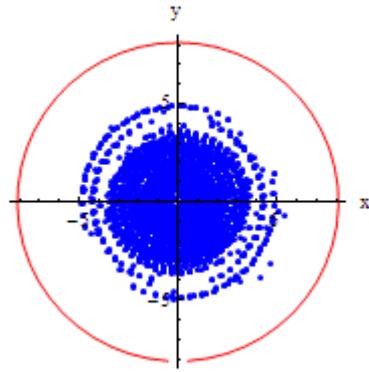
Para o caso de um autômato de comprimento $L = 51$ correndo até $t = 5000$, com condição inicial um ponto no centro da sequência. A largura do gráfico tridimensional é nula, e surgiu no caso por uma questão de precisão da máquina. Note-se a formação de colunas, o surgimento dessas estruturas implica que o autômato encontrou uma região estacionária. Agora para uma condição inicial aleatória:

Regra 22 com condição inicial aleatória.



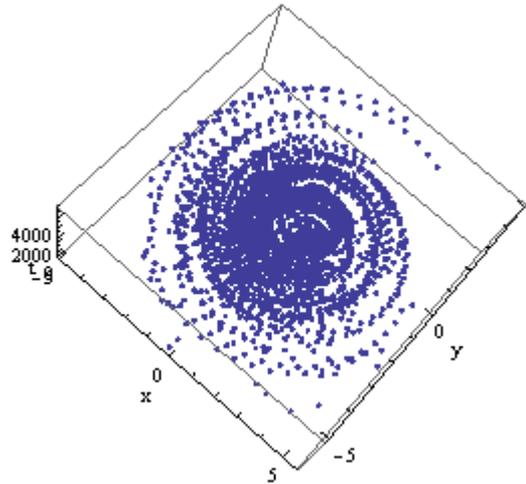
a)

Pontos azuis são a distribuição 2D dos centros de massa. A linha vermelha é o polígono de N lados circunscritor dos centros de massa.



b)

Evolução temporal dos centros de massa. Existe a formação de uma estrutura espiral, que é a circulação do centro de massa em torno de uma região média.

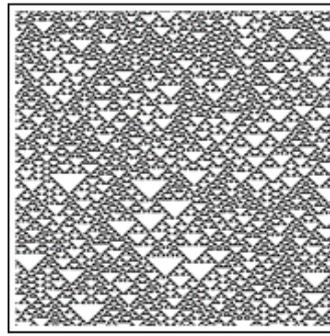


c)

Figura 55 - Evolução do Centro de Massa (condição inicial aleatória). Idem anterior.

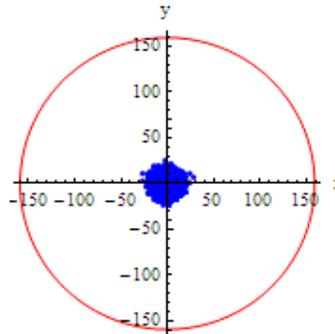
Neste caso a condição inicial aleatória criou um arranjo menos facilmente classificável, o que refletiu numa distribuição mais complexa do centro de massa. A circulação do mesmo em torno de um ponto e evoluindo no tempo gerou um conjunto de espirais. Elas são estruturas e também expressam a manifestação de ciclos ou de um estado estacionário. O mesmo comprimento $L = 51$ correndo sobre $t = 5000$ foi utilizado. Quando o autômato é demasiado grande, tempos nessa ordem de grandeza não são suficientes para gerar ciclos, portanto não o aparecimento de estruturas não será verificado:

Regra 22 com condição inicial aleatória e comprimento muito grande.



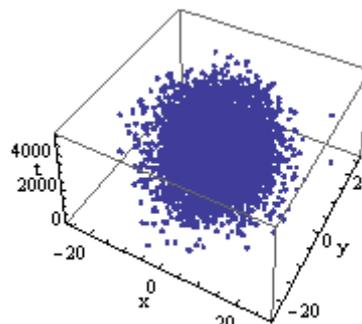
a)

Pontos azuis são a distribuição 2D dos centros de massa. A linha vermelha é o polígono de N lados circunscritor dos centros de massa.



b)

Evolução temporal dos centros de massa. Não existe a formação de estruturas, os centros de massa são distribuídos aleatoriamente.



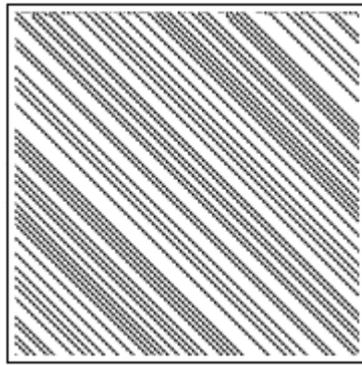
c)

Figura 56 - Evolução do Centro de Massa (Autômato muito grande e condição inicial aleatória). Idem anterior.

Quando o comprimento do autômato é grande o suficiente para impedir o aparecimento de estados estacionários, as estruturas bem comportadas também não surgem, tornando a evolução temporal da posição do centro de massa uma distribuição com características aleatórias. Neste caso foi utilizado $N = 1000$ e $t = 5000$.

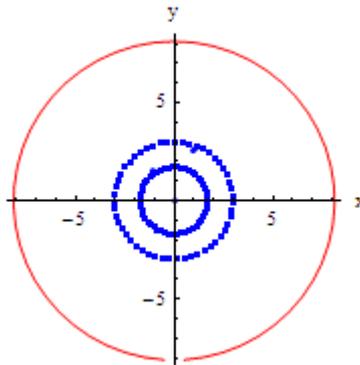
Em todos esses casos o centro de massa se movimenta sobre o plano do anel, o movimento é mais regular ou menos regular dependendo do tamanho do autômato e do tipo de regra utilizada. A regra 22 é complexa, do tipo III, e espera-se que forme padrões de movimentação mais complicados que regras periódicas, por exemplo. Segue mais dois exemplos para ilustração:

Regra 20 com $L = 51$.



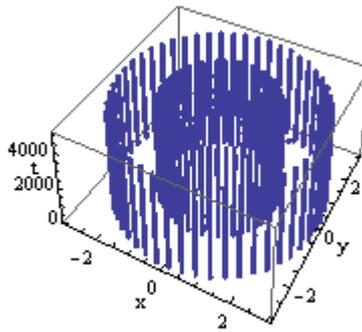
a)

Duas regiões de centro de massa distintas.



b)

Evolução temporal com formação de estruturas.

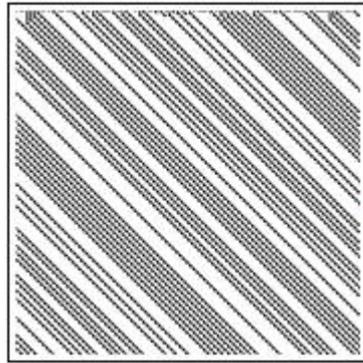


c)

Figura 57 - Evolução do Centro de Massa (Regra 20 e $L = 51$). Idem anterior.

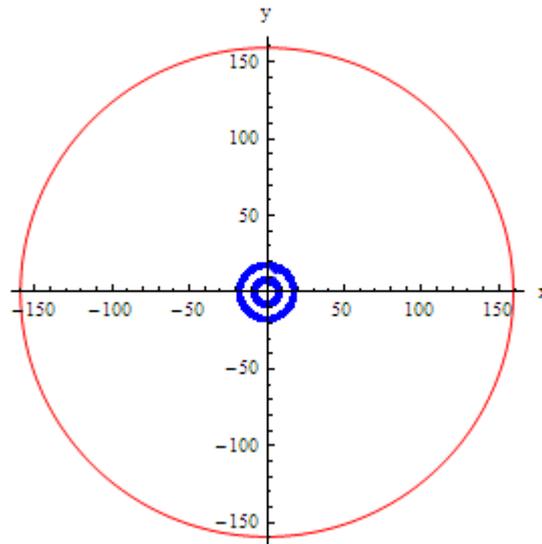
A mesma regra agora L muito grande:

Regra 20, $L = 1000$.



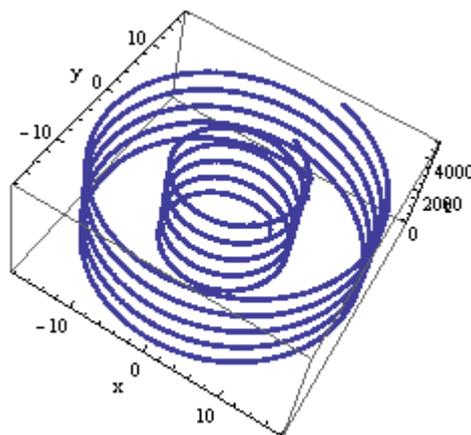
a)

A proporção entre o tamanho da região do centro de massa e o apótema do polígono aumenta com o aumento do tamanho do autômato com condição inicial aleatória.



b)

Evolução temporal com formação de estruturas.

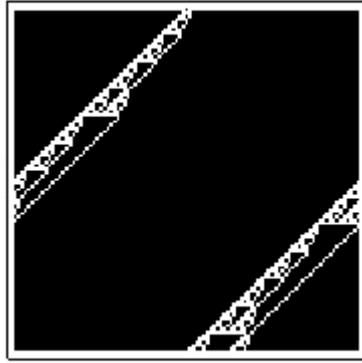


c)

Figura 58 - Evolução do Centro de Massa (Regra 20 e $L = 1000$). Idem anterior.

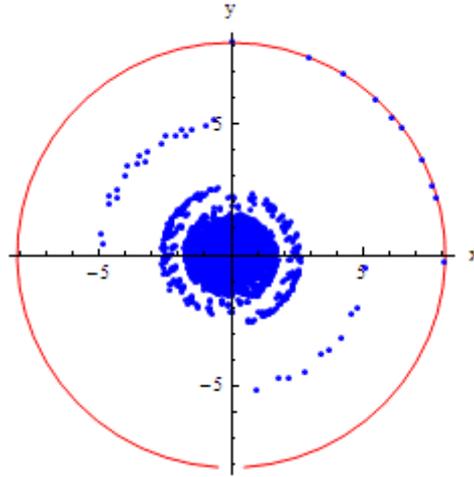
A regra 169 seria uma mescla de periódica com caótica e complexa:

Regra 169, $L = 51$, condição inicial 1 ponto no meio.



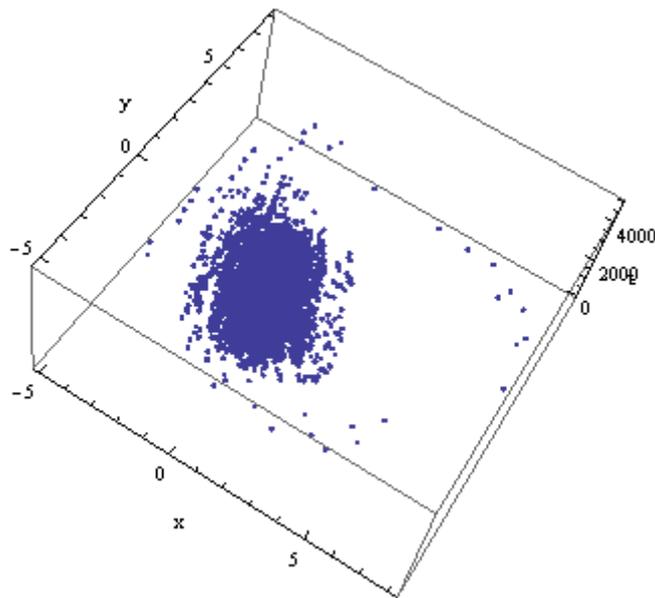
a)

A movimento do centro de massa é semelhante à periódica, mas de organização mais complicada.



b)

Distribuição quase aleatória dos pontos no espaço 3D. Não há a formação de estruturas, isso implica que não se apresentou estado estacionário.

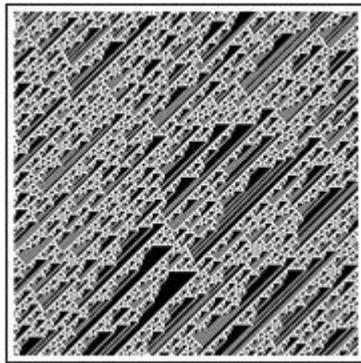


c)

Figura 59 - Evolução do Centro de Massa (Regra 169 e $L = 51$). Idem anterior.

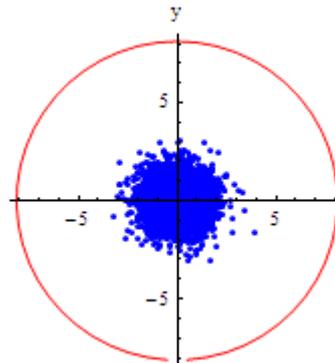
Com uma condição inicial aleatória:

Regra 169, $L = 51$,
condição inicial 1
aleatória.



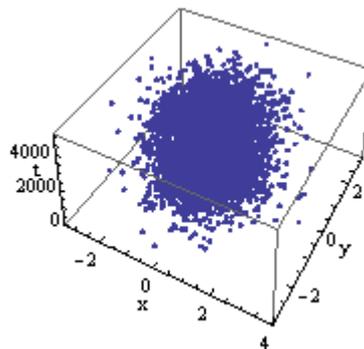
a)

O centro de massa é
aleatoriamente distribuído.



b)

E não há formação de
estruturas



c)

Figura 60 - Evolução do Centro de Massa (Regra 169 e $L = 51$). Idem anterior.

A movimentação dos centros de massa acontece de maneira tão variada quanto são os autômatos, mas o importante dos exemplos é notar dois comportamentos. O primeiro é que a regra age sobre as sequências movimentando o centro de massa, e é equivalente a um agente externo o alterando, é uma força motriz que o leva de uma posição a outra, e a complexidade desse movimento está ligada à complexidade dos diagramas.

O segundo ponto é que a distribuição temporal desses centros forma estruturas, colunas, circunferências ou espirais, e a presença de tais estruturas indica a ocorrência de estados estacionários.

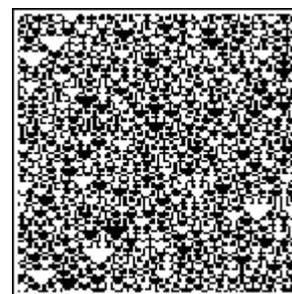
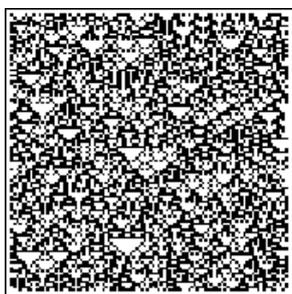
A movimentação do centro de massa dos autômatos é uma maneira gráfica de visualizar o surgimento de ciclos, a dinâmica da máquina, a forma com que a regra atua sobre as cadeias. As cadeias por si mesmas são estados, e desenhar a atuação da regra sobre cada vizinhança isoladamente seria inconcebível, sobretudo na infinitude de passos necessários para a ocorrência de movimentos periódicos. É um modo simples de enxergar a dinâmica e útil na percepção da complexidade dessas máquinas.

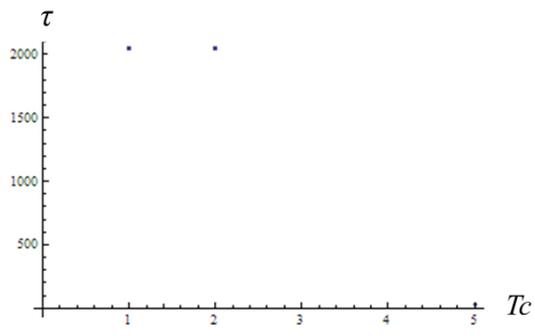
O que se disse até este momento sobre a complexidade das distribuições e sua relação com a complexidade dos arranjos de acordo com a classificação de Wolfram tem a ver com o modo com que a regra atua sobre as sequências em um autômato. O comportamento geral das classes menos complexas é convergir rapidamente ao estado estacionário, pelo menos mais rapidamente que as classes mais complexas. Isto impacta diretamente na variedade e duração dos períodos encontrados, bem como no próprio número de passos até aquela convergência.

Sendo assim, de acordo com a classificação de Wolfram, regras complexas devem exibir diagramas mais complicados, confundindo-se na própria definição das tais. Isto, evidentemente, sem citar as inúmeras situações intermediárias, figuras as quais exibem comportamento misto de toda sorte.

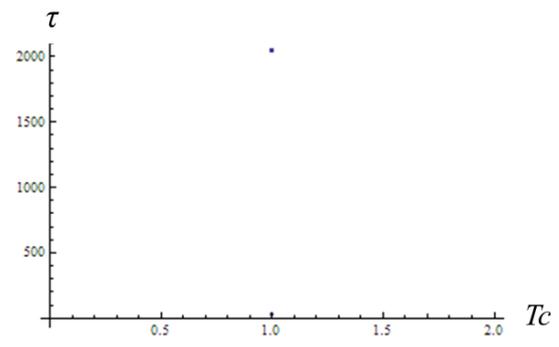
Contudo, a natureza algorítmica das regras, de sua construção, tem como consequência o surgimento de algumas regras ditas aditivas, colocado em II.3, quatro das quais são nulas e fixas (0, 170, 204 e 240), e duas são complexas (Classe III). A regra 90 e a regra 150 são elas. Não são simétricas entre si, portanto não podem ser consideradas equivalentes – ambas são pares –, mas se notar o conjunto de distribuições colocadas para os exemplos:

Classe III





Regra 90



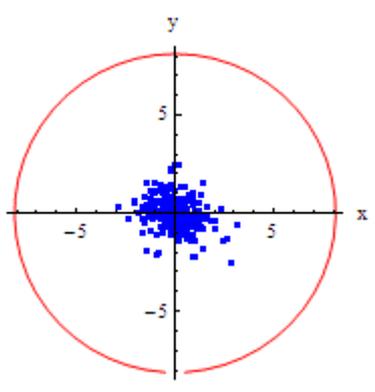
Regra 150

Figura 61 - Distribuição dos Períodos x Tempos Iniciais (Regras Aditivas). Períodos contra tempos mostrando a pequena variedade da distribuição nessas regras.

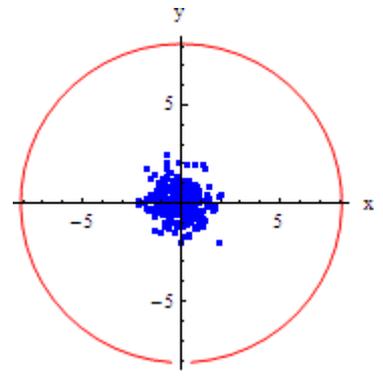
Concluir-se-á que a variedade de períodos τ é parca, em ambas, e igualmente em ambas o alcance dos instantes T_c é diminuto.

Essas regras complexas, diferentemente das outras, são fortes, convergem rapidamente ao estado estacionário:

Regra 90



Regra 150



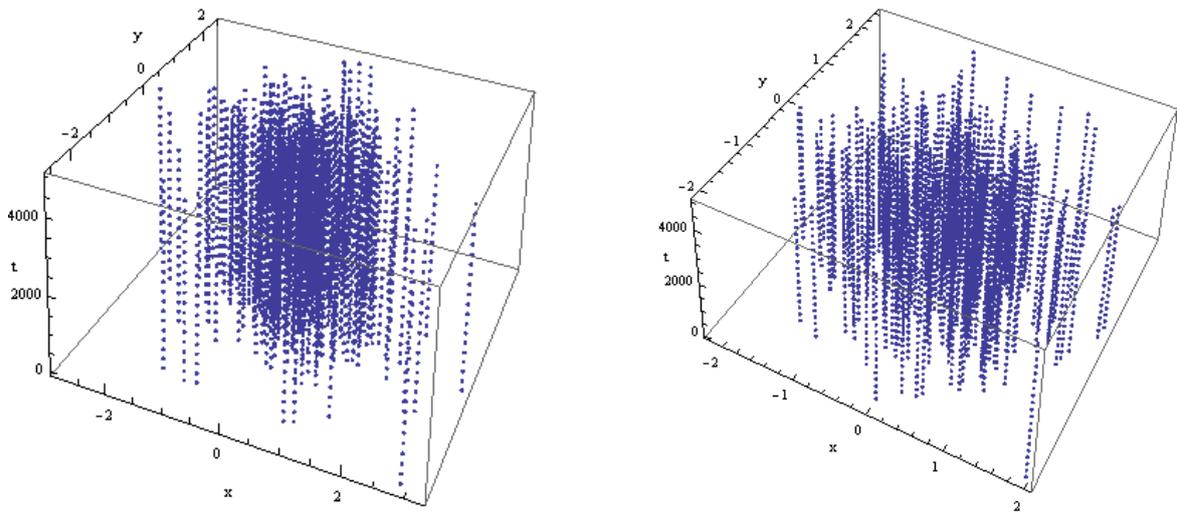


Figura 62 - Centros de Massa das Regras Aditivas. Distribuição e evolução dos centros de massa nessas regras.

Embora atuem complexamente sobre as cadeias. Contudo, tal atuação é notadamente organizada, o que é percebido pela formação de colunas no diagrama espaço-tempo do centro de massa.

Já foi dito aqui que a variância (ou o desvio padrão) dos períodos e instantes críticos devem estar para quantidades classificativas dos autômatos. Agora, com a constatação de que diagramas complexos podem exibir convergência rápida e ordenada, não obstante serem exceção, é razoável admitir que os valores esperados (aqui também médios) dos períodos τ e dos instantes críticos T_c dos autômatos também devem fornecer medida classificativa para os mesmos.

Um autômato ou uma regra parecem tão complexos quanto é a variedade dos τ e dos T_c , também assim é quão grandes são essas grandezas em média. Assim, uma rapidez de convergência F teria uma forma como:

$$F \propto \frac{1}{\sigma_\tau \sigma_{T_c} T_c}, \quad (\text{IV.b})$$

com:

$$\sigma_\tau^2 = \langle \tau^2 \rangle - \langle \tau \rangle^2$$

$$\sigma_{T_c}^2 = \langle T_c^2 \rangle - \langle T_c \rangle^2$$

Irrelevante, neste caso, se desvio padrão ou variância. Contudo, o conjunto de regras complexas aditivas diz que tal rapidez na convergência também precisa ser proporcional aos valores dos seus períodos. Uma regra, então, pode ser classificada de acordo com a atuação média sobre uma amostra suficientemente grande de condições iniciais aleatórias. Desta forma, é adequado escrever:

$$F \propto \left(\frac{\langle \tau \rangle}{\sigma_\tau \sigma_{T_c} \langle T_c \rangle} \right)^\delta, \quad (\text{IV.c})$$

na qual δ é um expoente de normalização adequado positivo e menor que a unidade.

Esta é uma conjectura e necessita refinamento, porém, é um indício da força de atuação de uma regra sobre as cadeias dentro de uma classe. Caso assuma valores tendentes ao infinito, expressará forte convergência, valores maiores ou iguais que à unidade estarão para diagramas estáveis com aspecto periódico, e quando maiores que zero e menores que a unidade, deverão representar as classes de lenta convergência e de complexa atuação sobre as sequências nelas encontradas.

Mas conjectura como conjectura auxilia apenas como norte, e neste caso particularmente se mostra como uma ilustração numérica para o comparativo entre as regras. Tratar a respeito da complexidade de atuação de uma regra de modo a classificá-las, como Wolfram fez com o aspectos diagramático, quer dizer matematizá-la e separar por grupos com características matemáticas semelhantes. E a complexidade adequada para este caso é a de Lempel-Ziv LZ76 por todas as razão que já foram discutidas anteriormente.

Mas a complexidade de Lempel-Ziv é aplicável somente a sequências de dados. Dado que uma fonte ergódica emite, calcular a complexidade de emissão quer dizer calcular a complexidade do ordenamento da emissão, e será a complexidade da máquina que emite. Contudo, um autômato celular não emite caracteres isoladamente, e calcular a complexidade LZ76 dos estados somente não faz sentido por dois motivos em particular. O primeiro é, obviamente, porque essas máquinas não são fontes dos caracteres isoladamente das cadeias que compõem os estados em cada instante. Autômatos Celulares Elementares se desenvolvem por regras síncronas, ou seja, atuam simultaneamente em todas as células do autômato a cada momento, e isso é feito da mesma forma para todas. Portanto, os estados completos é que são emitidos a cada passo.

O segundo motivo trata da natureza cíclica dos autômatos. O fato de serem máquinas circulares quer dizer necessariamente que nenhuma posição em particular tem preferência sobre as demais. A representação em diagramas, como é feita, pode levar a esse equívoco. E, se a complexidade de Lempel-Ziv necessita de um ponto de partida na sequência para começar a contagem, é natural que a escolha automática seja o início da cadeia no diagrama de estado-tempo. Porém, dentro de todo o conjunto de rotações disponíveis para o anel de células, não há nenhuma de maior primazia sobre as outras. De modo que estabelecer a complexidade de um autômato pela complexidade de suas sequências não é adequado, mesmo porque, rigorosamente, nem sequer sequências são, embora aqui se chame para mero título de denominação.

O que se faz frequentemente para contornar essa dificuldade é tratar sobre autômatos de tamanhos infinitos, de forma a fazer com que o conjunto de complexidades de todas as rotações disponíveis não difira muito entre si, isto quer dizer desprezar as bordas do diagrama. Outra maneira que é possível sugerir é utilizar a complexidade RLE nessa classificação. Esta complexidade tem a propriedade de ser invariante por rotações, contudo, é necessário estar atento ao escrever o algoritmo de cálculo para levar em conta a vizinhança das bordas no vetor.

A complexidade RLE, juntamente com uma entropia do tipo Shannon para as subcadeias geradas, pode ser grandeza de extrema utilidade para a estatística e geometria de autômatos celulares. Isto facilmente justificável pela evidência de que as subcadeias isoladamente dentro da sequência maior apresentam dimensão fractal nula (I.6.o), ergo são pontos isolados dentro da estrutura, portanto elementos de geometria distintos.

O segundo motivo, então, explica que os elementos das cadeias sozinhos não são expressão suficiente de sua complexidade, pela própria natureza do autômato. E o primeiro justifica não somente a mesma coisa, como também sugere como deve ser feito. A partir da constatação de que um autômato celular emita cadeias inteiras simultaneamente, portanto estados, ele é uma fonte egódica destes estados desde que um tempo suficientemente grande ou um número considerável de passos aconteça. Então se LZ76 medir a complexidade da série temporal de estados geral, ela medirá também a complexidade da máquina, e por consequência a complexidade da regra.

Por essa forma, diga que a regra 90, por exemplo, evolua de $t = 0$ até $t = 20$:

$$\{\{0,0,1,1,0,1,1,1,0,1\},$$

$\{1,1,1,1,0,1,0,1,0,0\}$,
 $\{1,0,0,1,0,0,0,0,1,1\}$,
 $\{1,1,1,0,1,0,0,1,1,0\}$,
 $\{1,0,1,0,0,1,1,1,1,0\}$,
 $\{0,0,0,1,1,1,0,0,1,0\}$,
 $\{0,0,1,1,0,1,1,1,0,1\}$,
 $\{1,1,1,1,0,1,0,1,0,0\}$,
 $\{1,0,0,1,0,0,0,0,1,1\}$,
 $\{1,1,1,0,1,0,0,1,1,0\}$,
 $\{1,0,1,0,0,1,1,1,1,0\}$,
 $\{0,0,0,1,1,1,0,0,1,0\}$,
 $\{0,0,1,1,0,1,1,1,0,1\}$,
 $\{1,1,1,1,0,1,0,1,0,0\}$,
 $\{1,0,0,1,0,0,0,0,1,1\}$,
 $\{1,1,1,0,1,0,0,1,1,0\}$,
 $\{1,0,1,0,0,1,1,1,1,0\}$,
 $\{0,0,0,1,1,1,0,0,1,0\}$,
 $\{0,0,1,1,0,1,1,1,0,1\}$,
 $\{1,1,1,1,0,1,0,1,0,0\}$,
 $\{1,0,0,1,0,0,0,0,1,1\}$

À primeira configuração chama-se estado 1, à segunda chama-se estado 2 e assim por diante.
Então:

$\{0,0,1,1,0,1,1,1,0,1\}$,	1
$\{1,1,1,1,0,1,0,1,0,0\}$,	2
$\{1,0,0,1,0,0,0,0,1,1\}$,	3

{1,1,1,0,1,0,0,1,1,0},	4
{1,0,1,0,0,1,1,1,1,0},	5
{0,0,0,1,1,1,0,0,1,0},	6
{0,0,1,1,0,1,1,1,0,1},	1
{1,1,1,1,0,1,0,1,0,0},	2
{1,0,0,1,0,0,0,0,1,1},	3
{1,1,1,0,1,0,0,1,1,0},	4
{1,0,1,0,0,1,1,1,1,0},	5
{0,0,0,1,1,1,0,0,1,0},	6
{0,0,1,1,0,1,1,1,0,1},	1
{1,1,1,1,0,1,0,1,0,0},	2
{1,0,0,1,0,0,0,0,1,1},	3
{1,1,1,0,1,0,0,1,1,0},	4
{1,0,1,0,0,1,1,1,1,0},	5
{0,0,0,1,1,1,0,0,1,0},	6
{0,0,1,1,0,1,1,1,0,1},	1
{1,1,1,1,0,1,0,1,0,0},	2
{1,0,0,1,0,0,0,0,1,1},	3
{1,1,1,0,1,0,0,1,1,0},	4
{1,0,1,0,0,1,1,1,1,0},	5
{0,0,0,1,1,1,0,0,1,0},	6
{0,0,1,1,0,1,1,1,0,1},	1
{1,1,1,1,0,1,0,1,0,0},	2
{1,0,0,1,0,0,0,0,1,1}	3

Os estados repetindo-se quando as configurações correspondentes retornam.

A complexidade do autômato, assim, será a complexidade da série temporal dos estados emitidos. Neste caso:

$$C_{\text{Autômato}} = C(123456123456123456123) = \{7, 1.2.3.4.5.6.123456123456123\}$$

Independente de quantos passos sejam emitir, a partir do estado estacionário a complexidade não aumentará pois as configurações geradas não mudarão, nem o seu

ordenamento. Então um tempo suficientemente grande de emissões deverá ser suficiente para caracterizar a complexidade de cada máquina sem que ela aumente, como tem de ser.

Contudo, a complexidade real do autômato, que é a regra aplicada à condição inicial, necessita da ocorrência de períodos, caso contrário ela será na ordem do tempo estabelecido, e portando a complexidade da série diagramática. Graficamente a evolução da máquina pode ser colocada:

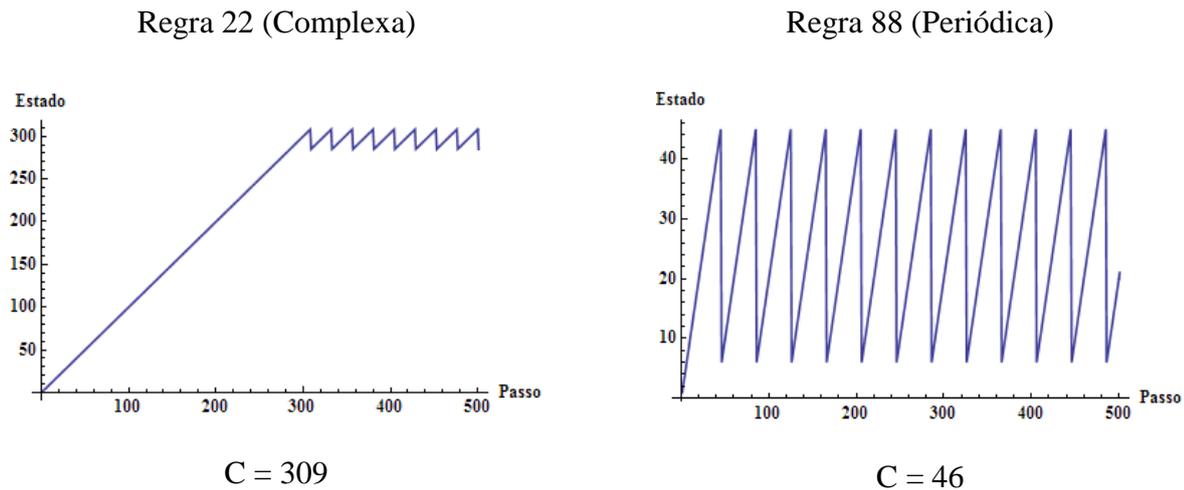


Figura 63 – Estado x Passo. Oscilações dos Autômatos. Essas oscilações representam a ocorrência de ciclos em determinado autômato, i. e., uma determinada regra aplicada em uma determinada condição inicial.

Para condições iniciais aleatórias e $L = 40$. Perceba-se que a ocorrência de oscilações nos gráficos representam o surgimento de ciclos, regiões estacionárias.

A complexidade de um autômato, desta forma, será uma quantidade que depende tanto do número de passos T_c até o ciclo como do tamanho τ deste ciclo, sendo igual ao somatório dos dois. De fato, se E_M representa o valor do maior estado alcançado:

$$C = T_c + \tau = E_M + 1 \tag{IV.d}$$

De acordo com os indícios anteriores, a complexidade de uma regra está relacionada ao número de passos que ela gasta para levar uma condição inicial até o estado estacionário, visto que é a complexidade de atuação do algoritmo que está para a complexidade da máquina. Isto impacta diretamente na organização dos diagramas. Estruturas e periodicidade podem aparecer mais ou menos explicitamente desde que oscilações nos estados ocorram.

Ao fim de tempo suficientemente longo, todas as regras convergem para os ciclos, independentemente do tamanho das sequências, posto que são construções determinísticas e atuarão sempre da mesma maneira sobre a mesma cadeia de dados. Isto entendido, fica claro que faz apenas sentido se falar sobre a complexidade de um autômato em particular, quando uma regra atua sobre uma condição definida de um tamanho definido. Porém, interessante aos estudos são os autômatos que convergem, de modo que cadeias pequenas comparadas ao tempo decorrido sejam melhor compreendidas no tocante à complexidade.

Nada impede, contudo, que se estude cadeias infinitas com recursos computacionais adequados. Mas, por enquanto, pode-se ter uma ideia da distribuição média das complexidades dos Autômatos Celulares Elementares de tamanhos menores:



Figura 64 - Distribuição das complexidades para todas as regras elementares.

A figura 64 mostra a distribuição das complexidades médias para todas as regras elementares. As complexidades, no caso, foram comparadas ao tamanho dos autômatos de 50 caracteres. O valor médio foi realizado para um conjunto de 100 condições iniciais pseudoaleatórias idênticas para todas as regras. É possível que se se usar outro comprimento de cadeia, o gráfico mude de aparência, visto que, de acordo com as distribuições nas Tabela 3 – Tabela 10, o tempo de convergência e o período do ciclo crescem de maneira diferente para regras diferentes.

Contudo, este foi um comprimento adequado e permitiu o surgimento de ciclos para a grande maioria dos autômatos. A região em verde do desenho é a região para a qual a razão $\langle C \rangle / 50$ assume valores até 0.15, a região em amarelo vai de 0.15 até 1.5, e a azul acima de 1.5. Na região verde se encontra a maioria das regras nulas e periódicas (185/190). Na parte azul do desenho estão todas as regras complexas e caóticas (35/35). Na região amarela estão as demais, a maioria dos autômatos com comportamento misto (30/31), complexo-periódico, caótico-periódico, e algumas nulas e periódicas.

De modo geral, a distribuição respeitou a classificação de Wolfram com algumas exceções, o que seria perfeitamente esperado. Isto porque a classificação de Wolfram, por sua própria generalidade, é insuficiente para descrever os autômatos que exibem comportamento misto de classes. É evidente que para obter informação mais precisa a respeito de como se distribuem as regras, seria necessário efetuar a avaliação para um alcance maior de comprimentos, a fim de se obter um comportamento. Mas a técnica fornece indícios de que a complexidade de um diagrama está para a complexidade da regra relacionada e sua aplicação à condição inicial dada.

A complexidade da máquina, então, neste caso, a complexidade de Lempel-Ziv da emissão de estados por um tempo suficientemente grande, implica uma descrição do tempo de convergência a um estado estacionário, bem como a duração deste estado, e isto reflete diretamente no surgimento de padrões nos diagramas dos autômatos, fornecendo uma quantificação da complexidade das regras, e dizendo que sua atuação é a responsável pelo surgimento ou não de estruturas nas figuras, coincidindo com o comportamento geral da classificação de Wolfram.

CONCLUSÕES

Cadeias de dados e Autômatos Celulares Elementares foram estudados principalmente no que concerne à sua complexidade, entretanto outras propriedades foram também analisadas.

Apoiando-se em um estudo feito Tél, Fülöp e Viscek sobre estruturas chamadas de multifractais geométricos [13], aplicou-se seu resultado às sequências de dados. Multifractais geométricos foram entendidos como sendo estruturas fractais constituídas por matérias de diferentes naturezas, ou cujo crescimento vai se alterando ao longo de sua composição. Com isso foi concluído que sua estatística seria adequada para aplicação em cadeias de dados, devido a sua possibilidade de se compor por elementos de ampla variedade. Cálculos consistentes com a dimensionalidade de Hausdorff para os conjuntos de Cantor foram obtidos para o paralelo com sequências de forma muito mais amistosa do que usualmente é feito para as séries temporais – como o descrito em [12], [21], [22], apesar de os métodos de [12] e [23] futuramente poderem se mostrar úteis para as máquinas de estado finito –, fornecendo uma estatística precisa no caso de sequências com aspecto não fractal.

Após isso, aquelas sequências de dados foram colocados no âmbito da complexidade de Lempel-Ziv LZ76 [6], um algoritmo capaz de quantificar o quão compressível pode ser uma cadeia, sobretudo infinita pela suposição de ergodicidade da sua fonte, e que fornece um fundamento para a construção das chamadas sequências de Máxima Complexidade de Lempel-Ziv (MLZs). Essas sequências foram descritas Estevez-Rams e colaboradores em [2], sendo providas de muitas propriedades, as quais foram objetos de estudo submetido para [2.1].

Dentre as propriedades pontuadas sobre as MLZs está o seu caráter não aleatório. Por existir um algoritmo capaz de gerá-las, um LZ76 reverso, elas são não randômicas por construção. Entretanto foi observado que elas não respeitam certos requisitos de aleatoriedade, como o teste de Martin Löf [20], para a incompressibilidade dos prefixos das *strings*, e a distribuição normal dos fatores (dada uma fatorização) no sentido da normalidade de Émile Borel, reforçando a premissa de determinismo. Assim, foi demonstrado que sequências com grande complexidade medida por um algoritmo específico podem apresentar pequena complexidade de Kolmogorov-Chaitin (KC), por haver um pequeno programa que as gera. Mas isto não contraria o senso de que sequências geradas por fontes ergódicas de cadeias aleatórias possuem máximas complexidades KC, já que elas devem parecer aleatórias para qualquer algoritmo que as tente reproduzir, portanto com máximas complexidades algorítmicas; não

contraria pois a recíproca não é verdadeira, haja vista que máximas complexidades medidas por algoritmos não necessariamente implicam máximas complexidades KC, mas sim o contrário; o que é mostrado nas MLZs.

Com isso foi sugerido um ansatz (III.d) que modela suas complexidades com base no número de símbolos de que são compostas, com o ajuste de curva adequado. Também, os fatores de LZ76 dessas MLZs foram analisados com a intenção de identificar o padrão de crescimento dessas cadeias, e outro ansatz experimental (III.e) foi inferido para esse crescimento como função da ordem da sequência gerada. Esse modelo é sensível às diversas bases numéricas, não somente às binárias, e permitem previsão do tamanho e da complexidade dessas estruturas com pequeno desvio percentual, como pode ser visto na Tabela 2.

As cadeias de dados foram mencionadas no texto como sendo a representação dos estados de um autômato celular. Apesar de essa não ser uma declaração precisa, entendido que os autômatos aqui estudados são de natureza circular, bordas de cujos estados são considerados em uma mesma vizinhança, a ilustração serve na visualização de seu funcionamento, sendo uma boa consideração, mesmo não rigorosa. E a princípio foi notado que os padrões gráficos gerados por essas máquinas, aqui Autômatos Celulares Elementares, teria a ver com a maneira com que determinada regra atuaria sobre as cadeias. Wolfram discriminou os diversos padrões gerados em quatro classes descritas no texto, cada qual com propriedades semelhantes.

Para compreender então a maneira como uma regra escolhida atua sobre as configurações celulares, primeiramente analisou-se o tempo de convergência de um autômato, ou seja, o número de passos que este leva para sair da condição de irreversibilidade, inicial, e encontrar uma sequência cuja evolução tornará a si mesma, passado um período de tempo, também analisado. E constatou-se que esses tempos de convergência T_c eram meramente dependentes do tamanho do autômato, ou o número de elementos dos estados, pois a forma de suas curvas de crescimento não se alterava quando colocadas em função da complexidade LZ76 e da complexidade RLE, podendo somente ser uma função média de seu tamanho. Também se comprovou as próprias formas das curvas eram diferentes quando vistas para diferentes regras em diferentes classes, não se sendo em primeiro momento passíveis de formulação matemática geral.

Contudo, os experimentos mostraram que o tempo de convergência T_c era tão grande quanto era complexo o padrão do arranjo relacionado, em média, bem como o tempo que a máquina levava para circular o estado e completar o ciclo estacionário, em média. Em média, pois o conjunto de regras aditivas da Classe 3, 90, 150 e derivadas, exibiam pequenos tempos de

convergência, contudo mostraram estado estacionário de grande período, o que se deve à sua construção aditiva por si mesma. De modo que foi sugerida uma grandeza que representasse a ação de uma regra sobre as sequências, que leva mais ou menos rapidamente uma condição inicial até o equilíbrio, ela sendo função daqueles tempos médios bem como de seus desvios padrões (IV.c); mas apenas como conjectura, necessita muito refinamento.

A hipótese de que as regras agem sobre as configurações, forçando-as à convergência, foi exposta escrevendo-se uma dinâmica específica para os autômatos, sempre em mente a evidência de que a complexidade dos arranjos classificados por Wolfram era função da complexidade de atuação da regra sobre os estados. Então colocou-se esses estados, essas combinações celulares, dentro da sua forma natural de construções circulares. Atribui-se a cada sequência com um certo número de caracteres um polígono regular com o mesmo número de vértices, fechando a máquina, e definiu-se que cada célula constituída por 1 era um vértice presente, e cada marcada com 0 era um vértice ausente. Assim, cada combinação de cadeia foi representada no plano cartesiano com uma distribuição de pontos marcados em posições correspondentes às posições dos caracteres nas sequências.

Essa distribuição de pontos definiu um sistema de partículas, cada qual com peso estatístico 1, e isto teve como consequência a definição do centro de massa de cada anel, cada polígono, cada configuração do autômato em cada instante de tempo. Destarte, a dinâmica de evolução do centro de massa foi visualizada para as diversas regras, já que cada estado teve seu centro particular, e constatou-se que de fato eles não permanecem estáticos, mudam de posição com o passar do tempo, implicando serem sistemas submetido à uma ação externa, justamente pela não conservação desses seus centros estatísticos. Foi concluído que essa ação é causada pela regra, e o surgimento de estruturas regulares identificado com a presença de oscilações dentro da condição estacionária, surgimento de ciclos reversíveis.

Assim, mais uma vez o trabalho se resumiu em compreender a complexidade de atuação de uma regra, visto que agente exterior ao sistema de caracteres. Foi então visualizada a evolução dos estados dos autômatos ao longo do tempo, colocando cada configuração celular distinta como sendo um estado distinto com valor dependente da sua primeira ocorrência na evolução da máquina. Após isso, construída uma série temporal com os valores dos estados, aplicou-se o método de Lempel-Ziv LZ76 para verificar a complexidade do arranjo.

Foi entendido isso ser razoável pois, se por um lado a LZ76 mede a complexidade de uma cadeia de dados, por outro lado ela estará medindo com isto a complexidade da máquina que gera

essa saída em regime ergódico. E, desde que os elementos emitidos por um autômato são os estados em cada instante de tempo, a complexidade LZ76 calculada sobre essa série temporal de emissão, ergódica, deve caracterizar a complexidade da regra que gera aquelas sequências; e não a complexidade das sequências por si, posto que todos seus elementos são emitidos de maneira síncrona em cada passo. A complexidade LZ76 sobre a série temporal de estados foi então calculada para autômatos suficientemente pequenos para permitir a ocorrência de ciclos, porque, como os ciclos se repetem ao longo do tempo, a complexidade não deverá se alterar a partir de então. E se ciclos não ocorrem, então a complexidade será tão grande quanto o próprio tempo do autômato.

Uma distribuição dessas complexidades médias foi colocada na Figura 64 para diversas condições iniciais aleatórias. E com isso pôde-se notar que de fato a complexidade dos arranjos está relacionada à complexidade de atuação da regra, aqui quantificada por sua LZ76, e as diferentes classes apresentadas por Wolfram tendem a corresponder à discriminação numérica apresentada.

Isto cumpre o objetivo de trabalho de estabelecer uma classificação quantitativa para os autômatos celulares elementares, à luz da complexidade Lempel-Ziv, mostrando que a complexidade dos arranjos gerados pelos autômatos está para a complexidade de atuação de uma regra, a qual nada é diferente de quão difícil é classificar a emissão dos estados pelas mesmas, entendida sua atuação dinâmica sobre as configurações, sempre convergindo de maneira determinística as máquinas, mas cada qual à sua construção.

Diversas formas de diferenciar numericamente as regras podem ser encontradas, por exemplo a distância entre os vértices do hipercubo gerado pelo espaço das 256 regras, a qual é a distância de Hamming entre as regras por si [24], ou também a entropia da randomização booleana de Hütt e Marr [25], ou mesmo a própria classificação de Wolfram para os Autômatos Celulares Elementares. Mas esta é uma abordagem diferente, provavelmente não melhor, provavelmente não mais útil, construída sobre os alicerces da Complexidade Algorítmica, mas cujas bases podem indicar aplicação aos sistemas físicos reais – algo como a sugestão de autômatos naturais, já extensamente estudados pela comunidade – como o rascunhado no anexo que se segue. Tal conceito pode vir a ser a útil, embora fuja um pouco da elementaridade tratada neste trabalho, e se aproxima um pouco das novas máquinas abstratas emergindo da pesquisa moderna.

Compreendidas os resultados, e que estudos teóricos sobre a cinética de transição de fases e ordenamento de fases, usando modelos de dinâmicas estocásticas unidimensionais em $T = 0K$, tem sido de interesse recente, e ainda que há uma crescente compreensão de que vários problemas tridimensionais podem ser efetivamente modelados em termos de sistemas 1D [42], este trabalho representará uma abordagem diferente ao estudo de defeitos estruturais em sólidos cristalograficamente desafiadores. O uso da teoria da Informação para tal finalidade é, até onde se sabe, inédito. Este trabalho permite começar a discussão sobre a aplicação desta metodologia para a investigação dos mesmos.

REFERÊNCIAS

- [1] COVER, Thomas M.; THOMAS, Joy A. **Elements of Information Theory**. New Jersey: Wiley, 2006. 2nd ed. 748 p.
- [2] RAMS, E. E. et al. On the non-randomness of maximum Lempel Ziv complexity sequences of finite size. **Chaos**, v. 23, 02318, 2013.
- [2.1] Artigo submetido: “*Properties of maximum Lempel-Ziv complexity strings*”. C. A. J. Nunes, E. Estevez-Rams, R. Lora-Serrano and B. Aragón Fernández. Submetido a Journal of Physics A: Mathematical and Theoretical
- [3] ARORA, S.; BARAK, B. **Computational Complexity: a modern approach**. New York: Cambridge University, 2009. 546 p.
- [4] TETURO, K. On Kolmogorov’s complexity and information. **Osaka Journal of Mathematics**, v. 10, p. 305-307, 1973.
- [5] KOLMOGOROV, A. N. Three approaches to the concept of the amount of information, **Problems of Information Transmission**, v. 1, p. 1–7 (1965).
- [5] KOLMOGOROV, A. N. On Tables of Random Numbers. **The Indian Journal of Statistics**, v. 25, n. 4, p. 369-373 (1963).
- [6] LEMPEL, A.; ZIV, J. On the Complexity of Finite Sequences. **IEEE Transactions on Information Theory**, v. IT-22, n. 1, p. 75-81 (1976).
- [7] LEMPEL, A.; ZIV, J. Compression of Individual Sequences via Variable-Rate Coding. **IEEE Transactions on Information Theory**, v. IT-24, n. 5, p. 530-536 (1978).
- [8] SALOMON, D. **Data Compression: the complete reference**. New York: Springer, 2000. 2nd ed. 898 p.
- [9] GOLDSTEIN, Herbert.; POOLE, Charles; SAFKO, John. **Classical Mechanics**. San Francisco: Addison-Wesley, 2002.
- [10] MACEDO, Maysa et al. **Cálculo da Dimensão Fractal: Método de BoxCounting**. Disponível em: <<http://www.cbpf.br/~maysagm/>>.
- [11] BACKES, A. R.; BRUNO, O. M. **Técnicas da Estimativa da Dimensão Fractal: Um Estudo Comparativo**. USP. 2005.

- [12] THEILER, J. Estimating the Fractal Dimension of Chaotic Time Series. **The Lincoln Laboratory Journal**, v. 3, n. 1, p. 63-86. 1990.
- [13] TÉL, T.; FÜLÖP A.; VICSEK T. Determination of Fractal Dimension for Geometrical Multifractals. **Physica A**, n. 159, p. 155-166. 1989.
- [14] PICKOVER, C. A. **The Math Book: From Pythagoras to the 57th Dimension, 250 Milestones in the History of Mathematics**. Sterling. 528 p. 2009.
- [15] GARDNER, M. Mathematical Games: the fantastic combinations of John Conway's new solitaire game "life". **Scientific American**, v. 223, p. 120-123. 1970.
- [16] WOLFRAM, S. Statistical Mechanics of Cellular Automata. **Reviews of Modern Physics**, v. 55, n. 3, p. 600-644. 1983.
- [17] KUNKLE, D. R. **Automatic Classification of One-Dimensional Cellular Automata**. Master of Science Thesis. Rochester Institute of Technology. 2003.
- [18] WOLFRAM, S. **A New Kind of Science**. Champaign: Wolfram Media, p. 952-953. 2002.
- [19] ROWLAND, T.; WEISSTEIN, E. W. **Additive Cellular Automaton**. Disponível em: <<http://mathworld.wolfram.com/AdditiveCellularAutomaton.html>>
- [20] MARTIN-LÖF, Per. On the definition of random sequences. **Information and Control**, n. 9, p. 602-619. 1966.
- [21] LIAW, S. S.; CHIU F. Y. Fractal Dimensions of Time Sequences. **Physica A**, n. 388, p. 3100-3106. 2009.
- [22] EMBRECHTS, M. J.; DANON Y. **Determining the Fractal Dimension of a Time Series with a Neural Net**. New York: ASME, 1993, v. 2, p. 897-902.
- [23] LI, M. Fractal Time Series: a tutorial review. **Mathematical Problems in Engineering**, v. 2010, 26 p. 2009.
- [24] LI, W; PACKARD, N. H.; LANGTON, C. G. Transition Phenomena in Cellular Automata Rule Space. **Physica D**, v. 45, p. 77-94. 1990.
- [25] MARR, C.; HÜTT, M. T. Outer-totalistic cellular automata on graphs. **Physics Letters A**. v. 373, p. 546-549. 2009.
- [26] J. P. Crutchfield. *Nature Physics*, 8:17–24, 2012.
- [27] Adami C. *BioEssay*, (24):1085–1094, 2002.

- [28] D. P. Varn, G. S. Canright, and J. P. Crutchfield. *Phys. Rev. B*, 66:174110–174113, 2002.
- [29] W. M. Goncalves, R. D. Pinto, J. C. Sartorelli, and M. J. Oliveira. *Physica A*, 257:385–389, 1998.
- [30] R. W. Clarke, M. P. Freeman, and N. M. Watkins. *Phys. Rev. E*, 67:160–203, 2003.
- [31] C. E. A. Shannon. *Bell Sys. Tech.*, 27:379–423, 1949.
- [32] D. Schechtman, I. Blech, D. Gratias, and J. W. Cahn. *Phys. Rev. Lett.*, 53:1951-1953, 1984.
- [33] Palosz, B., Grzanka, E., Gierlotka, S., Stelmakh, S. (2010), *Z. Kristallogr.* 225. pp 588–598.
- [34] Amigó, J. M. et al. (2003). *BioSystems* 68, pp. 57–66.
- [35] D. Abósolo, M. Aboy and D. Alvarez (2006). *IEEE Trans. Biomed. Eng.* 53, pp. 2282–2288.
- [36] **The basics of crystallography and diffraction**, Christopher Hammond, Second edition, Oxford science publishers, IUCr, pag. 28;
<<http://www.iue.tuwien.ac.at/phd/ayalew/node20.html>>.
- [37] Haag G. (1943) *Arkiv. fur Kemi. Mineralogi och Geologi* 16B, 1-6.
- [38] Crutchfield, J. P. and K Young (1989). *Phys. Rev. Lett.* 63, pp. 105–108.
- [39] Crutchfield, J. P. and D. P. Feldman (1997). *Phys. Rev. E* 55, R1239–R1242 (2003); *Chaos* 13, pp. 25–54.
- [40] Zhang, Y. et al. (2009). *J. Math. Chem.* 46, pp. 1203–1212.
- [41] J. Ziv, “Coding theorems for individual sequences,” *IEEE Trans. Inf. Theory* 24, 405–412 (1978).
- [42] S. Ahmad and M. A. Khan. *Stabilities of Modulated Phases of Close Packed Materials.* *Phys. Stat. Sol.(b)* 218, 425 (2000).

ANEXO

ANEXO A

Uma das questões que eventualmente podem surgir com respeito à dinâmica dos Autômatos Celulares Elementares é sobre quão próximos de um sistema natural eles estão, ou, ainda, quais autômatos evoluem mais ou evoluem menos semelhantemente a um sistema natural.

Já foi descrito no tópico sobre a Entropia (I.4) como deve ser a termodinâmica de um sistema natural, o qual, dada uma condição de reversibilidade, a entropia tende a aumentar com o passar do tempo, sendo máxima para a condição de equilíbrio. Equivalente dizer que a energia interna deve ser mínima pela simetria do princípio variacional sobre as equações termodinâmicas.

Com efeito, se qualquer dentre aquelas máquinas obedecer, por exemplo, ao princípio da máxima entropia para o estado de equilíbrio, essa deve ser mais dinamicamente próxima de um sistema físico natural do que as que não obedecem. Contudo, se tal suposição não encontra muitos impedimentos argumentativos, a proposta de escrever a entropia para um sistema abstrato que corresponda à entropia de Boltzmann já é mais obscura.

Alguém poderia, de fato, dizer que a entropia de Shannon é uma boa candidata àquela correspondência, haja vista que carrega o conteúdo de informação de uma distribuição de caracteres, além do fato de, por uma mera transformação de base, ser capaz de gerar todo o conjunto de entropias em todos os ensembles estatísticos. Isto funciona. Entretanto já foi pontuado neste mesmo texto a incapacidade de uma mera entropia de Shannon descrever a informação contida no ordenamento dos símbolos, pois trata apenas da probabilidade de ocorrência dos mesmos, nunca de sua disposição ao longo das cadeias, e esta, em termos de organização termodinâmica, é mais importante que meramente a probabilidade ocorrência, entendida a não materialidade dos constituintes das cadeias de dados.

Também neste texto foi descrito que existe uma classe de entropias, as entropia de bloco, sensíveis ao ordenamento particular de cada sequência, carregando, desta forma, informação sobre a geometria das sequências e conseqüentemente sobre o seu estado de organização, se fazendo, portanto, muito adequadas na quantificação de quão complexo pode ser um arranjo de caracteres qualquer. Suponha-se, assim, que o estado de organização dos elementos em uma sequência esteja para o estado de organização de um sistema físico, e que sequências mais organizadas devam apresentar menor entropia termodinâmica, pois o aumento da entropia é

intimamente relacionado ao aumento da complexidade de organização de um sistema natural. Suponha-se isto.

Suponha-se também que os blocos de caracteres utilizados na entropia de bloco, que é do tipo Shannon, são gerados pela codificação RLE, descrita em (I.5); blocos razoáveis, que não misturam caracteres em uma disposição arbitrariamente escolhida. De forma que uma sequência como:

$$\begin{array}{c} 100101001111 \\ \downarrow \\ (1)(00)(1)(0)(1)(00)(1111) \end{array}$$

É levada a esta fatorização.

A escolha dos blocos por esta forma é justificável pelo seguinte. No tópico sobre a dimensão fractal de sequências de dados (I.6), mostrou-se que uma subsequência de elementos corridos, idênticos em uma vizinhança, possui dimensão fractal nula quando comparada à estrutura como um todo. Isto quer dizer que os blocos de elementos idênticos em uma cadeia de dados são enxergados como não tendo extensão quando dentro de uma estrutura maior com caracteres distintos. Fazendo com que os fatores distintos de RLE de uma sequência sejam pontos distintos, inclusive geometricamente, dentro da cadeia, tornando esse tipo de codificação, além de sensível à forma, adequada para a distribuição dimensional da mesma.

Blocos de RLE são, por este motivo, justificados por sua estrutura fractal. Então, aquela sequência acima colocada é equivalentemente escrita como:

$$\begin{array}{c} 100101001111 \\ \downarrow \\ (1)(00)(1)(0)(1)(00)(1111) \\ \downarrow \\ ABACABD \end{array}$$

E é sobre essa nova escrita que a entropia deve ser calculada, por exemplo, na base natural:

$$H = -\sum_j p_j \log p_j \quad (i)$$

Em que p_j é a probabilidade de ocorrência do elemento j na cadeia.

Quando se fala sobre autômatos celulares elementares, é necessária atenção ao fato de que os elementos presentes nas bordas das sequências são elementos vizinhos, ou seja, as sequências não são sequências propriamente ditas, mas configurações circulares; e isto já foi bastante descrito anteriormente. Então, se a sequência posta acima, por exemplo, for um estado de um autômato, a transformação mais adequada e correta tornaria:

$$\begin{array}{c}
 \underline{100101001111} \\
 \downarrow \\
 (1111)(00)(1)(0)(1)(00)(\dots) \\
 \downarrow \\
 ABCDCB(\dots)
 \end{array}$$

O que mostra como a codificação RLE pode ser invariante sobre rotações. Para qualquer rotação daquela cadeia, a fatorização será a mesma. Reforçando sua importância no tratamento de autômatos celulares elementares, que são máquinas circulares.

A despeito de toda essa consideração sobre que blocos de entropia mais adequados para a termodinâmica de autômatos, um questionamento pode ocorrer. Se uma configuração binária é levada a uma configuração em outro espaço de base, por conta da geometria de seus fatores, a própria entropia do tipo Shannon calculada também não deve sofrer uma transformação de base adequada à nova estruturação? Equivalente perguntar: é conveniente tratar um conjunto de elementos escritos em uma base numérica em termos de outra base numérica? Ou, ainda: é legítimo normalizar a entropia pelo novo alfabeto gerado?

Tente-se responder a isso pela experimentação, mas, primeiramente, recordando sobre as propriedades entrópicas que cada sistema deve exibir.

Um sistema que, dado um estado inicial de organização, é levado a um estado de organização mais complexo, no equilíbrio, deverá ter a sua entropia incrementada relativamente à condição inicial. No caso de a situação de equilíbrio apresentar configuração mais simples que a região inicial, então é fundado que houve um trabalho maior sobre o sistema e sua entropia é assim decrementada.

Atentando ao fato de a nova base possivelmente descrita ser o número de elementos distintos que compõe a sequência transformada:

$$\begin{aligned}
 &100101001111 \mapsto b = 2 \\
 &\quad \downarrow \\
 &(1111)(00)(1)(0)(1)(00) \\
 &\quad \downarrow \\
 &ABCDCB \mapsto b = 4
 \end{aligned}$$

A entropia base dependente H_b seria, apenas:

$$H_b = -\sum_j p_j \log_b p_j = -\sum_j \frac{p_j \log p_j}{\log b} = \frac{H}{\log b} \quad (\text{ii})$$

Para os j fatores de RLE. Deste modo, pode-se desenhar a evolução temporal da Entropia de Shannon calculada para os elementos geométricos isolados para uma regra dada:

Regra 88

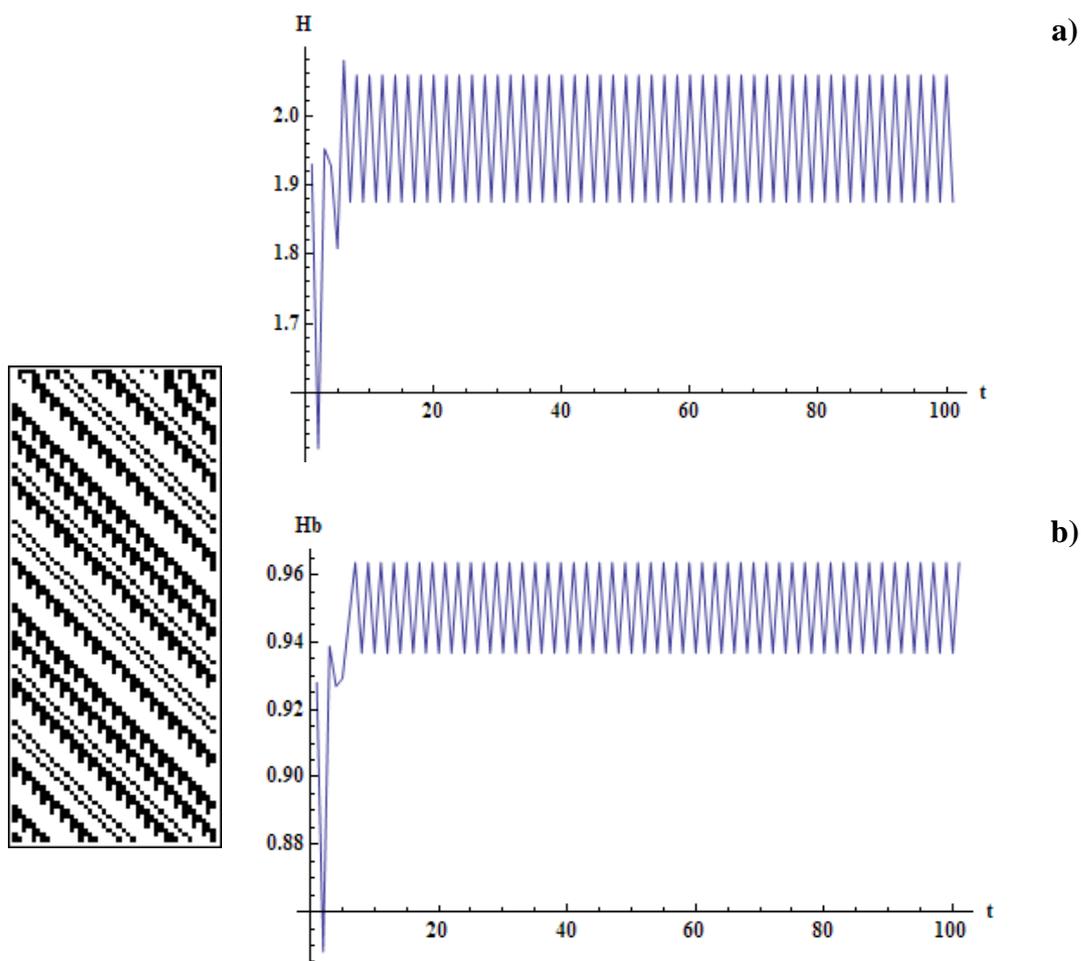


Figura 65 – Entropia de Bloco RLE

Base dependente H_b e base natural H . Entretanto, quando se diz sobre a entropia do estado de equilíbrio, passando um longo tempo, é mais interessante considerar a média temporal sobre as flutuações do que a grandeza em si, pela própria natureza macroscópica. Isto é, estatisticamente, o valor esperado da entropia naquela condição, e este valor esperado que deve representar a tendência variacional descrita. Estes mesmos gráficos tornariam, então:

Regra 88

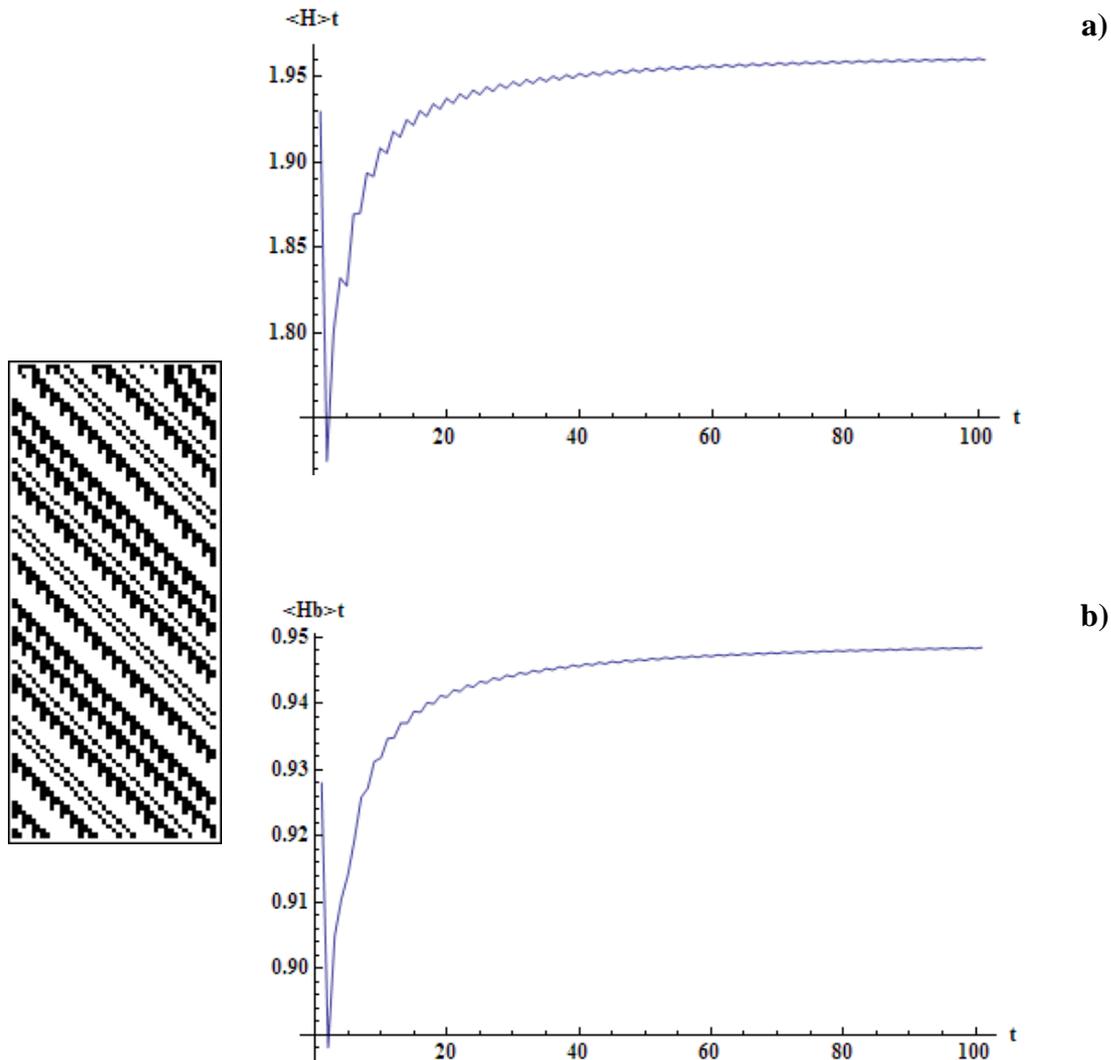


Figura 66 – Média temporal da Entropia de Bloco RLE

Representando, para cada t , o valor médio da entropia até aquele instante. Isto é o valor esperado para cada instante. E desta forma para qualquer autômato que se deseje.

Para um efeito estatístico mais adequado, é interessante tomar uma série de condições iniciais aleatórias, e para todas desenhar uma curva média. Aqui se utilizou autômatos de

tamanho igual a 43 correndo sobre 5000 passos, e 100 condições iniciais aleatórias foram colocadas todas iguais para todas as regras. Por exemplo, a curva média da regra **22** tem a forma:

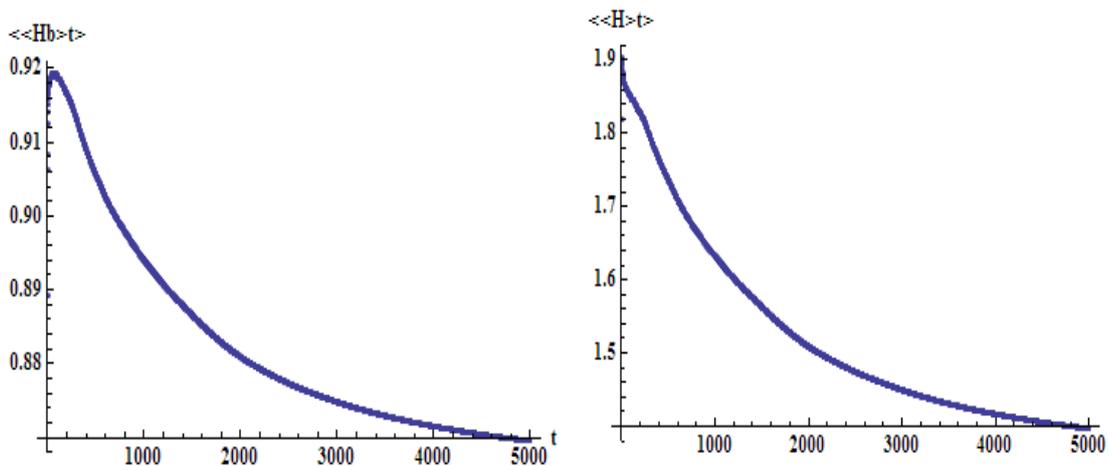


Figura 67 – Entropia Média para diversas condições iniciais aleatórias

Para ambas as entropias.

Dir-se-á um que uma regra é dinamicamente mais próxima do natural caso, passando um tempo muito longo para a condição de equilíbrio, o valor esperado da entropia seja maior do que o valor esperado da mesma no instante inicial. Ou:

$$\Delta h = \langle H \rangle_{t \rightarrow \infty} - \langle H \rangle_{t=0} > 0 \quad (\text{iii})$$

Caso a condição de equilíbrio seja um estado de organização mais complexa que o instante inicial, e a entropia que corresponder a isto será a fisicamente relacionada aos autômatos.

As condições supracitadas de condições iniciais e tamanho foram calculadas para todas as regras, e os valores de Δh verificados. Primeiramente para o caso base dependente:

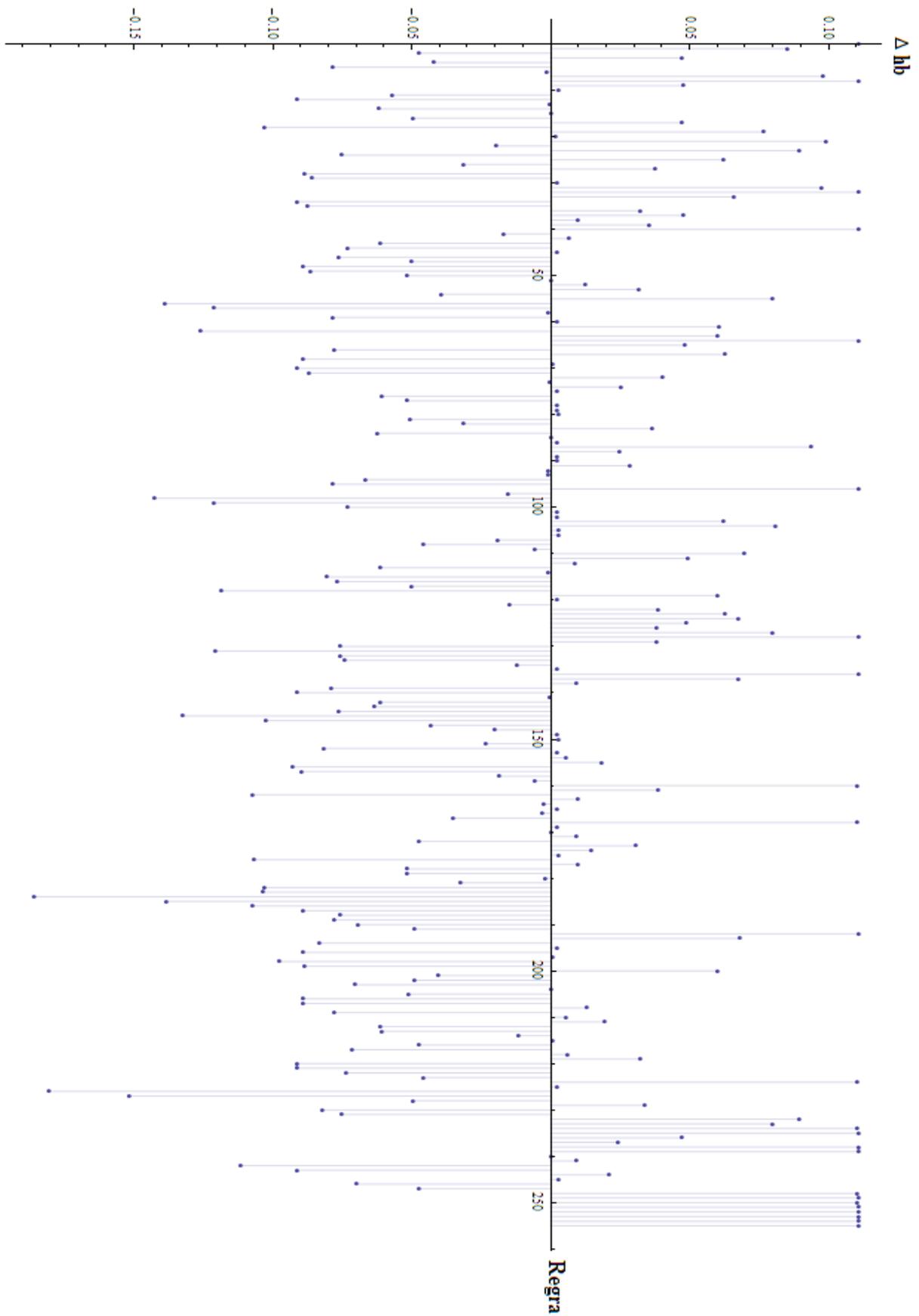


Figura 68 – Distribuição das entropias pelas regras (normalizadas)

E para a base natural:

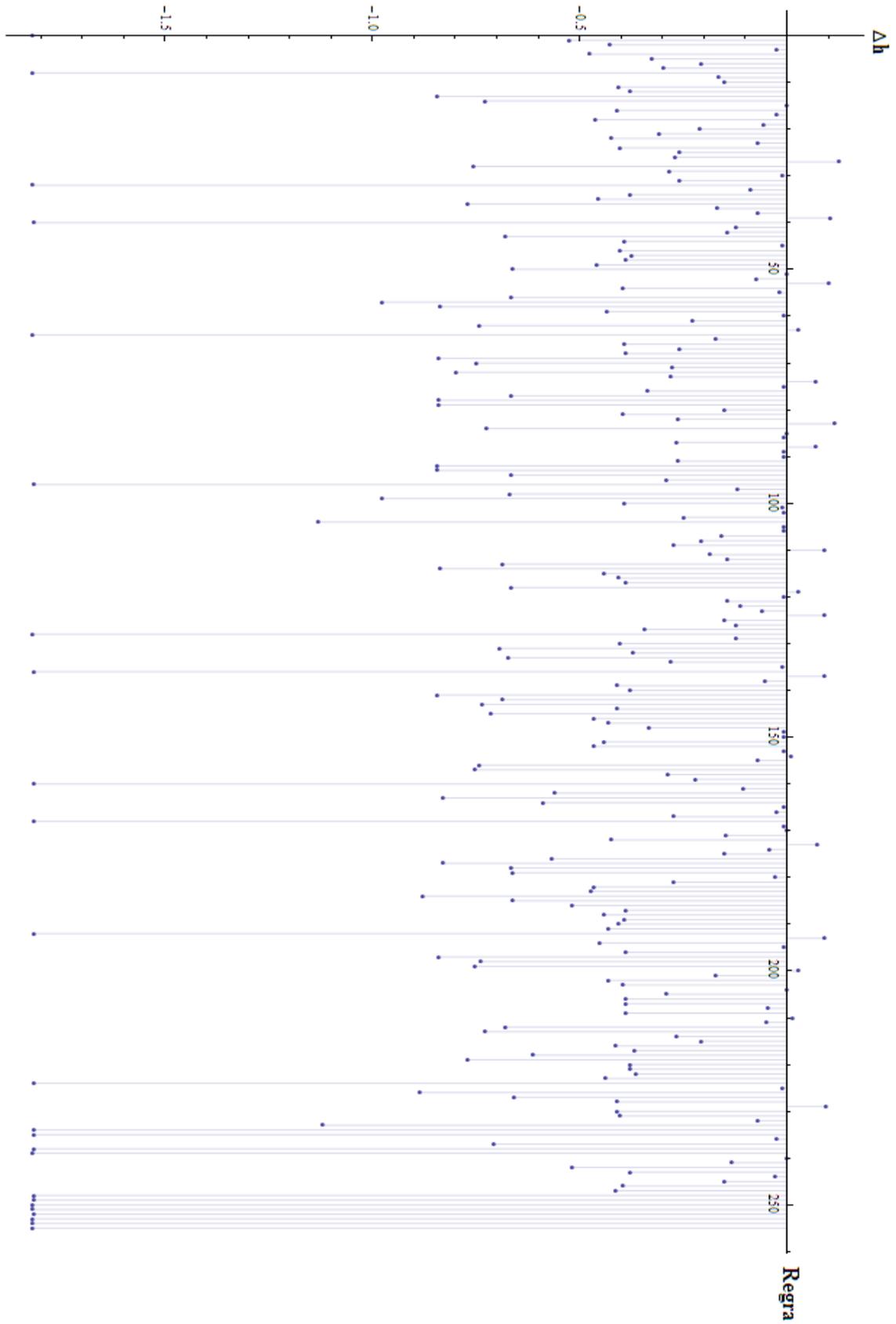
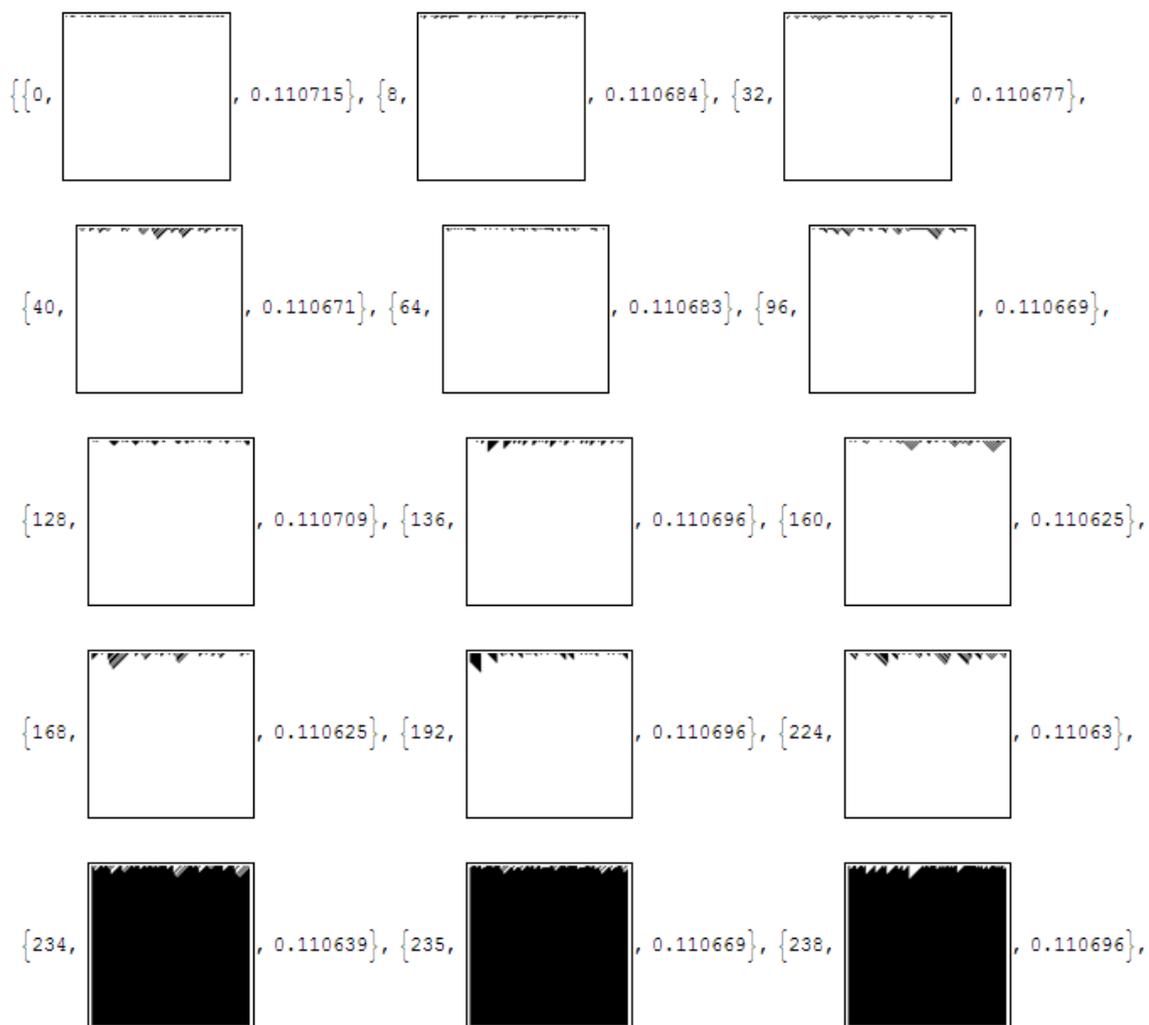
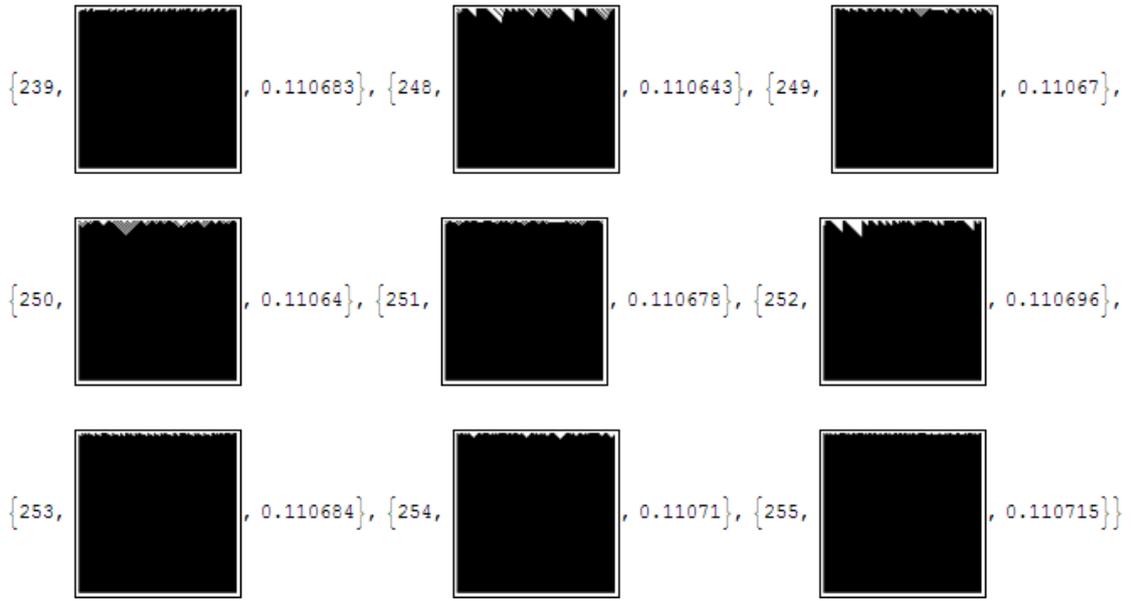


Figura 69 – Distribuição das entropias pelas regras

O caso base dependente apresentou uma distribuição mais homogênea entre os valores positivos e negativos de Δh , enquanto para a base natural a quantidade de autômatos fisicamente adequados é bastante exígua. Mas isto não responde a pergunta: qual das entropias é naturalmente mais correspondente para as cadeias de dados? Para responder a isto, primeiro selecione-se um conjunto de autômatos representativos, que expressem indubitavelmente o que é um estado de equilíbrio complexo e o que é um estado de equilíbrio simples e verifique-se suas entropias. Outra opção pode ser selecionar um conjunto de entropias expressivas e analisar os autômatos correspondentes. Escolha-se a segunda opção.

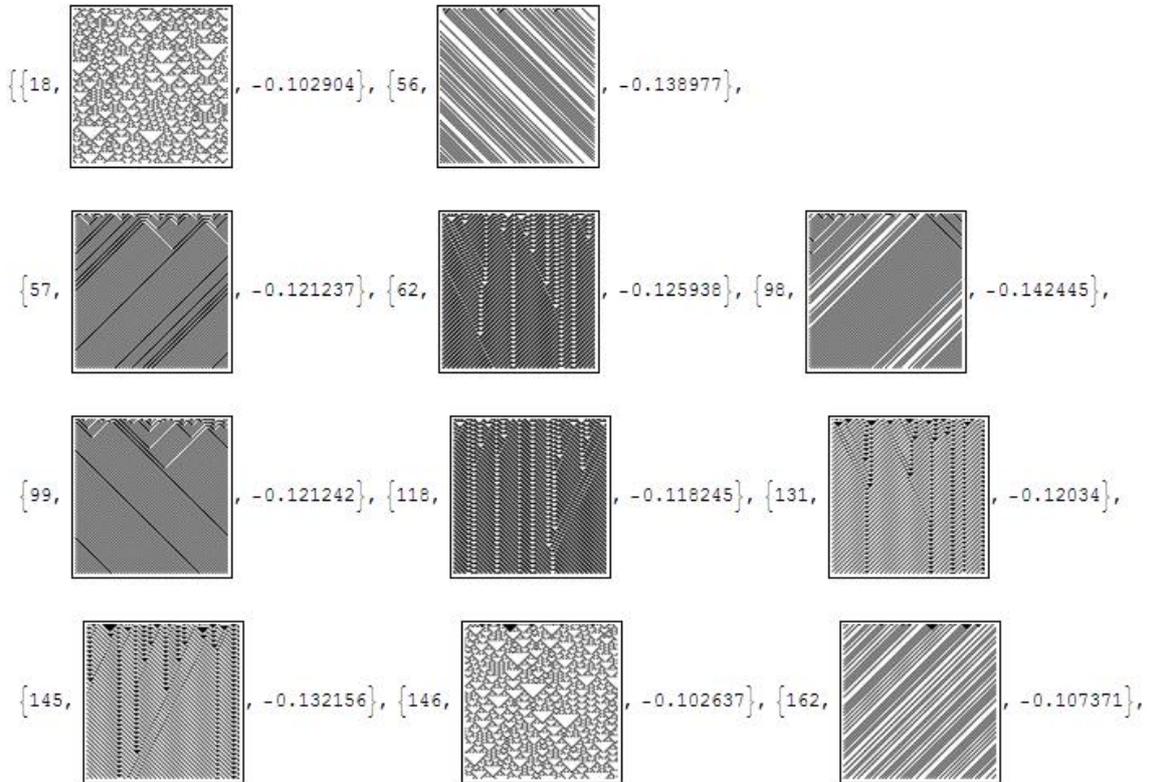
Diga-se que valores expressivos de Δh para o caso base dependente sejam valores maiores que 0,1 e valores menores que -0,1, pela simetria e escassez de regras nessas regiões. Diga-se, também, que para o caso da base natural, as regras selecionadas sejam correspondentes aos valores de Δh maiores que 0 (zero) e menores -1,5. Módulos maiores deverão exprimir mais fortemente as características termodinâmicas desejadas. Desta forma, o conjunto de autômatos selecionados se apresenta na forma {regra, arranjo, Δh }. Para base dependente:

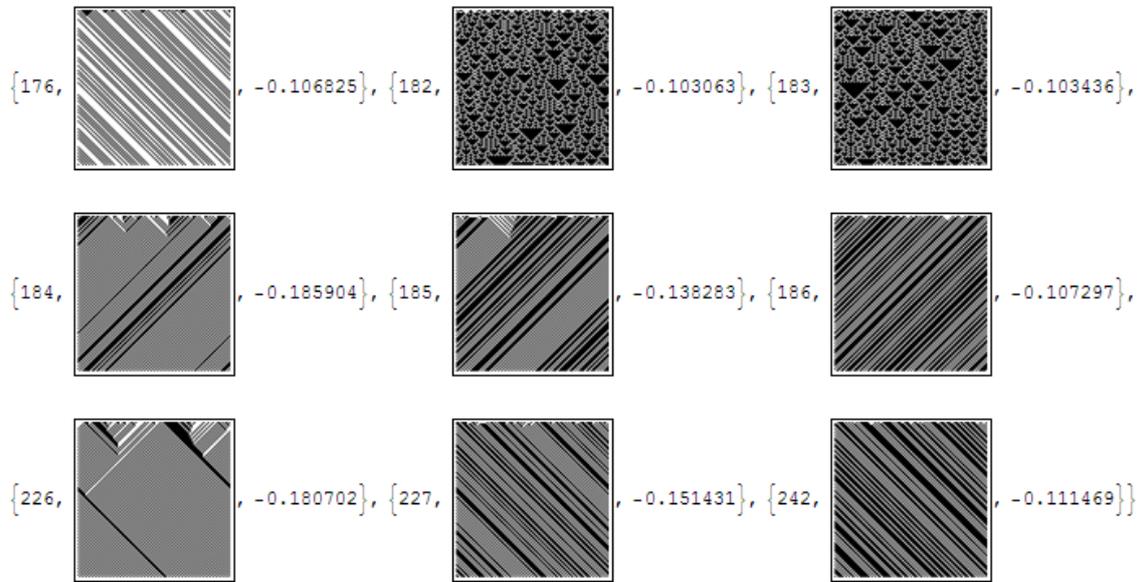




$\Delta h_b > 0$

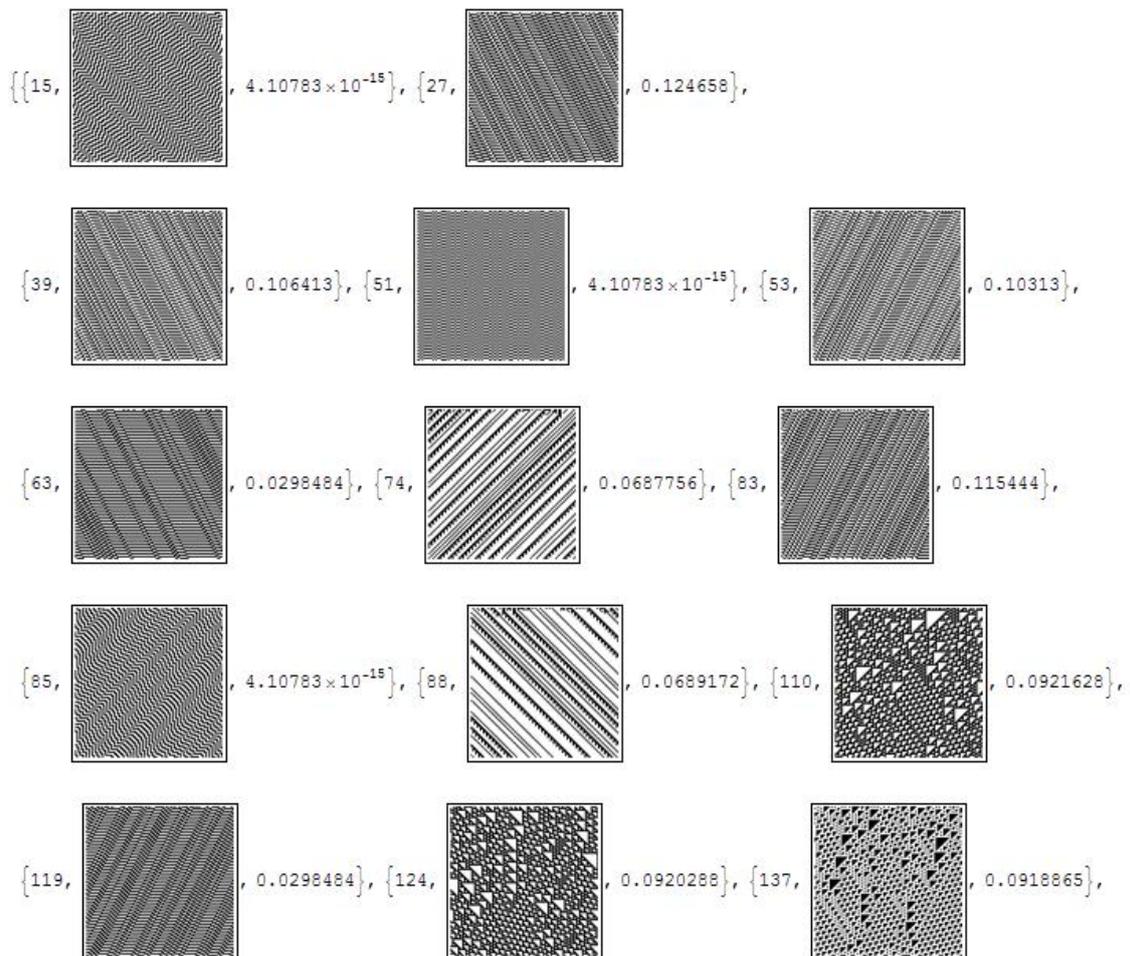
E:

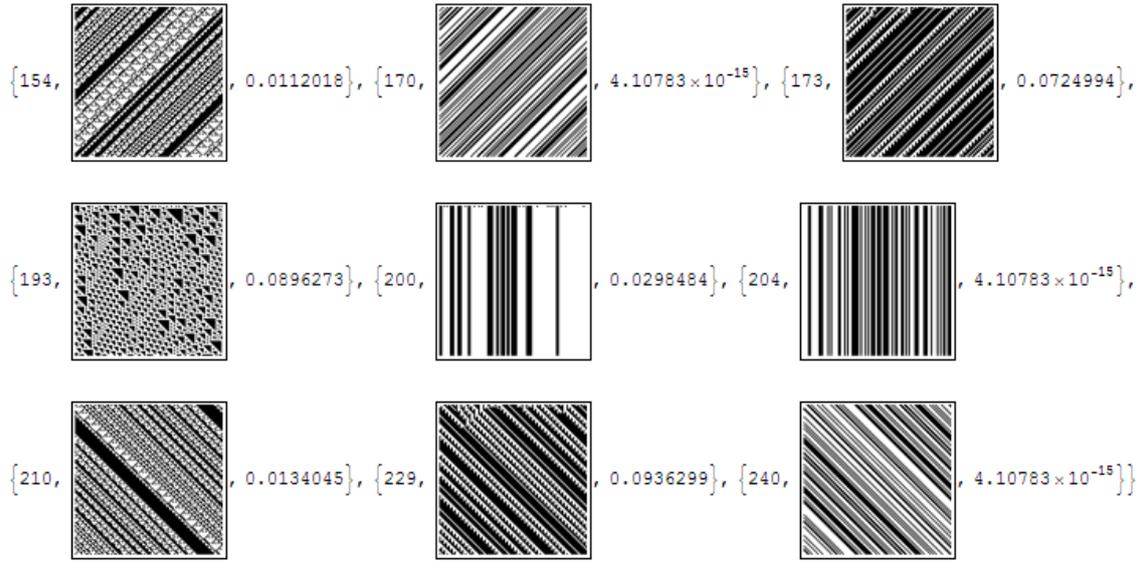




$$\Delta h_b < 0$$

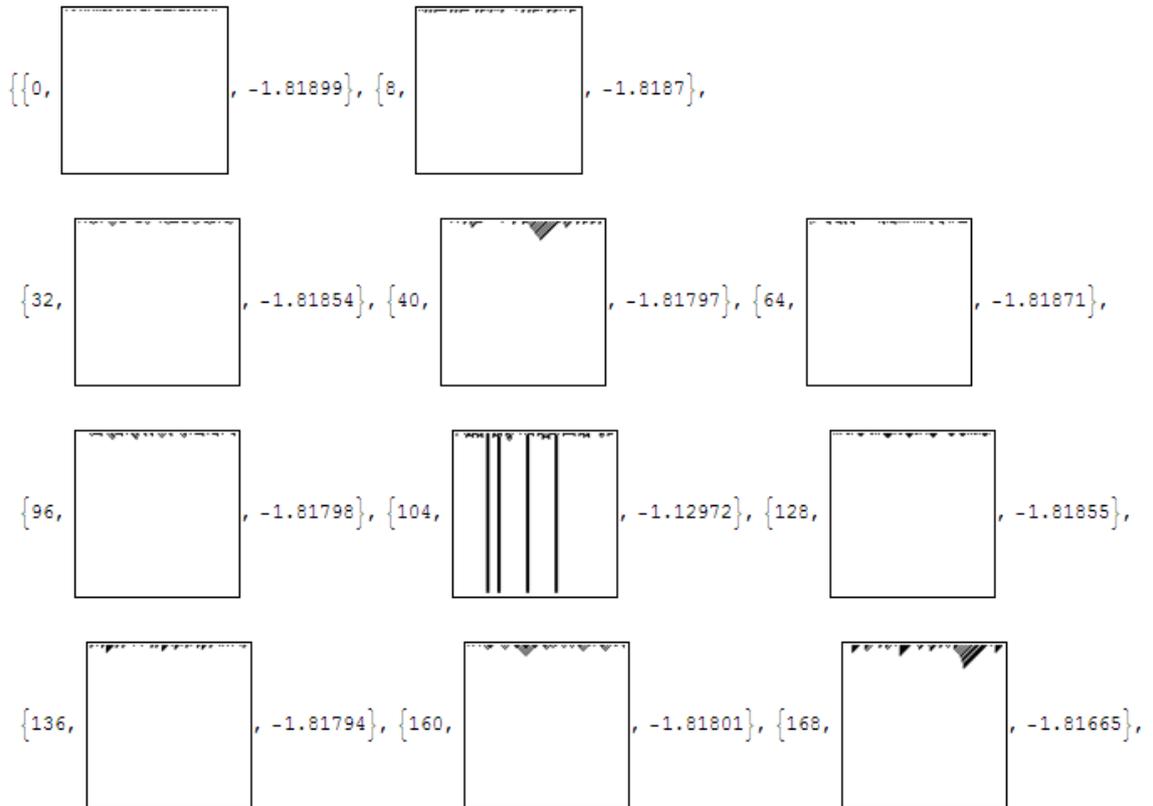
E para a base natural:

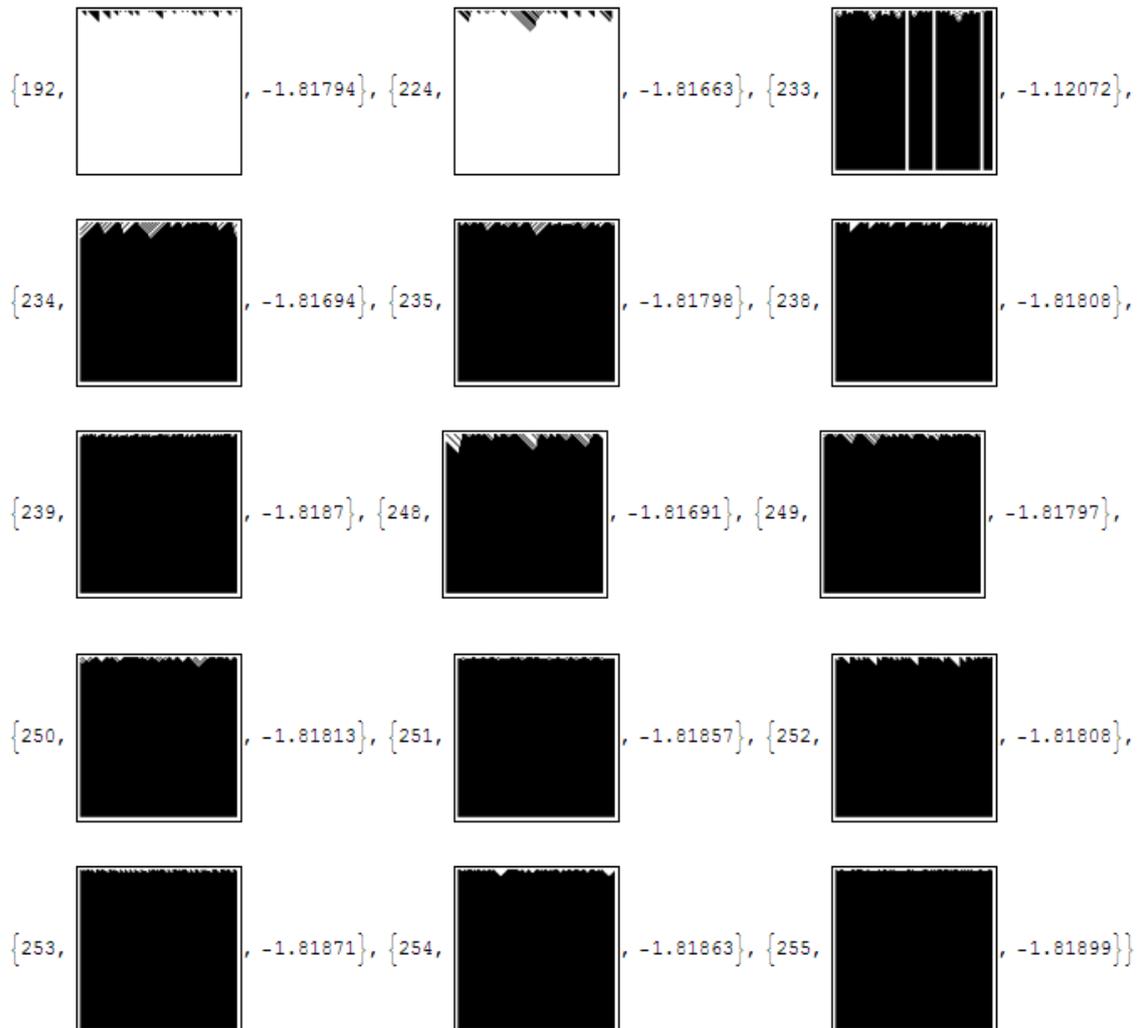




$\Delta h > 0$

E





$$\Delta h < 0$$

Esse conjunto de autômatos selecionados de acordo com os módulos expressivos de Δh mostram que quando a entropia calculada é base dependente, os autômatos de convergência nula exibem $\Delta h > 0$, e quando a entropia é base natural, o conjunto de autômatos de convergência nula apresentam $\Delta h < 0$. Autômatos de convergência nula levam um estado inicial qualquer, aleatório, até condições de equilíbrio de configuração muito simples, as quais deveriam exibir menor entropia que nos instantes iniciais.

Isto deverá ocorrer pois as regras atuam mais fortemente sobre as configurações, levando mais rapidamente até o estado estacionário do que regras mais complexas; isso, termodinamicamente, quer dizer que trabalhos são realizados sobre os sistemas, trabalhos maiores, tornando suas entropias menores que zero. De forma que a entropia em base natural (ou qualquer base que não seja normalizada) é mais fisicamente adequada para os autômatos que a

entropia base dependente, haja vista os Δh serem negativos para os autômatos de configuração mais simples, e positivos para autômatos de equilíbrio mais complexo.

Sabendo agora que a entropia RLE do tipo Shannon, considerando a comunicação entre as bordas de uma cadeia, é fisicamente legítima para cadeias de dados, e conhecendo, então, quais os autômatos exibem comportamento mais natural e quais o exibem menos, fica-se apto a construir um sistema termodinâmico real completamente baseado em autômatos.

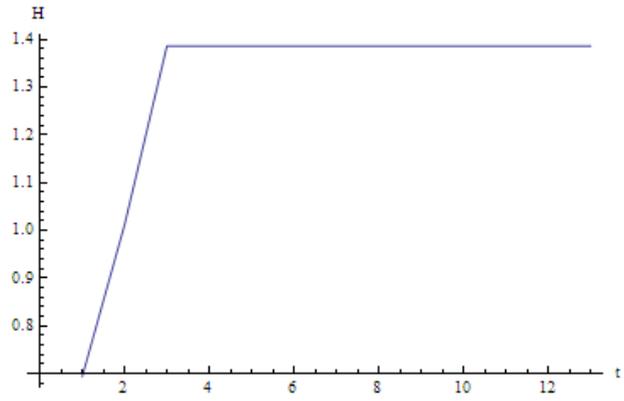
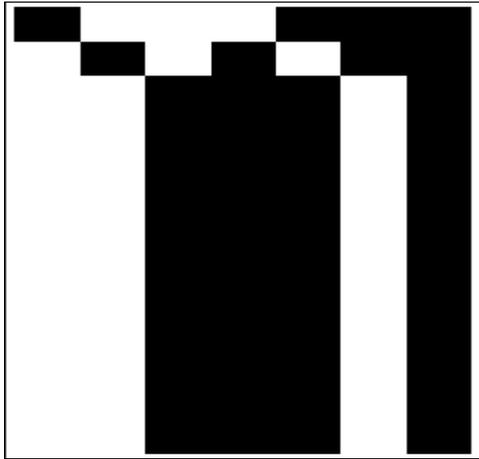
Ao longo de quase tudo que foi tratado a respeito de autômatos celulares elementares, deixou-se evidente que cada forma específica de arranjo se deve a uma maneira diferente com que cada regra atua sobre cada condição inicial distinta. Inclusive conjecturou-se que algo semelhante a uma força e de (IV.c) é que leva as cadeias até os estados de equilíbrio. De toda maneira, sempre pareceu claro que um agente externo atua sobre as sequências nos autômatos, o que ficou bastante explícito quando se calculou os equivalentes aos centros de massa para os anéis e viu-se que os mesmos se moviam por trajetórias que dependiam muito da classe à qual pertencia o arranjo relacionado.

Um teorema fundamental da mecânica clássica diz que para um sistema isolado de partículas, o centro de massa permanecerá estático até que o conjunto seja submetido a forças externas. Isto evidencia que para os autômatos estudados, existe uma ação externa que atua sobre as configurações de caracteres, pois seu centro de massa não sustenta sua posição, nem no estado estacionário.

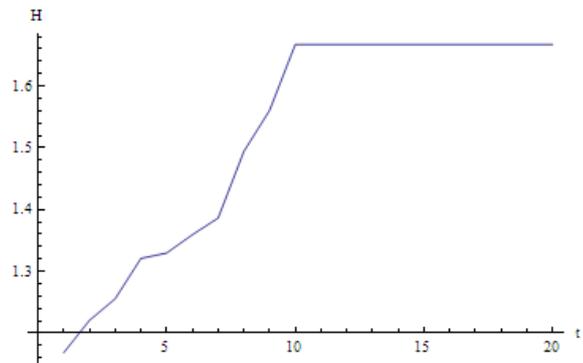
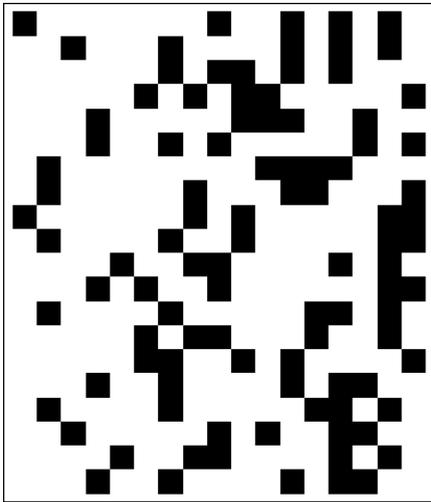
A proposta com esta pequena teoria progressiva é, baseando-se na conservação do centro de massa, construir um sistema automático que seja isolado de ações externas, mas que evolua termodinamicamente, de forma representar um sistema natural, real ou não.

Para tanto, esse sistema também deve ser isolado termicamente, não pode trocar constituintes – conserve seus elementos –, não pode alterar o centro de massa e deverá ter a sua entropia incrementada ao longo do tempo. Esse sistema deverá assemelhar-se a um gás dentro de um recipiente isolado, mas um gás de caracteres que respeite grandezas de informação específicas. Tal sistema, dada uma condição inicial aleatória, deverá permutar seus elementos de modo a nunca diminuir a entropia e, reiterando, sempre conservar o centro de massa do anel.

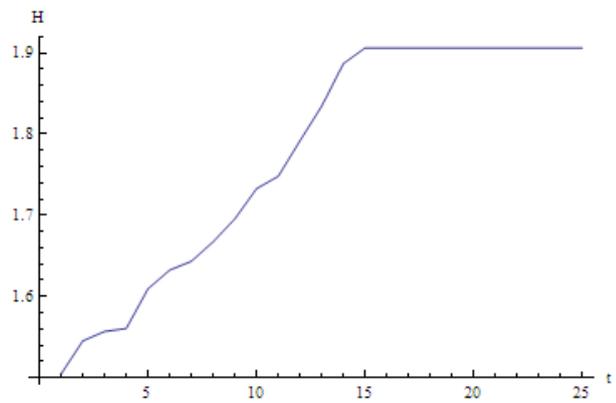
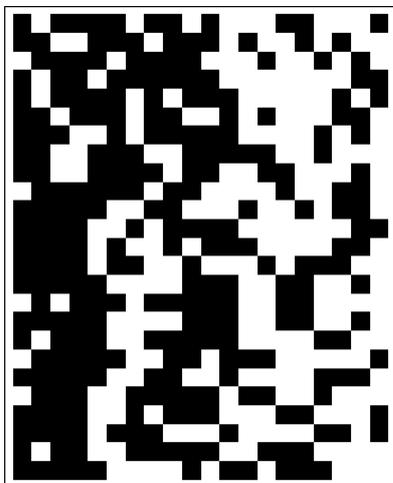
Se por ventura a evolução encontrar mais de uma configuração equivalente, dois arranjos com a mesma entropia e o mesmo centro de massa, tais configurações deverão ser equiprováveis, e a escolha deverá ser deixada ao acaso. Assim, um exemplo de arranjo com essas características:



Um pouco maior:



Ainda maior:



Essas estruturas obedecem à conservação do centro de massa, à conservação de caracteres e ao aumento da entropia com o passar do tempo, até o estado de equilíbrio quando ela é máxima. Se comportam como gases de dados sobre um anel, e os constituintes alternam posição discretamente a passos discretos. Não se pode dizer com certeza que representam sistemas físicos reais, e, se representam, quais são eles. Mas nitidamente são autômatos naturais e fornecem base para um estudo mais interessante.

É sabido que muitos autômatos bidimensionais, como os jogos da vida de Conway, são escolhidos para simular sistemas biológicos reais. Eles fazem parte de uma classe de regras de ordem superior, e sua evolução frequentemente é representada em animações. A proposta aqui submetida visa, futuramente, dentro da disponibilidade, simular sistemas bidimensionais como esses pelas regras termodinâmicas e mecânicas de conservação, eventualmente por físicas mais modernas, a fim de se verificar se é possível obter sistemas reais a partir de leis fundamentais. O que seria salutar.