Discrete Applied Mathematics 37/38 (1992) 193-203
North-Holland

A new look at the de Bruijn graph

Harold Fredricksen

Department of Mathematics, Code MA, Naval Postgraduate School, Monterey, CA 93943, USA

Received 5 July 1989
Revised 18 April 1990

Abstract

Fredricksen, H., A new look at the de Bruijn graph, Discrete Applied Mathematics 37/38 (1992) 193–203.

The Good–de Bruijn graph was originally defined to settle a question of existence of a certain shift register sequence, namely a binary cycle of length 2^n containing each of the different binary n-tuples. The further properties of the graph have been studied by several authors. Because the graph is non-planar and fairly complicated to draw for longer shift register lengths there have been several attempts to improve the description of the graph. We detail several of these attempts. Each description has its positive and negative features. Most assuredly each possible description of the graph illustrates some features and hides others. As the de Bruijn graph possesses dozens of interesting properties, each different presentation will have its own advantages.

Finally, while considering an esoteric property that the graph possesses, the ultimate depiction of the graph, in the author's view, has emerged. This version of the de Bruijn graph is the primary subject of the paper.

"Would you tell me, please, which way I ought to go from here?"

"That depends a good deal on where you want to get to," said the cat.

"I don't much care where–," said Alice.

"Then it doesn't matter which way you go," said the cat.

Chapter 7, Alices Adventures in Wonderland

1. Introduction

The binary de Bruijn graph of span n, B_n is a directed graph of 2^n nodes and 2^{n+1} arcs or directed edges. The nodes are labelled by the 2^n binary n-tuples. The graph is regular of degree 2 with two edges into and out of each node. There is an edge from $x_1x_2 \ldots x_n$ to $y_1y_2 \ldots y_n$ if and only if $x_2 \ldots x_n = y_1 \ldots y_{n-1}$. The k-ary de Bruijn graph is defined similarly.

The graph was constructed by de Bruijn [1] and independently by Good [6] while exhibiting a sequence of length 2^n bits containing all 2^n n-tuples of zeros and ones. Such sequences are defined as Hamiltonian paths through all of the nodes of B_n.

Elsevier Science Publishers B.V.

In Fig. 1 we exhibit the graphs B_3 and B_4. We can label the edge from $x_1x_2\dots x_n$ to $y_1y_2\dots y_n$ with the $(n+1)$-tuple $x_1x_2\dots x_ny_n = x_1y_1y_2\dots y_n$ since $x_2\dots x_n = y_1\dots y_{n-1}$. Thus the $(n-1)$-tuple $x_2\dots x_n$ determines an adjacency quadruple with two predecessors determined by $x_1 = 0$ and $x_1 = 1$ and two successors determined by $y_n = 0$ and $y_n = 1$. We can imagine the $(n+1)$-tuple labels on the edges of B_n as playing the role of the $(n+1)$-tuple labels on the nodes of B_{n+1}. Thus, an Eulerian circuit through all of the edges of B_n, which is guaranteed by the nature of the graph B_n, determines a Hamiltonian circuit through all of the nodes of B_{n+1}. Therefore, by interpreting the edges of B_n as nodes in B_{n+1} and adding the appropriate edges in B_{n+1}, the successively larger graphs can be obtained. Hamiltonian paths therefore exist through B_n for all $n > 1$ and the sequence 01 is a Hamiltonian path through B_1.

The graphs become increasingly more difficult to draw as n increases, perhaps because the graphs B_n are nonplanar for $n > 3$. In fact, if B_n^k is the k-ary de Bruijn graph of span n containing k^n nodes and k^{n+1} edges, B_n^k is only planar for B_n^1, B_1^2, B_1^3, B_1^4, B_2^2, B_2^3 and B_3^2 [7]. The genus of B_n^k is still an open question.

There have been several attempts to give an alternative description of the de Bruijn graph. For example, to resolve some of the difficulties with the planar property of B_n the graph has been displayed on/in a sphere as in Fig. 2. There are two arcs from the node $x_1x_2\dots x_n$. We call the arc $x_1x_2\dots x_n \to x_2\dots x_nx_1$ the pure cycle transition and $x_1x_2\dots x_n \to x_2\dots x_n\bar{x}_1$ the jump cycle transition. The pure cycle transition determines cycles of nodes of length d for each d dividing n.

These cycles can be constructed in an efficient fashion using the θ operation defined as follows. Let $(a_1a_2\dots a_n)$ be a cycle of length n. Then the n different nodes $a_1a_2\dots a_n, a_2a_3\dots a_na_1, a_3\dots a_na_1a_2, \dots, a_na_1\dots a_{n-1}$ are all on the cycle of length n. We pick a representative for the cycle as the "largest" node on the cycle. That is, $a_1a_2\dots a_n$ is the representative of its cycle if for every i there is some j such that $a_1 = a_i$, $a_2 = a_{i+1}, \dots, a_{j-1} = a_{i+j-2}$, and $a_j = 1$ while $a_{i+j-1} = 0$ where all of the

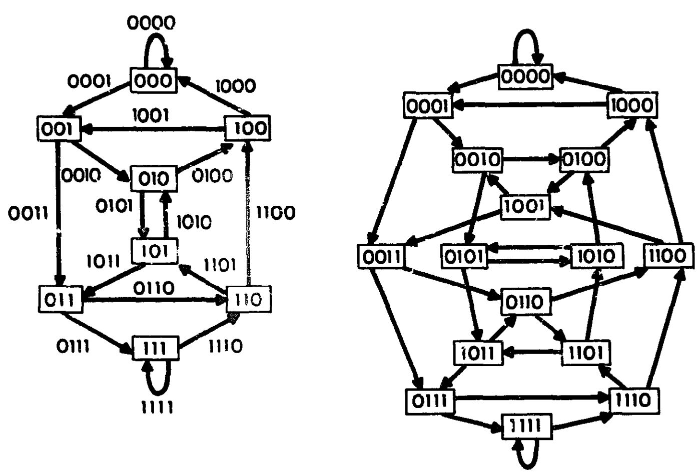

Fig. 1. The Good-de Bruijn graph, $n = 3, 4$.

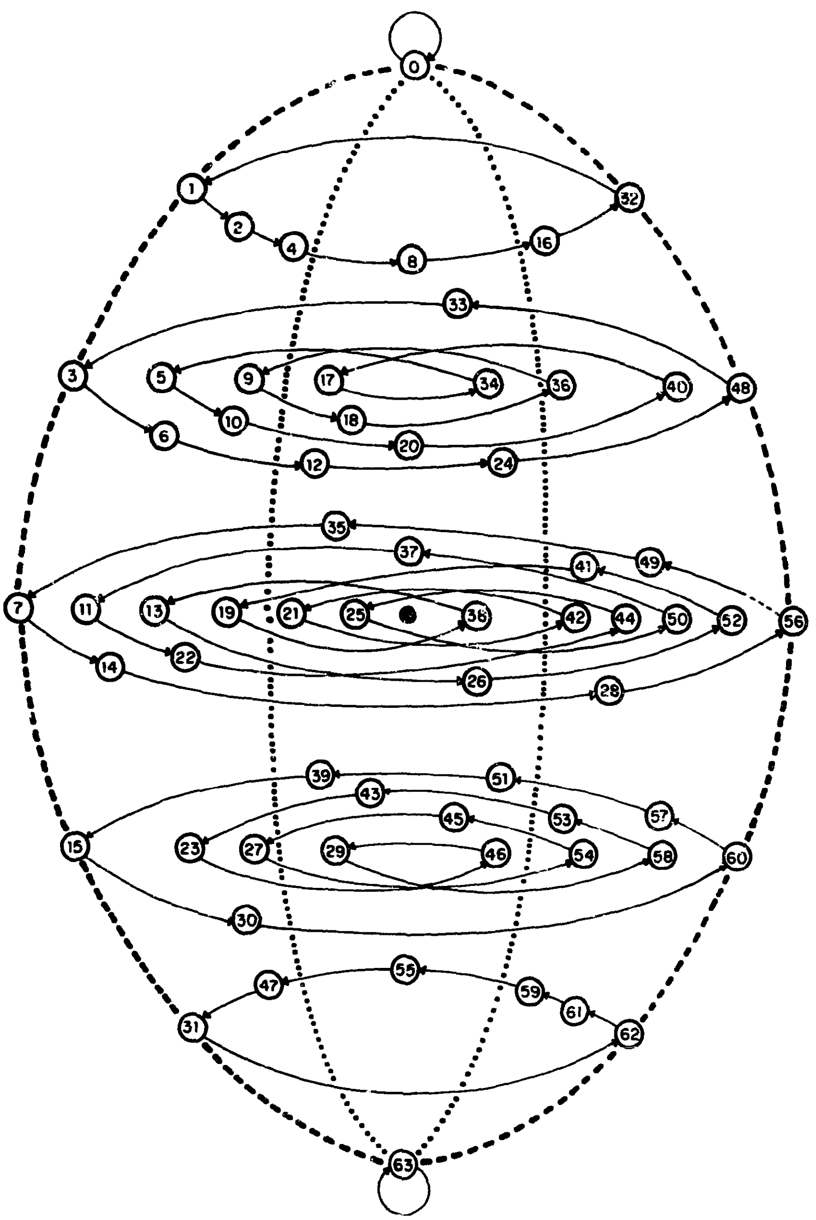

Fig. 2. The Good-de Bruijn graph displayed on a sphere, $n=6$.

subscripts bigger than n are reduced by n. Also the representation of $a_1a_2\ldots a_n$ by the integer $\sum_{j=1}^{n} a_j2^{n-j}$ is the largest integer appearing on the cycle of integers. Now if $a_1a_2\ldots a_n$ is the representative of its cycle generated by the pure cycle transition, then the next cycle is given as $\theta[(a_1a_2\ldots a_n)]=(b_1b_2\ldots b_n)$ which is formed by first finding the largest subscript j such that $a_j=1$ and $a_k=0$ for all $k>j$. Then $(b_1b_2\ldots b_n)=([a_1\ldots a_{j-1}0]^l a_1\ldots a_s)$ where $0\le s<j$ and $n=l\cdot j+s$. If $s=0$, then $a_1\ldots a_s$ is the empty string ε. Then, as is shown in [4], $(b_1b_2\ldots b_n)$ is the "next"

representative of its cycle, determined by the pure cycle transition, iff $s=0$. If $s\neq 0$, then θ is again applied to the resulting vector until s does finally equal 0 at which point the next representative of its cycle is found. If we start with the first (largest) cycle (11...1) and continue until the last (smallest) cycle (00...0) is reached every pure cycle is encountered. Note, θ[(00...0)] is undefined as (00...0) is the last cycle in the list. As an illustration we apply θ to the cycles of length $n=5$. The cycle (11111) is the first cycle. θ[(11111)] = (11110). As $j=5$ and $s=0$, (11110) is the representative of its cycle. Continuing we produce θ[(11110)] = (11101) which has $j=4$ and $s=1$ so is not the representative of its cycle. Applying θ again we have $\theta[\theta$[(11110)]] = (11100) which is the representative of its cycle. Finally we produce the representatives (11010), (11000), (10100), (10000) and (00000) in order, at which point the process ends. After forming the cycle sets by the pure cycle transition we display the pure cycles on/in the sphere.

At successively lower latitudes from the north pole to the south pole we display cycles of successively greater weight from weight 0 to weight n where by the weight of a cycle we mean the number of ones in n/d copies of the pure cycle of length d.

For weights w with $2\leq w\leq n-2$ there is more than one cycle of weight w for $n>3$. We display on the surface of the sphere the cycle of weight w which has all of its ones contiguous in the cycle. Other cycles of weight w are displayed in the interior of the sphere at the same latitude. The cycles with fewest contiguous ones are the cycles most near the diameter of the sphere which runs from the north pole to the south pole. In Fig. 2 we show the pure cycle transitions for the graph B_6 where the decimal representations of the binary n-tuples are given. We suppress drawing the jump cycle transition edge for each node as they make the figure impossibly difficult to understand. Though this makes the graph easier to view, the resulting graph displayed on a sphere is still nonplanar.

2. The result of Golomb. Lempel and Mykkeltveit

One of the problems considered for the de Bruijn graph is the maximum number of simultaneous, node-disjoint cycles into which the graph can be decomposed. Under the pure cycle mapping $x_1\ldots x_n\to x_2\ldots x_n x_1$ the graph is decomposed into Z_n cycles where

$$Z_n=\frac{1}{n}\sum_d \phi(d)2^{n/d}.$$

Here the summation is over all divisors d of n and ϕ is Euler's totient function. The cycles in the decomposition are all the cycles of length d for any d dividing n. In the decimal representation, the elements of a pure cycle are the numbers $i, 2\cdot i, 2^2\cdot i,\ldots, 2^{n-1}\cdot i$ all reduced modulo 2^n-1.

Golomb [5, p. 174] conjectured that it was never possible to decompose B_n into more cycles than Z_n. Lempel [9], trying to prove Golomb's conjecture made a conjecture of his own which implied Golomb's. Lempel conjectured that the minimal number of nodes (and all edges incident on those nodes) which could be removed from B_n and leave the resulting graph acyclic is Z_n.

The remaining problem was to find a set of Z_n nodes to be removed from B_n to leave the resulting graph acyclic for each n. Clearly, no fewer than Z_n nodes would do as the pure cycle mapping shows. One node surely must be removed from every cycle of the pure cycle map if the conjecture is valid. Mykkeltveit [10] gave an algorithm to find a node for each cycle of the pure cycle map which when removed would leave the graph without cycles. This result proves Lempel's conjecture and thereby Golomb's conjecture. We call such a minimal set of nodes a decycling set, D.

3. Another version of the de Bruijn graph

The decycling set of the previous paragraph can be used to generate a new depiction of the de Bruijn graph. It is this version of the graph that the current author most favors. Using the decycling set D we determine the cycle adjacency array. The cycle adjacency array is formed by first finding a decycling set D of Z_n nodes from the graph B_n. One node is chosen from each of the Z_n cycles formed by the pure cycle map. The set D is then written in a column. A new decycling set is formed from D by finding each pair of nodes in the decycling set of the form j and $2^{n-1}+j$. For each choice of j, these are the two possible predecessors of the nodes $2j$ and $2j+1$

a

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	4	4	4	4	4	4	4	4	4	4	4	8	8	16
17	3	3	6	6	6	12	12	12	12	12	24	24	24	24	17
18	18	5	5	10	10	10	10	10	10	10	10	20	9	18	18
19	19	19	7	7	7	7	14	14	28	28	25	25	25	19	19
21	21	21	21	11	22	13	13	13	13	26	26	21	21	21	21
27	27	27	27	27	23	23	15	30	29	27	27	27	27	27	27
31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31

b

0															
1	2	4											8		16
	3		6			12					24				17
		5		10								20	9	18	
			7				14		28		25			19	
				11	22	13				26		21			
					23		15	30	29	27					
								31							

Fig. 3. The cycle adjacency array, $n=5$.

under the left cycle mappings. Here both j and $2^{n-1}+j$ have the possible successors $2j$ and $2j+1$ taken modulo 2^n-1. Then D is replaced by D' when we replace j and $2^{n-1}+j$ by the new nodes $2j$ and $2j+1$ in their respective rows in the next adjacent column. Other elements of the decycling set D that do not change are recopied in their respective rows to complete D'. From D' a new pair i and $2^{n-i}+i$ is found to change to produce D^2 in the same manner. Then in order by the same process are found D^3, D^4, etc. The process must either terminate or become ultimately periodic. When a proper decycling set D is chosen then the process ultimately becomes periodic of period p, say, and $D^p=D$.

An example is given of a cycle adjacency array for $n=5$ in Fig. 3(a). Any decycling set defines a cycle adjacency array; and any column of the cycle adjacency array is a decycling set for the graph. For emphasis, only the elements of D which change in D' are included in Fig. 3(b). In Fig. 3(b), it is sometimes possible to exchange both the pair i and 2^n+i and the pair j and $2^{n-1}+j$ at the same time. The order chosen is of no importance. In fact, both changes can be made simultaneously if no confusion results.

In Figs. 4(a) and 4(b), the $n=7$ graph is depicted. In Fig. 4(a) only one change is made in each column while in Fig. 4(b) all possible changes are taken simultaneously in the same column. No essential confusion results as the reader can observe.

One may view the cycle adjacency array as written on a cyclinder. Then the cycle adjacency array as described in Fig. 3(b) is the description of the de Bruijn graph most favored by the author. The 2^n distinct numbers appearing correspond to the

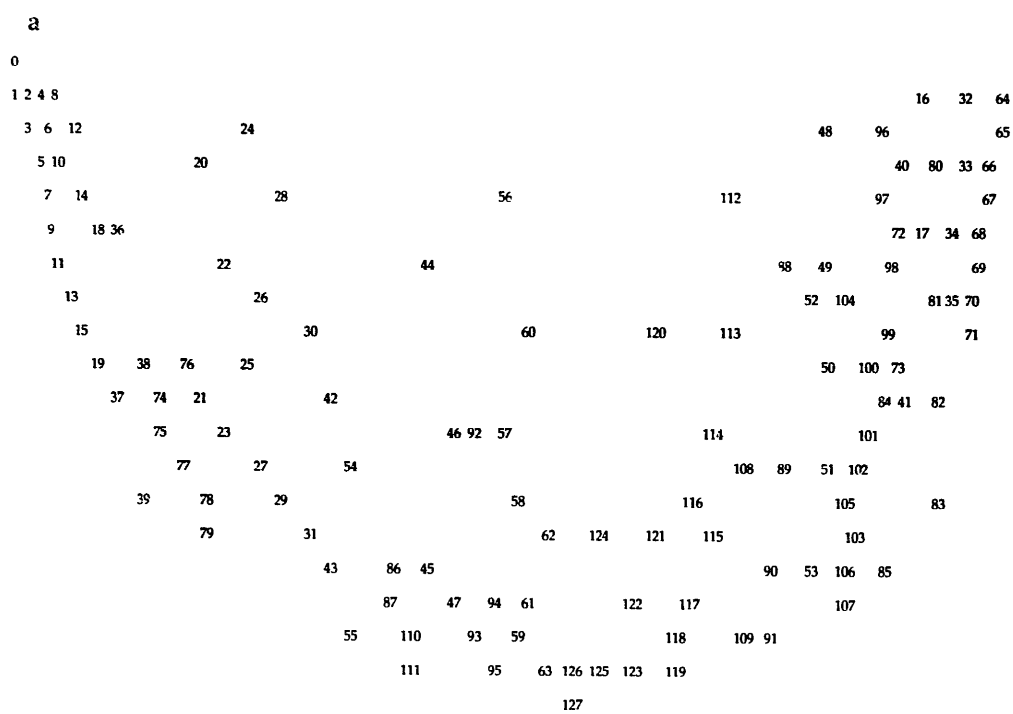

Fig. 4. (a) The best de Bruijn graph, $n=7$.

b

0																									
1	2	4	8																			16		32	64
	3	6	12				24												48	96					65
		5	10				20															40	80	33	66
		7	14				28					56				112				97					67
			9	18	36																72	17		34	68
			11				22			44								88	49	98					69
			13				26												52	104			81	35	70
			15				30					60			120	113				99					71
				19	38	76	25												50	100	73				
					37	74	21	42													84	41	82		
						75	23			46	92	57				114				101					
						77	27	54									108	89	51	102					
					39	78	29					58				116				105			83		
						79	31					62		124	121	115				103					
								43	86	45								90	53	106	85				
									87	47	94	61			122	117				107					
								55	110		93	59				118	109	91							
									111		95	63	126	125	123	119									
													127												

Fig. 4. (b) The best de Bruijn graph with overlapping columns, $n = 7$.

2^n nodes of the graph. The two edges from any element j in the graph are shown in the array as the "next" element in the same row, $(2j \pmod{2^n - 1})$ and the "other" element $(2j+1 \pmod{2^n - 1})$ which is the unique other element appearing in the same column as the "next" element. The element $2j+1 \pmod{2^n - 1}$ appears in the same row as the other predecessor element $2^{n-1}+j$. The edges of the graph to successor nodes are thus well enough defined in the array that we need not even draw them on the array. Their appearing only clutters the picture.

By its definition this description of the graph shows up the cycles of the pure cycle map in a much better way than that of the description of the graph in Fig. 1. Also, the cycle adjacency description of B_n yields a relatively easy way to describe a de Bruijn cycle.

De Bruijn cycles are Hamiltonian paths in the graph. It was a desire to produce such a cycle that spurred the work of Good and de Bruijn to study the original graph. For more information on de Bruijn sequences the reader is referred to the survey paper [3].

Table 1. Decycling sets for $n = 1, 2, \ldots, 10$

n	D_n = decycling sets of B_n
1	0, 1
2	0, 2, 3
3	0, 4, 5, 7
4	0, 8, 9, 10, 11, 15
5	0, 16, 17, 18, 19, 21, 27, 31
6	0, 32, 33, 34, 35, 36, 37, 38, 39, 42, 43, 54, 55, 63
7	0, 64, 65, 66, 67, 68, 69, 70, 71, 73, 82, 101, 102, 83, 103, 85, 107, 91, 119, 127
8	0, 128–143, 146, 147, 202, 165, 203, 204, 166, 205, 206, 167, 207, 170, 171, 214, 215, 219, 238, 239, 255
9	0, 256–271, 273, 324, 393, 290, 394, 325, 395, 396, 326, 397, 291, 398, 327, 399, 292, 293, 294, 295, 297, 338, 405, 406, 421, 407, 307, 358, 422, 461, 339, 462, 359, 423, 463, 341, 427, 347, 363, 471, 438, 439, 375, 495, 511
10	9, 512–543, 546, 547, 570, 786, 649, 787, 788, 650, 789, 581, 790, 651, 791, 792, 652, 793, 582, 794, 653, 795, 796, 654, 797, 583, 798, 655, 799, 585, 658, 805, 806, 659, 807, 660, 916, 714, 661, 917, 613, 662, 918, 715, 663, 919, 614, 615, 681, 860, 870, 845, 923, 924, 691, 846, 925, 851, 926, 871, 847, 927, 682, 683, 685, 855, 858, 859, 861, 862, 863, 731, 951, 887, 990, 991, 1023

In the original depiction of the de Bruijn graph, cycles are relatively easy to discover. A disadvantage to the current description of the de Bruijn graph is that it is difficult to find the cycles of length 2, 3, 4, etc. However the cycles of lengths d for d a divisor of n are still relatively easy to see in this description. These cycles show up very readily in the original Good–de Bruijn graph. Their location is somewhat further pinpointed by a result of Jump and Marathe [8]. They show the dimension n of the graph satisfies $n < l_1 + l_2$ if there is a node in B_n which lies on both a cycle of length l_1 and a cycle of length l_2. In other words, short cycles are physically far apart in the original Good–de Bruijn graph for large enough n. It is not so easy to detect these distances in the other descriptions of the de Bruijn graph.

In Table 1 we list a decycling set of B_n for $n = 1, 2, \ldots, 10$. These sets are not necessarily the sets formed by Mykkeltveit's algorithm. For $n = 8$, 9 and 10 the dash - indicates that all of the intervening integers are to be included as well.

There is no limit on how big a picture of the array describing the de Bruijn graph can be constructed by this method given sufficient stamina and a large enough piece of paper.

In particular the de Bruijn graph defined using the cycle adjacency array for $n = 7$ is shown in Figs. 4(a) and 4(b). For each node j in the graph, the pair of successors, $2j$ and $2j+1$ appear in their own column in Fig. 4(a). Note also that $2^{n-1} + j$ is found by reading back along the row from $2j+1$. In Fig. 4(b), the same graph is drawn more compactly with more than a single pair appearing in a given column. It is not difficult to differentiate two successors of a given pair j and $2^{n-1} + j$. The graph for $n = 9$ is shown in Fig. 5, again with overlapping columns.

The choice of decycling set is not unique for every span n. In particular, the decycling sets given in Table 1 are not (necessarily) the decycling sets chosen by

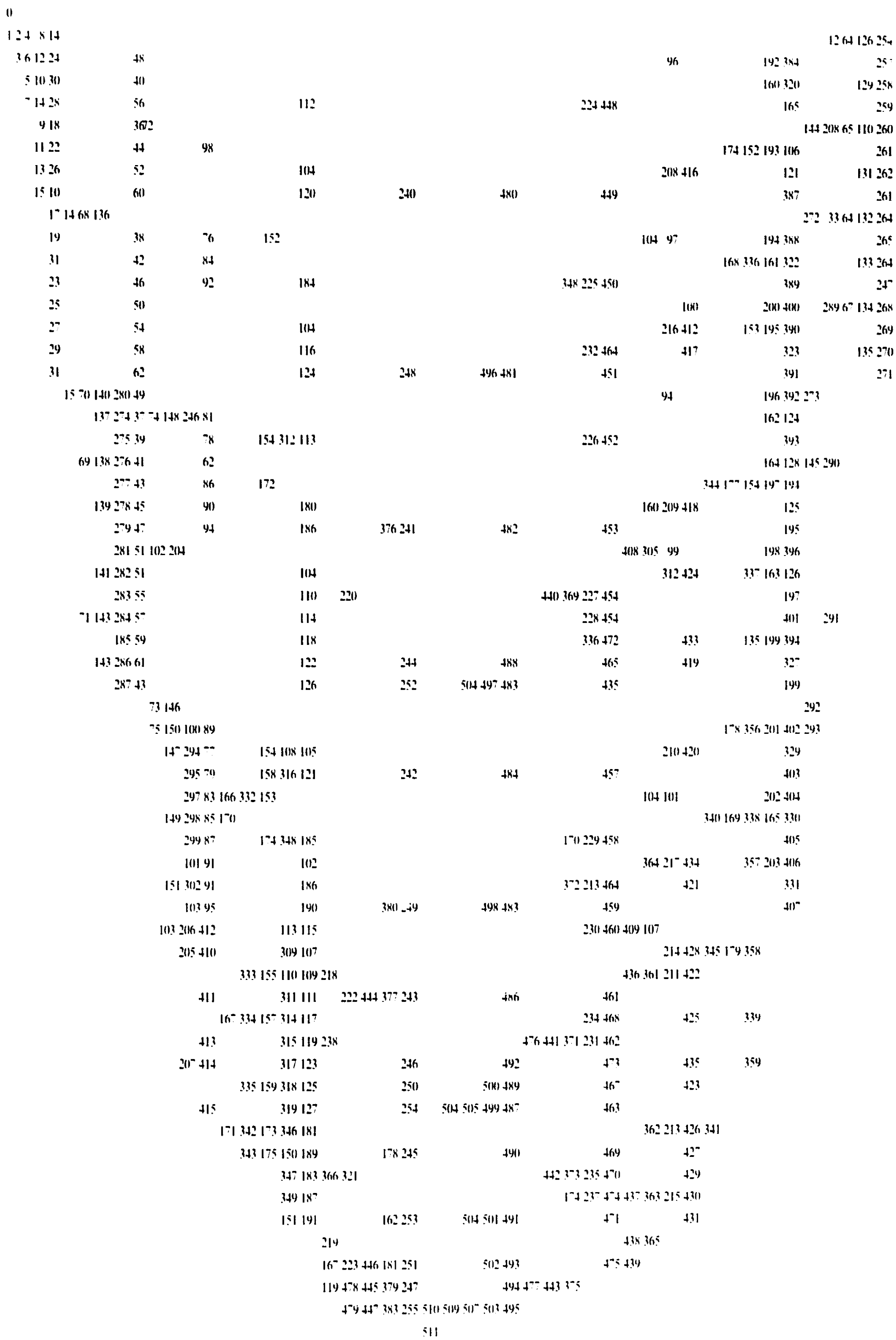

Fig. 5. The de Bruijn graph, $n = 9$.

(a)

0	0					
8	1	2	4			8
9		3	6		12	9
10			5			10
11			7	14	13	11
15				15		

(b)

0	0							
8	1	2					4	8
9		3	6		12			9
5						10	5	
11			7	14	13	11		
15				15				

Fig. 6. (a) Cycle adjacency array, $n=4$, $p=6$. (b) Cycle adjacency array, $n=4$, $p=8$.

Mykkeltveit and do not determine the same cycle adjacency array as his decycling sets. For example, we see in Figs. 6(a) and 6(b) two cycle adjacency arrays for $n=4$. The first cycle adjacency array has period 6 and the second has period 8. The first is "symmetric" in that the ith column is the complement (modulo $2^4-1=15$) of the $(i+3)$rd column. Precisely stated, the first column of the cycle adjacency array of Fig. 6(a) contains the elements 0, 8, 9, 10, 11 and 15. The elements which can change to produce the next decycling set are 0 and 8 which change to 0 and 1, respectively. In the fourth column appear the elements 0, 4, 5, 6, 7 and 15. The elements which can change are 7 and 15 to produce 14 and 15 in the next decycling set. The elements 0 and 8 are the complements of the elements 7 and 15 which appear three columns later in the array. As three is half the period of six, the array is called symmetric. The second array is not symmetric. In Table 2 some periods of cycle adjacency arrays are given. The examples leading to symmetric cycle adjacency arrays are underlined.

Table 2. Periods of cycle adjacency arrays

n	p
2	2
3	4
4	6, 8
5	12
6	14, 15, 16
7	26
8	34, 36, 39
9	46, 48, 54, 59, 60
10	64, 73

The reader may peruse the different descriptions of the de Bruijn graph to find the one that best reflects his own favorite feature of the graph.

References

[1] N.G. de Bruijn, A combinatorial problem, Nederl. Akad. Wetensch. Proc. Ser. A 49 (1946) 758-764.

[2] H. Fredricksen, A class of nonlinear de Bruijn cycles, J. Combin. Theory 19 (1975) 192-199.

[3] H. Fredricksen, A survey of full length nonlinear shift register cycle algorithms, SIAM Rev. 24 (1982) 195-221.

[4] H. Fredricksen and J. Maiorana, Necklaces of beads in k colors and k-ary de Bruijn sequences, Discrete Math. 23 (1978) 207-210.

[5] S.W. Golomb, Shift Register Sequences (Aegean Park Press, Laguna Hills, CA, 1982).

[6] I.J. Good, Normal recurring decimals, J. London Math. Soc. 21 (1946) 167-169.

[7] D.M. Johnson and N.S. Mendelson, Planarity properties of the Good-de Bruijn graphs, in: Combinatorial Structures and Their Applications (Gordon and Breach, London, 1970) 177-183.

[8] J.R. Jump and S. Marathe, On the length of feedback shift registers, Inform. and Control 19 (1971) 345-352.

[9] A. Lempel, On extremal factors of the de Bruijn graph, J. Combin. Theory 11 (1971) 17-27.

[10] J. Mykkeltveit, A proof of Golomb's conjecture for the de Bruijn graph, J. Combin. Theory Ser. A 13 (1973) 40-45.