The Collatz conjecture and De Bruijn graphs

Thijs Laarhoven*, Benne de Weger

Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

In memory of Prof. N.G. de Bruijn (1918–2012)

Abstract

We study variants of the well-known Collatz graph, by considering the action of the $3n + 1$ function on congruence classes. For moduli equal to powers of 2, these graphs are shown to be isomorphic to binary De Bruijn graphs. Unlike the Collatz graph, these graphs are very structured, and have several interesting properties. We then look at a natural generalization of these finite graphs to the 2-adic integers, and show that the isomorphism between the resulting infinite graphs is exactly the conjugacy map previously studied by Bernstein and Lagarias. Finally, we show that for generalizations of the $3n + 1$ function, such as the family of $an + b$ functions and Collatz-like functions, we get similar relations with 2-adic and p-adic De Bruijn graphs respectively.

© 2013 Royal Dutch Mathematical Society (KWG). Published by Elsevier B.V. All rights reserved.

Keywords: $3n + 1$ problem; Collatz conjecture; De Bruijn graphs; Shift map; Conjugacy; 2-adic integers

1. Introduction

The $3n + 1$ or Collatz conjecture is a long-standing open problem in mathematics. Let the $3n + 1$ function T be defined on the integers by

$$T(n) = \begin{cases} (3n + 1)/2 & \text{if } n \text{ is odd,} \\ n/2 & \text{if } n \text{ is even.} \end{cases} \tag{1}$$

* Corresponding author. Tel.: +31 621377276.
  E-mail addresses: t.m.m.laarhoven@tue.nl, mail@thijs.com (T. Laarhoven), b.m.m.d.weger@tue.nl (B. de Weger).
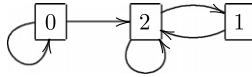 *URLs:* http://thijs.com (T. Laarhoven), http://www.win.tue.nl/~bdeweger/ (B. de Weger).

Fig. 1. The modular Collatz graph with modulus $m = 3$.

The Collatz conjecture states that, starting from any positive integer n, repeated application of the function T will eventually produce the number 1, after which it will end in the cycle $\{1, 2\}$. This conjecture is true if and only if, on the positive integers, there are no divergent paths (i.e., $T^k(n)$ is bounded for all positive integers n, where $T^0(n) = n$ and $T^{k+1}(n) = T(T^k(n))$ for $k \geq 0$) and there are no other cycles besides the trivial cycle $\{1, 2\}$ (i.e., there are no natural numbers $n \geq 3$ with $T^k(n) = n$ for some $k \geq 1$). Though easy to state, this problem seems very hard, if not impossible to solve.

Because of its simple formulation, researchers from many different branches of mathematics have at one time or another encountered this problem and have become fascinated by it. This has lead to hundreds of papers in the last few decades, with each researcher using his own area of expertise to shed a new light on this problem. An excellent overview of many of these papers was given by Lagarias [16,17], while extensive surveys of previous work on this problem can be found in books by Lagarias [15] and Wirsching [28].

Three of those branches of mathematics that have been used to study the Collatz conjecture are those of graph theory, modular arithmetic and 2-adic integers. This paper aims to show connections between these three approaches.

We start in Section 2 with studying modular Collatz graphs, i.e., finite graphs that capture the behavior of the $3n + 1$ function on congruence classes of integers. It turns out that there is an intimate relation to binary De Bruijn graphs when the modulus is a power of 2. Letting this modulus tend to infinity, in Section 3 we are led to studying these problems on the 2-adic integers. This leads to a natural generalization of the binary De Bruijn graphs to the 2-adic integers. In Section 4 we look at the structure of this infinite graph, and we try to describe how various Collatz graphs are embedded in it. In Section 5 we briefly indicate possible generalizations.

2. Binary Collatz graphs and binary De Bruijn graphs

One particular approach to the $3n + 1$ problem that caught our attention is using directed graphs to visualize the action, and in particular iteration, of the function T. We denote a directed graph by $G = (V, E)$, where V is the set of vertices, and E is the set of directed edges. Since we will not be dealing with undirected graphs, we will refer to directed graphs simply as graphs. Consider the graph $C(\mathbb{N}_+) = (V, E)$ with vertices $V = \mathbb{N}_+$ and edges $E = \{n \to T(n) : n \in V\}$. This graph is known in the literature as the Collatz graph [2,13,24, 28], and allows for a simple visual explanation of the $3n + 1$ problem to a broad audience [26].

Since the Collatz graph has infinitely many vertices and looks very chaotic, we introduce a new family of related, finite graphs. Given some modulus m, we define the *modular Collatz graph with modulus m* as the graph $G = (V, E)$ with vertices $V = \{0, 1, \ldots, m - 1\}$, and a directed edge runs from a to b if there exist some numbers $a_1, b_1 \in \mathbb{N}_+$, with $a_1 \equiv a \pmod{m}$ and $b_1 \equiv b \pmod{m}$, such that $T(a_1) = b_1$. For instance, taking $m = 3$ leads to the graph on three vertices in Fig. 1.

As can be seen from this graph, there are no edges going into the vertex 0. So once we 'leave' the set $3\mathbb{Z}$, we will never return. The only way not to leave this set is to have $T^k(n) \equiv 0 \pmod{6}$ for all n, which implies $n = 0$. Hence proving that all positive integers $n \equiv \pm 1 \pmod{3}$ iterate
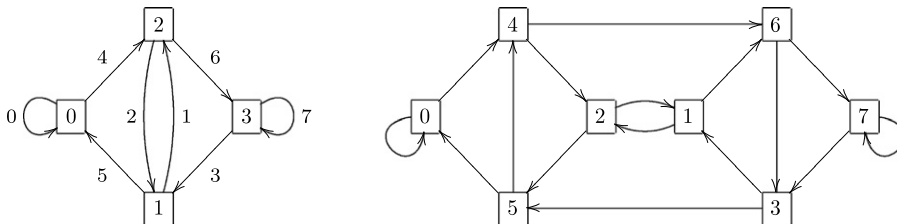
Fig. 2. The binary Collatz graphs $C(2)$ with modulus 4 (left) and $C(3)$ with modulus 8 (right).

to 1 is sufficient to prove the Collatz conjecture. Although this conclusion may seem trivial, and similar but stronger results have been derived by the Monkses [22,24], it shows that studying these graphs may be useful.

Upon further inspection, it turns out that in general, these graphs do not look particularly nice or structured. But when we take the modulus m to be some power of 2, these graphs do have a nice structure. From now on we will therefore focus on what we call *binary modular Collatz graphs*, or simply *binary Collatz graphs*. We write $C(k)$ for the modular Collatz graph with modulus $m = 2^k$, and we refer to this graph as the *binary Collatz graph of dimension k*. For convenience, we write T_k for forward iteration in the graph $C(k)$, i.e., $T_k(n)$ is the set of vertices v in the graph $C(k)$ that are connected to n by an edge $n \rightarrow v$. This relation can be explicitly written as[1]

$$T_k(n) = \begin{cases} \{(3n+1)/2, (3n+1)/2 + 2^{k-1}\} & \text{if } n \text{ is odd,} \\ \{n/2, n/2 + 2^{k-1}\} & \text{if } n \text{ is even.} \end{cases} \tag{2}$$

For $k = 2, 3$ we get the graphs $C(2)$ and $C(3)$ shown in Fig. 2. These can also be found in a recent paper of Monks et al. [24, Figs. 7.1, 7.2].

Looking at these graphs, we can immediately see a lot of structure. Both graphs have several symmetries, every vertex has two incoming and two outgoing edges, and reversing the direction of each edge leads to a graph isomorphic to the original graph. In the graph $C(2)$ we have also labeled each edge with the corresponding congruence class modulo 8, e.g., the edge from 2 to 3 has a label 6, because the numbers $a \equiv 2 \pmod 4$ satisfying $T(a) \equiv 3 \pmod 4$ are exactly all numbers $a \equiv 6 \pmod 8$. With this labeling, we can see a connection between $C(2)$ and $C(3)$: the latter can be formed by taking the so-called *line graph* of the former, associating edges in $C(2)$ to vertices in $C(3)$ and connected edges in $C(2)$ to edges in $C(3)$.

Seeing these beautiful graphs, one may wonder what is known about these graphs, and the title of this paper gives most of it away. Given an alphabet $\Sigma = \{0, 1, \ldots, p-1\}$ of size p and a word length k, the *p-ary De Bruijn graph of dimension k* [6] is defined as the graph $B(p, k) = (V, E)$ with vertex set $V = \Sigma^k$, and an edge runs from the word $a_0 a_1 \cdots a_{k-1}$ to the word $b_0 b_1 \cdots b_{k-1}$ if and only if $a_{i+1} = b_i$ for $i = 0, \ldots, k-2$. Thus, an edge runs from one word to another if the last $k-1$ symbols of the first word are equal to the first $k-1$ symbols of the second word. When $p = 2$ we also refer to these graphs as *binary De Bruijn graphs*. Besides viewing the vertices as words of a fixed length over some finite alphabet, it is also convenient to associate the numbers from 0 to $p^k - 1$ to the vertices. For this we identify words $b_0 b_1 \cdots b_{k-1}$ with numbers

---

[1] In this expression, numbers should be calculated modulo 2^k, as all vertices correspond to congruence classes modulo 2^k.
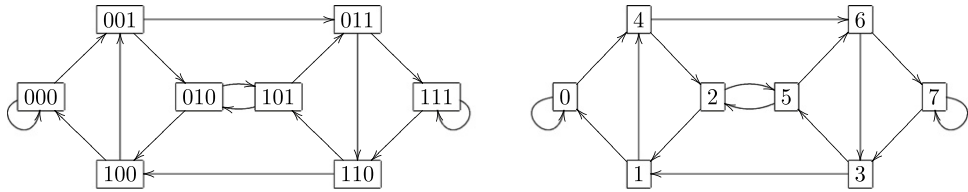
Fig. 3. The binary De Bruijn graph of dimension 3, with binary (left) and numeric (right) labels on vertices.

$\sum_{i=0}^{k-1} b_i p^i$. Fig. 3 shows the two different labelings of the binary De Bruijn graph $B(2, 3)$ of dimension 3. For the remainder of this paper, we will choose to label the vertices with these numbers rather than with words over a finite alphabet.

We write $\sigma_{p,k}$ for forward iteration in the p-ary De Bruijn graph of dimension k. In terms of finite words over Σ, this corresponds to $\sigma_{p,k}(b_0 b_1 \cdots b_{k-1}) = \{b_1 b_2 \cdots b_{k-1} x : x \in \Sigma\}$. For now we will focus on the case $p = 2$, when the relation $\sigma_{2,k}$ can be described in terms of numbers as

$$\sigma_{2,k}(n) = \begin{cases} \{(n-1)/2, (n-1)/2 + 2^{k-1}\} & \text{if } n \text{ is odd,} \\ \{n/2, n/2 + 2^{k-1}\} & \text{if } n \text{ is even.} \end{cases} \tag{3}$$

Since this relation shifts the bits to the left and appends a new bit, we will refer to this relation as the *(binary) shift relation*.

De Bruijn graphs are closely related to, and mostly studied for, finding *De Bruijn sequences*: cyclic sequences of symbols over a finite alphabet containing each sequence of length k exactly once as a subsequence. For instance, a De Bruijn sequence for $k = 3$ is given by 00010111(00). These sequences correspond precisely to Hamiltonian paths in De Bruijn graphs. The above sequence corresponds to the path starting at 000, and following the path $000 \xrightarrow{1} 001 \xrightarrow{0} 010 \xrightarrow{1} 101 \xrightarrow{1} 011 \xrightarrow{1} 111 \xrightarrow{(0)} 110 \xrightarrow{(0)} 100$, visiting each vertex of the graph exactly once, hence containing each sequence of length 3 as a subsequence exactly once. Since De Bruijn graphs are Hamiltonian, such De Bruijn sequences exist for any value of p and k, and can be constructed from a Hamiltonian path in a De Bruijn graph.[2]

Comparing Figs. 2 and 3, it is clear that the graph $B(2, 3)$ has exactly the same structure as the binary Collatz graph $C(3)$. In fact, for any value of k, the graphs $C(k)$ and $B(2, k)$ have exactly the same structure. This can be seen as follows. Let us use the notation $(x_i(n))$ for the 0–1 sequence defined by $x_i(n) \equiv T^i(n) \pmod{2}$, for $i \in \mathbb{N}$, as in [13]. For example, since the orbit of $n = 3$ under iterating T is $(3, 5, 8, 4, 2, 1, 2, 1, 2, \ldots)$, the sequence $(x_i(n))$ is this orbit taken modulo 2, i.e., $(1, 1, 0, 0, 0, 1, 0, 1, 0, \ldots)$. Then it is immediate that $x_i(T(n)) \equiv T^i(T(n)) = T^{i+1}(n) \equiv x_{i+1}(n) \pmod{2}$, i.e., applying T to n means shifting the sequence $(x_i(n))$ one position to the left. The first k terms of $(x_i(n))$ depend only on the value of $n \pmod{2^k}$, and $x_{k+1}(n)$ and $x_{k+1}(n + 2^k)$ always are different, as shown in [13, Eq. (2.9)]. So if we consider the (iterated) action of T on a congruence class of some $n \pmod{2^k}$, then we find that the first k elements of $(x_i(n))$ are identical for all numbers in this congruence class, and that the first k elements of $(x_i(T(n)))$ are those of $(x_i(n))$ shifted by one to the left, with at the end added a 0 or 1, both occurring equally often. But this shows exactly that one forward iteration in the binary

---

[2] Due to the exponential size of the De Bruijn graphs, this method of finding De Bruijn sequences is not very efficient. In Section 4 we will see a more practical method for generating these sequences.
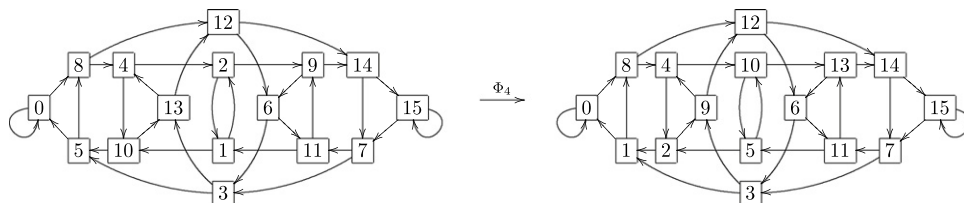
Fig. 4. The graphs $C(4)$ (left) and $B(2, 4)$ (right), and the corresponding isomorphism $\Phi_4 = (1, 5)(2, 10)(9, 13)$.

Collatz graph of dimension k can be described as the shift relation on the bit sequences of length k, i.e., as the forward iteration on the binary De Bruijn graph of dimension k.

To summarize the result, we introduce the function $\Phi_k : \{0, \ldots, 2^k - 1\} \to \{0, \ldots, 2^k - 1\}$, defined by

$$\Phi_k(n) = \sum_{i=0}^{k-1} x_i(n) 2^i. \tag{4}$$

Then we have the following statement.

Theorem 2.1. *For any $k \geq 1$, the function Φ_k is an isomorphism between $C(k)$ and $B(2, k)$.*

For the binary Collatz graphs, we introduced the notation T_k to indicate forward iteration in the graph, and for binary De Bruijn graphs we wrote $\sigma_{2,k}$ for walking along edges in these graphs. Since Φ_k is an isomorphism between the two graphs, we can describe the relation between T_k and $\sigma_{(2,k)}$ via Φ_k by

$$T_k \equiv \Phi_k^{-1} \circ \sigma_{2,k} \circ \Phi_k, \tag{5}$$

where $\Phi_k^{-1}(\{a, b\}) = \{\Phi_k^{-1}(a), \Phi_k^{-1}(b)\}$. This is why Φ_k may be called a (*k-dimensional*) *conjugacy map*.

While the relation between De Bruijn graphs and the Collatz conjecture does not appear in the literature, these bijections Φ_k have been explicitly studied before by Bernstein and Lagarias [3,4,13], but in a context different from graph isomorphisms. In [13, Table 2] these bijections were explicitly given as permutations on numbers between 0 and $2^k - 1$, e.g., for $k = 3$ we get the permutation $\Phi_3 \equiv (1, 5)$ cf. Figs. 2, 3, and for $k = 4$ we get the permutation $\Phi_4 \equiv (1, 5)(2, 10)(9, 13)$ cf. Fig. 4.

Since the graphs $C(k)$ and $B(2, k)$ are isomorphic, they share several properties. As mentioned before, the line graph of a k-dimensional binary Collatz graph is isomorphic to the $(k + 1)$-dimensional binary Collatz graph, and the transpose graph of $C(k)$, obtained by reversing the direction of each edge, is again isomorphic to $C(k)$. One other remarkable property of these graphs is that there is exactly one way to end up at any vertex in the graph after exactly k steps (including the vertex itself). This can be expressed in terms of the adjacency matrix of the graph, A_k, by $(A_k^k)_{i,j} = 1$ for all indices i, j. This was previously noted by Feix et al. [7], and Feix and Rouet [8]. Since every vertex has two outgoing edges, it follows that $(A_k^\ell)_{i,j} = 2^{\ell-k}$ for $\ell \geq k$. In particular, this value does not depend on i or j. Thus, if we start at any vertex in the graph, and take k or more random steps in the graph, we can be at all vertices with equal probability. So if we only know the k least significant bits of a number n, then we know absolutely nothing about $T^k(n)$. This also gives some intuition why the $3n + 1$ problem is so hard: unless we know

exactly what number we started out with, during each iteration we lose one bit of information about the resulting number.

We are ultimately looking for more insight into the $3n + 1$ problem, but these De Bruijn graphs are not quite the same as the Collatz graphs. The vertices are congruence classes rather than numbers, and the graphs are finite. Still, there are relations between these binary Collatz graphs and the regular Collatz graph. For instance, if we simply restrict the Collatz graph to the vertex set $\{0, \ldots, 2^k - 1\}$, we get a subgraph of $C(k)$. Also, each edge in the Collatz graph corresponds in each $C(k)$ to one edge, e.g., the edge $5 \to 8$ in the Collatz graph corresponds to the edges $1 \to 0$ in $C(1)$ and $C(2)$, the edge $5 \to 0$ in $C(3)$, and the edge $5 \to 8$ in the graphs $C(k)$ for $k \geq 4$. By considering the sequence of graphs $C(k)$, and taking only those edges that appear in infinitely many of these graphs, we exactly get the Collatz graph on the natural numbers (including 0).

3. The 2-adic Collatz graph and the 2-adic De Bruijn graph

While considering finite Collatz graphs on congruence classes leads to nicely structured graphs, we would like to know more about the infinite Collatz graph on the natural numbers. Therefore it would be interesting to consider extensions of these binary Collatz graphs and binary De Bruijn graphs to larger (infinite) vertex sets, by in some way letting k go to infinity. Then vertices may actually correspond to numbers rather than to residue classes.

Although it is not exactly clear what "$\lim_{k \to \infty} C(k)$" means, we can naturally find an answer to this question by taking a detour along the De Bruijn graphs. First, we can easily extend the concept of De Bruijn graphs to infinite sequences of symbols over a finite alphabet. We define the *infinite binary De Bruijn graph* as the graph $G = (V, E)$, where the vertex set is defined by $V = \{b_0 b_1 \cdots : b_i \in \{0, 1\}\}$, and an edge runs from a vertex $a_0 a_1 \cdots$ to a vertex $b_0 b_1 \cdots$ if and only if $a_{i+1} = b_i$, for each $i \in \mathbb{N}$. Similar to the previous section, to these infinite sequences we may associate numbers, by considering the set of 2-adic integers $\mathbb{Z}_2$ [11]. We will identify sequences $b_0 b_1 \cdots$ with 2-adic integers $\sum_{i=0}^{\infty} b_i 2^i$. With this labeling of the vertices, we will also refer to the infinite binary De Bruijn graph as the 2-*adic De Bruijn graph*, or $B(\mathbb{Z}_2)$. Note that in this graph, each vertex has only one outgoing edge. So while the shift relation $\sigma_{2,k}$ for finite binary De Bruijn graphs, mapping vertices m to its set of neighbors, is not a proper function, forward iteration in the 2-adic De Bruijn graph can be seen as a proper function from $\mathbb{Z}_2$ to $\mathbb{Z}_2$. We denote this function by σ_2, and it satisfies

$$\sigma_2(n) = \begin{cases} (n-1)/2 & \text{if } n \text{ is odd,} \\ n/2 & \text{if } n \text{ is even.} \end{cases} \tag{6}$$

In terms of binary sequences, this function σ_2 simply removes the first bit of a number and shifts the 'remaining' bits one position to the left. This function is therefore known in the literature as the *shift map* [1,3,4,12,23].

Before going back to the $3n + 1$ problem, consider the limit of the isomorphisms Φ_k, for $k \to \infty$. Using the definition $\Phi_k(n) = \sum_{i=0}^{k-1} x_i(n) 2^i$, we can just let k tend to infinity to obtain $\Phi(n) = \sum_{i=0}^{\infty} x_i(n) 2^i$. In $\mathbb{Z}_2$, this is a convergent series, and for any $n \in \mathbb{Z}_2$, $\Phi(n)$ is a well-defined 2-adic integer. By investigating $\Phi^{-1}(B(\mathbb{Z}_2))$ it becomes clear what "$\lim_{k \to \infty} C(k)$" could be, as $\Phi^{-1}(B(\mathbb{Z}_2))$ corresponds exactly to the Collatz graph extended to the 2-adic integers. The extension of the $3n + 1$ function T to the 2-adic integers was previously studied in [1,3,4,12, 14,19,23,28]. We will refer to this graph, with vertex set $\mathbb{Z}_2$ and edge set $\{n \to T(n) : n \in \mathbb{Z}_2\}$, as the 2-*adic Collatz graph*, and denote it by $C(\mathbb{Z}_2)$. The following result is immediate.
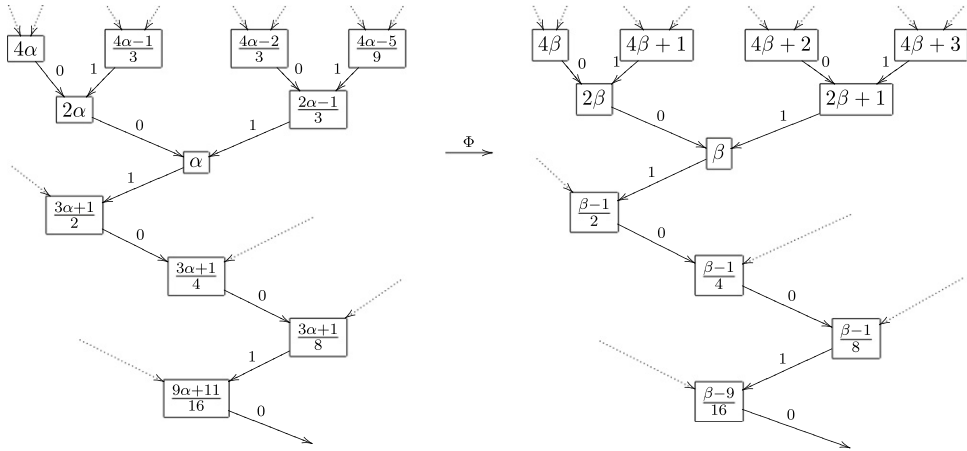
Fig. 5. A divergent component in $C(\mathbb{Z}_2)$ with an irrational number $\alpha = 10110\ldots$ (left), and the corresponding component in $B(\mathbb{Z}_2)$ with the irrational number $\Phi(\alpha) = \beta = 10010\ldots$ (right). Note that the first five bits of β follow from $\Phi_5(10110) = 10010$, or equivalently, $\Phi_5(13) = 9$.

Theorem 3.1. *The graphs $C(\mathbb{Z}_2)$ and $B(\mathbb{Z}_2)$ are isomorphic, and the function $\Phi : \mathbb{Z}_2 \to \mathbb{Z}_2$, defined by*

$$\Phi(n) = \sum_{i=0}^{\infty} x_i(n) 2^i, \tag{7}$$

is an isomorphism from $C(\mathbb{Z}_2)$ to $B(\mathbb{Z}_2)$.

The function Φ is known in the literature as the *conjugacy map* [3,4,12,13,15,23,28], and it clearly satisfies

$$T \equiv \Phi^{-1} \circ \sigma \circ \Phi. \tag{8}$$

For values of n for which the behavior of iterates of n is completely known, one can easily compute $\Phi(n)$. For instance, $\Phi(1) = 101010\ldots = -1/3$, $\Phi(5) = -13/3$, and all numbers ending in the cycle $\{1, 2\}$ correspond to rational numbers in the 2-adic De Bruijn graph with denominator 3. The Collatz conjecture is equivalent to the statement $\Phi(\mathbb{N}_+) \subseteq \frac{1}{3}\mathbb{Z} \setminus \mathbb{Z}$ [13].

4. Structure inside the 2-adic Collatz graph

Let us now further investigate the graph $C(\mathbb{Z}_2)$. First, the graph $C(\mathbb{Z}_2)$ is not connected, and contains uncountably many components. It contains two types of components: cyclic components, corresponding to cycles of the function T; and divergent components, which do not contain a cycle, but extend infinitely far in both directions.

The subgraph of $C(\mathbb{Z}_2)$ containing all divergent components, which we will denote by $C^{\mathrm{div}}(\mathbb{Z}_2)$, is arguably the least interesting part of the graph. Although it contains uncountably many components, they are all pairwise isomorphic, and the vertices in each of these graphs are all indistinguishable, in the sense that the forward mapping is an automorphism. Fig. 5 shows one of these components, containing an irrational 2-adic integer $\alpha = 10110\ldots$, and the corresponding component in $B(\mathbb{Z}_2)$.
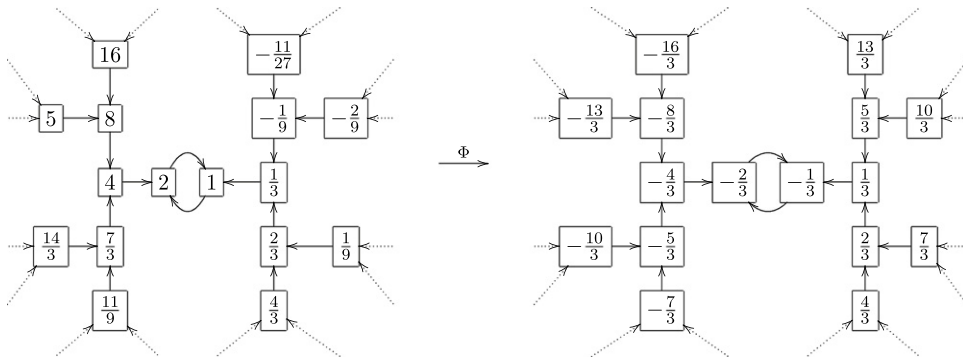
Fig. 6. The components corresponding to the cycle (01), in $C(\mathbb{Z}_2)$ (left) and in $B(\mathbb{Z}_2)$ (right). The Collatz conjecture states that the component on the left contains all positive integers.

Note that theoretically, it is possible that a divergent component arises from some *rational* 2-adic integer, i.e., a rational number with an odd denominator. It has been conjectured [13, Periodicity Conjecture] that all rational 2-adic integers are part of a cyclic component, which can be formulated as $\Phi(\mathbb{Z}_2 \cap \mathbb{Q}) = \mathbb{Z}_2 \cap \mathbb{Q}$. But so far, we only know for certain that cyclic components must arise from rational numbers, i.e., $\Phi(\mathbb{Z}_2 \cap \mathbb{Q}) \supseteq \mathbb{Z}_2 \cap \mathbb{Q}$ [3, Corollary 1].

The other part of the graph, formed by the 2-adic integers that end in a cycle, has different properties. The number of components is only countably infinite, since we can enumerate all possible cycles to find all components. For instance, there are two components with a cycle of length 1 (containing the cycles $0 \overset{0}{\to} 0$ and $-1 \overset{1}{\to} -1$), and there is exactly one component with a cycle of length 2, shown in Fig. 6. The Collatz conjecture states that this latter component contains the entire Collatz graph on the positive integers as a subgraph.

Writing M_k for the number of components in $C(\mathbb{Z}_2)$ with a cycle of length k, by counting words of length k and computing their cyclic orders, it follows that

$$2^k = \sum_{d|k} d\, M_d. \tag{9}$$

Applying Möbius inversion to the above, we obtain a direct formula for M_k as

$$M_k = \frac{1}{k} \sum_{d|k} \mu(d) 2^{k/d} = \frac{1}{n} 2^n + O(2^{n/2}). \tag{10}$$

The sequence $(M_k)_{k \geq 1} = (1, 2, 1, 2, 3, 6, \ldots)$ has been encountered before in the context of the Collatz conjecture [14], but has also been studied in different contexts [27] and is known as Moreau's necklace-counting function [25]. What this function M is actually counting is the number of different Lyndon words of length k: strings of length k that are inequivalent modulo cyclic rotations, and with period equal to k [18].[3] Somewhat surprisingly, a connection between Lyndon words and De Bruijn sequences has been made before in the literature [10,9]: one may obtain a De Bruijn sequence of order k by appropriately concatenating Lyndon words of length some divisor d of k. For instance, for $k = 4$, the Lyndon words of lengths $d|k$ are given by

---

[3] Lyndon words are actually the *representatives* of the classes of words that are equivalent through cyclic rotation, where the representative of a class is chosen as the lexicographic first word.
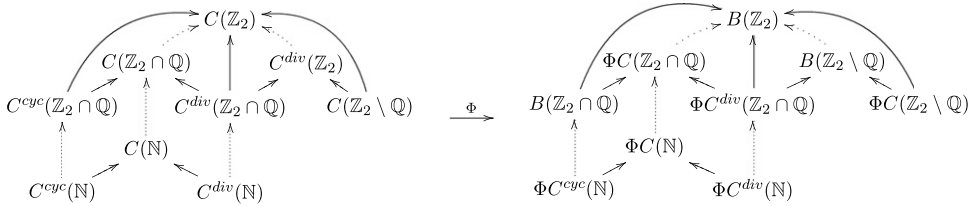
Fig. 7. The structure inside $C(\mathbb{Z}_2)$ (left) and the corresponding structure inside $B(\mathbb{Z}_2)$ (right). Solid arrows pointing to the same larger graph imply that the large graph can be partitioned into these smaller graphs. Dotted lines correspond to graph inclusions.

0, 1, 01, 0001, 0011, 0111. By extending these words to words of length 4, we may order them lexicographically as 0(000), 0001, 0011, 01(01), 0111, 1(111). Concatenating the Lyndon words in this order, we get the sequence

$$0|0001|0011|01|0111|1, \tag{11}$$

which is indeed a De Bruijn sequence of order 4, corresponding to a Hamiltonian path in the graph $B(2, 4)$. This algorithm is known in the literature as the FKM algorithm, after its inventors Fredricksen, Kessler and Maiorana [10,9].

Another fascinating property of the cyclic part of the 2-adic Collatz graph, is the fact that there is a one-to-one correspondence between cycles in $C(\mathbb{Z}_2)$, and integer cycles of the family of $3n + b$ functions $T^{(3,b)}$ defined by

$$T^{(3,b)}(n) = \begin{cases} (3n + b)/2 & \text{if } n \text{ is odd,} \\ n/2 & \text{if } n \text{ is even,} \end{cases} \tag{12}$$

for odd b coprime to 3. More precisely, if there is a rational cycle in $C(\mathbb{Z}_2)$ with all numbers having denominator b, then this cycle corresponds exactly to an integer cycle of the $3n + b$ problem. This follows from the fact that $T(n/b) = T^{(3,b)}(n)/b$. This correspondence was previously noted by Lagarias [14, p. 40]. So together, the $3n + b$ problems on integers are very structured, as we know that every cycle (Lyndon word) corresponds to exactly one primitive value b, as previously observed by Lagarias [14]. Finding the value of b associated to a given Lyndon word is not so hard, but when we try to find all Lyndon words corresponding to a value b, we get stuck. Note that solving this problem would solve the Collatz conjecture, and much more.

Going back to the complete 2-adic Collatz graph $C(\mathbb{Z}_2)$, we saw that we may divide the graph in two parts: a cyclic part, and a divergent part. Furthermore, we know that all cycles must correspond to rational numbers, but we do not know whether all rational numbers also correspond to cycles. So theoretically, we have three different types of components: cyclic, rational components; divergent, rational components; and divergent, irrational components. We will denote these parts of the graph with $C^{\text{cyc}}(\mathbb{Z}_2 \cap \mathbb{Q})$, $C^{\text{div}}(\mathbb{Z}_2 \cap \mathbb{Q})$, and $C(\mathbb{Z}_2 \setminus \mathbb{Q})$ respectively. Each of these parts of the graph corresponds to a part of $B(\mathbb{Z}_2)$, under the map Φ.

The complete structure of $C(\mathbb{Z}_2)$ and its image under Φ in $B(\mathbb{Z}_2)$ are summarized in Fig. 7. The periodicity conjecture [13] states that the middle part, $C^{\text{div}}(\mathbb{Z}_2 \cap \mathbb{Q})$, is empty. If this is true, then the picture somewhat simplifies, as then we would get $\Phi C(\mathbb{Z}_2 \cap \mathbb{Q}) = B(\mathbb{Z}_2 \cap \mathbb{Q})$ and $\Phi C(\mathbb{Z}_2 \setminus \mathbb{Q}) = B(\mathbb{Z}_2 \setminus \mathbb{Q})$.

The challenge of the Collatz conjecture now can be seen as a study in more detail of the embedding of $C(\mathbb{N}_+)$ into $C(\mathbb{Z}_2)$. The idea is that $C(\mathbb{N}_+)$ is a rather chaotic object, which lies

somehow inside the beautifully structured object $C(\mathbb{Z}_2)$. The latter graph is isomorphic via the conjugacy map Φ to $B(\mathbb{Z}_2)$, which indeed has a rich and well understood structure. Similarly other interesting subgraphs of $C(\mathbb{Z}_2)$ can via Φ be embedded into $B(\mathbb{Z}_2)$.

5. Generalizations and the *p*-adic De Bruijn graphs

5.1. *an + b* problems and the 2-adic De Bruijn graph

In the previous section we briefly mentioned a generalization of the $3n + 1$ function T to $3n + b$ functions $T^{(3,b)}$. A frequently considered further generalization of these functions is the family of *an + b functions* [5,4,7,13,15]. For odd $a, b \in \mathbb{Z}$, we define the *an + b* function $T^{(a,b)}$ by

$$T^{(a,b)}(n) = \begin{cases} (an + b)/2 & \text{if } n \text{ is odd,} \\ n/2 & \text{if } n \text{ is even.} \end{cases} \tag{13}$$

For $a = 3$, we expect that no divergent paths exist, since with equal 'probability', we either multiply a number by $1/2$ or (roughly) by $3/2$. So on average, the numbers increase by a factor $\sqrt{3/4} < 1$. For $a \geq 5$, this heuristic suggests a different behavior. For instance, for the $5n + 1$ problem and the $7n + 1$ problem, we expect most paths to diverge [5,15].

Although in this sense, the *an + b* problems behave differently from the $3n + 1$ problem, the story described above related to De Bruijn graphs applies to all of these problems as well. Defining $C^{(a,b)}(\cdot)$ similarly as the Collatz graphs $C(\cdot)$ but for the *an + b* function $T^{(a,b)}$, and defining $x_i^{(a,b)}(n)$ analogously to $x_i(n)$, we obtain the following result. This similarity between the $3n + 1$ problem and general *an + b* problems was previously noted by Feix et al. [7].

Theorem 5.1. *For any $k \geq 1$, the graphs $C^{(a,b)}(k)$ and $B(2, k)$ are isomorphic, and the function $\Phi_k^{(a,b)} : \{0, \dots, 2^k - 1\} \to \{0, \dots, 2^k - 1\}$, defined by*

$$\Phi_k^{(a,b)}(n) = \sum_{i=0}^{k-1} x_i^{(a,b)}(n) 2^i, \tag{14}$$

is an isomorphism between $C^{(a,b)}(k)$ and $B(2, k)$. Furthermore, the graphs $C^{(a,b)}(\mathbb{Z}_2)$ and $B(\mathbb{Z}_2)$ are isomorphic, and the an + b conjugacy map $\Phi^{(a,b)} : \mathbb{Z}_2 \to \mathbb{Z}_2$, defined by

$$\Phi^{(a,b)}(n) = \sum_{i=0}^{\infty} x_i^{(a,b)}(n) 2^i, \tag{15}$$

is an isomorphism from $C^{(a,b)}(\mathbb{Z}_2)$ to $B(\mathbb{Z}_2)$, satisfying

$$T^{(a,b)} \equiv (\Phi^{(a,b)})^{-1} \circ \sigma_2 \circ \Phi^{(a,b)}. \tag{16}$$

For example, for the $5n + 1$ problem and $k = 3$, we get the isomorphism $\Phi_3^{(5,1)} \equiv (1, 3)$ $(2, 6)(5, 7)$.

It is of interest to observe that for any *an + b* problem, the binary modular graphs that appear are always isomorphic to the same binary De Bruijn graphs. Only the labeling of the graphs depends on the choice of a and b. In other words, the binary and 2-adic De Bruijn graphs themselves do not contain much information anymore about the generalized Collatz problems. The wanted information is contained in the isomorphisms (the labelings of the graph edges,

the conjugacy maps), which are different for each a, b, and not well understood at all. That these maps really are different can be seen by noticing that the conjugacy map for the $3n + 1$ function maps $C(\mathbb{N}_+)$ into (most probably) one cyclic connected component in $B(\mathbb{Z}_2)$, while for the $5n + 1$ function the image of $C^{(5,1)}(\mathbb{N}_+)$ under its conjugacy map ends up in at least 3 cyclic components, but (most probably) in infinitely many components, most of which are divergent. While $C^{\mathrm{div}}(\mathbb{Z}_2 \cap \mathbb{Q})$ is conjectured to be empty, $(C^{(5,1)})^{\mathrm{div}}(\mathbb{Z}_2 \cap \mathbb{Q})$ is conjectured to have infinitely many distinct components. This may be seen as an illustration of the profound difficulty of the Collatz conjecture and its siblings.

Note that for the choice $a = 1$ and $b = -1$, the isomorphism $\Phi^{(1,-1)}$ is extremely well understood, since $T^{(1,-1)} \equiv \sigma_2$ and $\Phi^{(1,-1)} \equiv \mathrm{id}$, the identity map. So one could also write $B(2, k) = C^{(1,-1)}(k)$ and $B(\mathbb{Z}_2) = C^{(1,-1)}(\mathbb{Z}_2)$.

5.2. *p-ary functions and the p-adic De Bruijn graphs*

A further generalization of the function T was considered in, e.g., [7,8,20,21,19,28]. For some appropriately chosen integers a_i and b_i, let f be defined by

$$
f(n) = \begin{cases} (a_0 n + b_0)/p & \text{if } n \equiv 0 \pmod{p}, \\ (a_1 n + b_1)/p & \text{if } n \equiv 1 \pmod{p}, \\ \dots & \dots \\ (a_{p-1} n + b_{p-1})/p & \text{if } n \equiv p - 1 \pmod{p}. \end{cases} \tag{17}
$$

For instance, a problem Lagarias attributes to Collatz [13] concerns the function f_0 defined by

$$
f_0(n) = \begin{cases} 2n/3 & \text{if } n \equiv 0 \pmod{3}, \\ (4n - 1)/3 & \text{if } n \equiv 1 \pmod{3}, \\ (4n + 1)/3 & \text{if } n \equiv 2 \pmod{3}. \end{cases} \tag{18}
$$

Iterating these functions leads to similar behavior as the $an + b$ problems. In this case, when considering graphs on congruence classes modulo p^k, we find a relation with p-ary and p-adic De Bruijn graphs as follows. Again, the similar behavior of p-ary functions compared to $3n + 1$ functions and $an+b$ functions (when restricted to appropriate congruence classes) was previously noted by Feix et al. [7].

Theorem 5.2. *For any* $k \geq 1$, *the graphs* $C^{(f)}(p, k)$ *and* $B(p, k)$ *are isomorphic, and the function* $\Phi_k^{(f)} : \{0, \dots, p^k - 1\} \to \{0, \dots, p^k - 1\}$, *defined by*

$$
\Phi_k^{(f)}(n) = \sum_{i=0}^{k-1} x_i^{(f)}(n) p^i, \tag{19}
$$

is an isomorphism between $C^{(f)}(p, k)$ *and* $B(p, k)$. *Furthermore, the graphs* $C^{(f)}(\mathbb{Z}_p)$ *and* $B(\mathbb{Z}_p)$ *are isomorphic, and the function* $\Phi^{(f)} : \mathbb{Z}_p \to \mathbb{Z}_p$, *defined by*

$$
\Phi^{(f)}(n) = \sum_{i=0}^{\infty} x_i^{(f)}(n) p^i, \tag{20}
$$

is an isomorphism from $C^{(f)}(\mathbb{Z}_p)$ *to* $B(\mathbb{Z}_p)$, *satisfying*

$$
f \equiv (\Phi^{(f)})^{-1} \circ \sigma_p \circ \Phi^{(f)}. \tag{21}
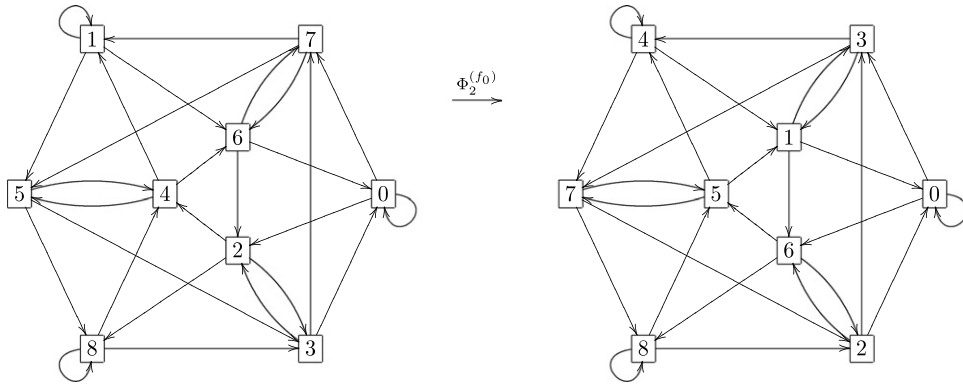$$

Fig. 8. The ternary modular graph of dimension 2 corresponding to the function f_0 in (18) and the ternary De Bruijn graph of dimension 2, $B(3, 2)$. The corresponding isomorphism $\Phi_2^{(f_0)}$ of order 7 is given by the permutation $\Phi_2^{(f_0)} \equiv (1, 4, 5, 7, 3, 2, 6)$.

As an example, Fig. 8 shows the 2-dimensional ternary graph corresponding to the function f_0 in Eq. (18), and the corresponding ternary De Bruijn graph.

Acknowledgment

References

[1] E. Akin, Why is the $3x + 1$ problem hard? Contemporary Mathematics 356 (2004) 1–20.

[2] P.J. Andaloro, The $3x + 1$ problem and directed graphs, Fibonacci Quarterly 40 (2002) 43–54.

[3] D.J. Bernstein, A non-iterative 2 -adic statement of the $3n + 1$ conjecture, Proceedings of the American Mathematical Society 121 (2) (1994) 405–408.

[4] D.J. Bernstein, J.C. Lagarias, The $3x + 1$ conjugacy map, Canadian Journal of Mathematics 48 (6) (1996) 1154–1169.

[5] R.E. Crandall, On the $3x + 1$ problem, Mathematics of Computation 32 (4) (1978) 1281–1292.

[6] N.G. de Bruijn, A combinatorial problem, Proceedings of the Section of Sciences of the Koninklijke Nederlandse Akademie van Wetenschappen 49 (1946) 758–764.

[7] M.R. Feix, A. Muriel, J.L. Rouet, Statistical properties of an iterated arithmetic mapping, Journal of Statistical Physics 76 (1994) 725–741.

[8] M.R. Feix, J.L. Rouet, The $(3x + 1)/2$ problem and its generalization: a stochastic approach, in: Proceedings of the International Conference on the Collatz Problem and Related Topics, 1999, pp. 1–18.

[9] H. Fredricksen, I.J. Kessler, An algorithm for generating necklaces of beads in two colors, Discrete Mathematics 61 (2–3) (1986) 181–188.

[10] H. Fredricksen, J. Maiorana, Necklaces of beads in k colors and k-ary De Bruijn sequences, Discrete Mathematics 23 (1978) 207–210.

[11] F.Q. Gouvea, p-Adic Numbers, Springer, 1997.

[12] B. Kraft, K. Monks, On conjugacies of the $3x+1$ map induced by continuous endomorphisms of the shift dynamical system, Discrete Mathematics 310 (13–14) (2010) 1875–1883.

[13] J.C. Lagarias, The $3x + 1$ problem and its generalizations, American Mathematical Monthly 92 (1) (1985) 3–23. Also Chapter 2 of [15], pp. 31–54, 2010.

[14] J.C. Lagarias, The set of rational cycles for the $3x + 1$ problem, Acta Arithmetica 56 (1990) 33–53.

[15] J.C. Lagarias (Ed.), The Ultimate Challenge: The $3x + 1$ Problem, American Mathematical Society, 2010.

[16] J.C. Lagarias, The $3x + 1$ problem: an annotated bibliography (1963–1999), 2011. arXiv:math.NT/0309224v13.

[17] J.C. Lagarias, The $3x + 1$ problem: an annotated bibliography (2000–2009) 2012. arXiv:math.NT/0608208v6.

[18] R.C. Lyndon, On Burnside's problem, Transactions of the American Mathematical Society 77 (2) (1954) 202–215.

[19] K.R. Matthews, Some Borel measures associated with the generalized Collatz mapping, Colloquium Mathematicum 63 (2) (1992) 191–202.

[20] K.R. Matthews, A.M. Watts, A generalization of Hasse's generalization of the syracuse algorithm, Acta Arithmetica 43 (1984) 167–175.

[21] K.R. Matthews, A.M. Watts, A Markov approach to the generalized syracuse algorithm, Acta Arithmetica 45 (1985) 29–42.

[22] K.M. Monks, The sufficiency of arithmetic progressions for the $3x + 1$ conjecture, Proceedings of the American Mathematical Society 134 (10) (2006) 2861–2872.

[23] K.G. Monks, J. Yazinski, The autoconjugacy of the $3x + 1$ function, Discrete Mathematics 275 (1–3) (2004) 219–236.

[24] K. Monks, et al., Strongly sufficient sets and the distribution of arithmetic sequences in the $3x + 1$ graph, Discrete Mathematics 313 (4) (2013) 468–489. http://dx.doi.org/10.1016/j.disc.2012.11.019.

[25] C. Moreau, Sur les permutations circulaires distinctes, Nouvelles Annales de Mathématique Séries 2 11 (1872) 309–314.

[26] R. Munroe, xkcd: Collatz conjecture, 2010. Published electronically at: http://xkcd.com/710.

[27] The On-Line Encyclopedia of Integer Sequences, OEIS, Sequence A001037, 2012. Published electronically at: http://oeis.org/A001037.

[28] G.J. Wirsching, The Dynamical System Generated by the $3n + 1$ Function, in: Lecture Notes in Mathematics, vol. 1681, 1998.