

Abstract

By decomposing the Collatz tree into a two-dimensional odd-even relation we show that it is sufficient to consider odd numbers (or a subset of even numbers) only using graph theory. A simple set of equations is used to build a connection graph which shows that all odd numbers are connected. We show that any valid proof that shows that all odd numbers (or a subset of even numbers) are connected without knowing their exact relation automatically proves the Collatz conjecture. Reasonable solutions solving the graph using graph theory or linear algebra are suggested.

1 Introduction

We start by describing that the even and odd numbers of the Collatz tree can be separated into a two-dimensional odd-even relation. The roles of odd numbers and subsets of even numbers are shown. Using that separation, we show that any possibly separated tree of odd numbers has to contain exactly one loop which implicates that, if all odd numbers are connected, the Collatz tree is built. Finally, a simple set of equations is used to build a network diagram with endpoints. The diagram and the equation set suggest that a solution should be possible using graph theory, rotation-shift-operations in 2D or similar operations in 3D.

2 The reverse path

The reverse path is built by 2 operations, implicated by the Collatz rules.

We define

$$M(x) = 2 \cdot x \quad (1)$$

which must be applied on any new number because it leads to an even number.

We define

$$D(x) = \frac{(x-1)}{3} \quad (2)$$

which must be applied on any new even number when the operation leads to an (odd) integer.

In the reverse path, every D splits the way into 2 paths. A_k shall be the indexed set of all integers on which D can be applied. The path starts at 1, which is multiplied by 2 infinitely often. On its way up, on every A_k the way is split. While A_k is further multiplied, the second way leads to an odd number. This odd number is, like the 1, multiplied by 2 infinitely often hitting other A_k on its way.

3 Properties of even numbers

The set of even numbers can be split into 3 sets each indexed with $k = 0, 1, 2, \dots$:

$$\begin{cases} 6k+4 & \text{set of even numbers } A_k \text{ that solve } D(A_k) \\ 6k+2 & \text{intermediary numbers between two } A_k \\ 6k+6 & \text{multiples of 3} \end{cases} \quad (3)$$

3.1 Properties of A_k

We define the set of even integers (only even integers are allowed) on which D can be applied as

$$A_k = 6k + 4 \quad (4)$$

with $k \geq 0$. So, they are 4,10,16,22,...

By applying D , we get the set of odd integers:

$$N_k = D(A_k) = 2k + 1 \quad (5)$$

We can make the following statements:

Lemma 1. *D is a bijective function between A_k and the set of all odd integers N_k .*

Lemma 2. *Each odd number N_k reached by $D(A_k)$ builds a new independent branch by multiplying by 2 infinitely often.*

Decimal	Binary
40	101000
20	10100
10	1010
5	101

Table 1: Binary decomposition

This follows from the fact that dividing back to the odd number is a unique path. In binary notation it can be seen that the odd number N_k is always present and a right-shift has to stop at the rightmost 1 (Table 1).

Lemma 3. *Together with the odd numbers, \mathbb{N}_+ is built by M -operations.*

This follows from the fact that any even number can be divided by 2 to finally reach an odd number.

3.2 Intermediary numbers

Intermediary numbers are even numbers that are between the double multiplication of two A_k (Lemma 4). They can be expressed as

$$I_k = 6k + 2 = 2, 8, 14, \dots \quad (6)$$

and can never be an A_k because

$$D(I_k) = \frac{6k}{3} + \frac{1}{3} = 2k + \frac{1}{3} \quad (7)$$

never is an integer.

3.3 Multiples of 3

Even numbers that are multiples of 3 are

$$J_k = 6k + 6 = 6, 12, 18, \dots \quad (8)$$

They can never be an A_k because

$$D(J_k) = \frac{6k+6}{3} - \frac{1}{3} = 2k + 2 - \frac{1}{3} \quad (9)$$

never is an integer.

4 Multiplication of A_k by 2^n

Multiplying A_k by 2^n can be viewed as

$$2^n \cdot A_k = 2^n \cdot 6k + 2^n \cdot 4 - 4 + 4 = 6 \cdot (2^n k + \frac{2}{3} \cdot 2^n - \frac{2}{3}) + 4 \quad (10)$$

The expression

$$\frac{2}{3} \cdot 2^n - \frac{2}{3} = 2 \cdot \frac{2^n - 1}{3} \quad (11)$$

contains Mersenne numbers which can be divided by 3 when n is an even number. So we can make the following statement:

Lemma 4. *For any k , $2^n \cdot A_k$ with even number $n = 2, 4, 6, \dots$ leads to another A_l with l being an even number.*

$$A_l = 6 \cdot (2^n k + 2 \cdot \frac{2^n - 1}{3}) + 4 \quad (12)$$

Between two A_k , there are intermediary even numbers $2^n \cdot A_k$ with $n = 1, 3, 5, \dots$

Lemma 4 is a statement about the relation between A_k and A_l only and not about the position of the first appearing A_k in the branch started from an odd number.

5 Relation between even and odd numbers

The general formula to reach an odd integer $N_k = 2k + 1$ from another odd integer $N_g = 2g + 1$ in the reverse path is

$$N_k = \frac{N_g \cdot 2^n - 1}{3} \quad (13)$$

or

$$2k + 1 = \frac{(2g + 1) \cdot 2^n - 1}{3} \quad (14)$$

or

$$6k + 4 = (2g + 1) \cdot 2^n \quad (15)$$

with $n = 1, 2, 3, \dots$, because for $n = 0$ we would apply the D -operation to the odd number N_g .

The left side of Eq. 15 is the set of A_k . The right side is \mathbb{N}_+ (Lemma 3).

5.1 Reachability of A_k from 3 sets of odd numbers

To find the first occurring A_k (which is the lowest n for an N_g solving Eq. 13) we split the set of odd numbers N_g into 3 sets and group them by index $s = 0, 1, 2, \dots$. The colors **Grey**, **Red** and **Blue** will be assigned to the sets.

$$\begin{cases} G_s = 3 + 6s & \text{for } N_g \equiv 0 \pmod{3} \\ R_s = 1 + 6s & \text{for } N_g \equiv 1 \pmod{3} \\ B_s = 5 + 6s & \text{for } N_g \equiv 2 \pmod{3} \end{cases} \quad (16)$$

5.1.1 Properties of G

$$x_n(s) = \frac{(3 + 6s) \cdot 2^n - 1}{3} = 2^n + 2s \cdot 2^n - \frac{1}{3} \quad (17)$$

never returns an integer, so the set G does not produce any A_k .

Lemma 5. *There are no A_k in set G .*

5.1.2 Properties of R

The red set is able to reach the odd numbers

$$x_n(s) = \frac{(1 + 6s) \cdot 2^n - 1}{3} = \frac{2^n - 1}{3} + 2s \cdot 2^n \quad (18)$$

While $2s \cdot 2^n$ is integer, the Mersenne number only divides to an integer on $n = 2, 4, 6, \dots$ (excluding $n = 0$).

Lemma 6. *All members of set R have their A_k located at $n = 2, 4, 6, \dots$*

5.1.3 Properties of B

The blue set is able to reach the numbers

$$x_n(s) = \frac{(2 + 3 + 6s) \cdot 2^n - 1}{3} = \frac{2^{n+1} - 1}{3} + (1 + 2s) \cdot 2^n \quad (19)$$

The Mersenne number only divides to an integer on $n = 1, 3, 5, \dots$

Lemma 7. *All members of set B have their A_k located at $n = 1, 3, 5, \dots$*

5.2 Self-mappings

To find possible self-mappings from g to k , we set $g = k$ in Eq. 14

$$2k + 1 = \frac{(2k + 1) \cdot 2^n - 1}{3} \quad (20)$$

and get

$$k = \frac{2^{n-1} - 2}{3 - 2^n} \quad (21)$$

For $n = 1, 2, 3, \dots$ k resolves to $k = -1, 0, -\frac{2}{5}, -\frac{6}{13}, \dots$, so the index is negative (and fractional except for $n = 1$).

Lemma 8. *The only self-mapping exists at $g = k = 0$ with $n = 2$. The offset $N_k - N_g$ is non-zero except for $g = k = 0$.*

5.3 Unique reverse path connections

As only the corresponding A_k can reach N_k (per definition), any N_k node can have only one incoming connection in the reverse path. The only option to reach some N_k over multiple ways in the reverse path would be to have multiple (g, n) -pairs reaching the corresponding A_k .

We assume two pairs, (g_1, n_1) and (g_2, n_2) reaching A_k , so Eq. 14 resolves to

$$(2g_1 + 1) \cdot 2^{n_1} = (2g_2 + 1) \cdot 2^{n_2} \quad (22)$$

or

$$\frac{2g_1 + 1}{2g_2 + 1} = 2^{n_2 - n_1} \quad (23)$$

The left side of Eq. 23 is able to generate odd numbers or fractions with odd remainder only. The right side is able to generate even numbers or fractions with even remainder except for $n_2 = n_1$ which implicates $g_2 = g_1$.

Lemma 9. *There exists a unique (g, n) -pair for every N_k . Only one (g, n) -pair can reach N_k in the reverse path.*

5.4 Unique forward path connections

We assume two different N_k reaching the same (g, n) -pair, so Eq. 14 resolves to

$$2k_1 + 1 = 2k_2 + 1 \quad (24)$$

which obviously leads to $k_1 = k_2$.

Lemma 10. *There exists a unique N_k for every (g, n) -pair. Only one N_k can reach an (g, n) -pair in the forward path.*

5.5 Loops

The Collatz conjecture defines a forward path for any positive integer including 1, so there is no stop condition. As we can always go back to the previous number via reverse path, there also is no starting point. The reverse path therefore cannot have a dangling incoming connection and we can state that there must exist at least one cycle.

Lemmas 9 and 10 do not forbid self-mappings ($N_k = N_g$). Lemma 8 restricts them to $g = k = 0$ with $n = 2$ and also states that all other N_k reached must be different from N_g . A loop is created only when either (g, n) or N_k has a duplicate connection, so we can state:

Lemma 11. *With the exception of $g = k = 0$ with $n = 2$:*

In the reverse path, any new N_k reached starting from 1 is unoccupied. In the forward path, any new (N_g, n) -pair reached from N_k is unoccupied.

The reverse tree started from 1 has exactly one cycle. A new connection in the reverse path cannot generate a cycle.

Lemma 12. *Lemma 11 is a statement about members connected to $N_k = 1$ only. Possibly unconnected nodes might form cycles on its own but as soon as it is shown that they connect to 1, lemma 11 would be violated.*

Because any N_g - N_k connection needs a corresponding A_k between them we can state:

Lemma 13. *If it can be shown that all odd numbers are connected or, alternatively, all A_k are connected, the Collatz conjecture is proved.*

Lemma 13 might seem obvious but still it is a lower requirement than showing that the whole \mathbb{N}_+ is connected.

6 Visualisation of even-odd relations

Diagram 1 shows the mapping of from N_g , passing A_k , to the corresponding N_k (green lines). The dotted green lines lead to endpoints in the reverse path. The N -axis consists of all odd numbers. The colored lines (R,G,B) contain all A_k which are reached by multiplying by 2. In the reverse path R,G,B-lines are followed from an odd number N_g on the N -axis up to any A_k . From there, the green line leads to the next odd integer N_k . In the forward path we start at any N_k or follow the R,G,B-lines to the corresponding N_k . The green line at N_k leads to the corresponding A_k . From there the red or blue line leads to the next odd integer N_g .

Note that, while in the reverse path the position of A_k in the diagram is unambiguous, in the forward path Eq. 13 cannot be solved easily without using complex numbers.

6.1 Interpretation in terms of n

Eq. 15 can be interpreted as a scan for all A_k resp. N_k (Eq. 14) over all odd numbers $2g + 1$ starting from $n = 1$ row by row and there must exist an (g, n) -pair for every k .

By resolving Eq. 15 for $n = 1, 2, 3$ we define 3 sets $S^{(n)}$. They represent the odd-to-odd connection in the reverse path for a given n . table 2 shows the resolved equation and parametric equations with independent variable $u = 0, 1, 2, 3, \dots$. $S^{(h)}$ was generated by numerical analysis of the linear equations of $S^{(n)}$. As all downward connections $g > k$ are already occupied by $S^{(1)}$, $S^{(h)}$ can only have upward connections $g < k$. According to Wolfram Alpha, $g(u)$ toggles between $3u$ and $3u + 2$ with increasing n .

Set	n	Eq. 15	$g(u)$	$k(u)$	g to k	$N_g(u)$	$N_k(u)$	occupied k
$S^{(1)}$	1	$3k = 2g - 1$	$3u + 2$	$2u + 1$	$g > k$	$6u + 5$	$4u + 3$	1, 3, 5, 7, 9, ...
$S^{(2)}$	2	$3k = 4g$	$3u$	$4u$	$g < k$	$6u + 1$	$8u + 1$	4, 8, 12, 16, ...
$S^{(3)}$	3	$3k = 8g + 2$	$3u + 2$	$8u + 6$	$g < k$	$6u + 5$	$16u + 13$	6, 14, 22, 30, ...
$S^{(h)}$	≥ 4		$8u + 2$	$8u + 2$	$g < k$		$16u + 5$	2, 10, 18, 26, ...

Table 2: Splitting the reverse path connections into 3+1 sets $S^{(n)}$ for $n = 1, 2, 3$ and $S^{(h)}$ for $h = n \geq 4$

We can further split the sets of reverse path connections $S^{(n)}$ and $S^{(h)}$ into 12 sets $S^{(nC)}$ and $S^{(hC)}$, with n as before and C being the destination color (Table 16) of the connection. The per-color index is $s = 0, 1, 2, 3, \dots$

7 Relation between odd numbers

Diagram 2 shows the mapping between g and k using table 2. The x -axis corresponds to the N_k -axis of Fig. 1 and vertical connections correspond to the lines of sets R and B. Connections to members of set G can be handled as endpoints and are shown as gray circles. Pink connections are the open entry points for $n \geq 4$.

The graph shows that all open connections are endpoints and/or connect to the set $S^{(h)}$. As any connection in $S^{(h)}$ connects strictly from left ($g < k$), we can iterate over all $S^{(h)}$ points and close the open connections.

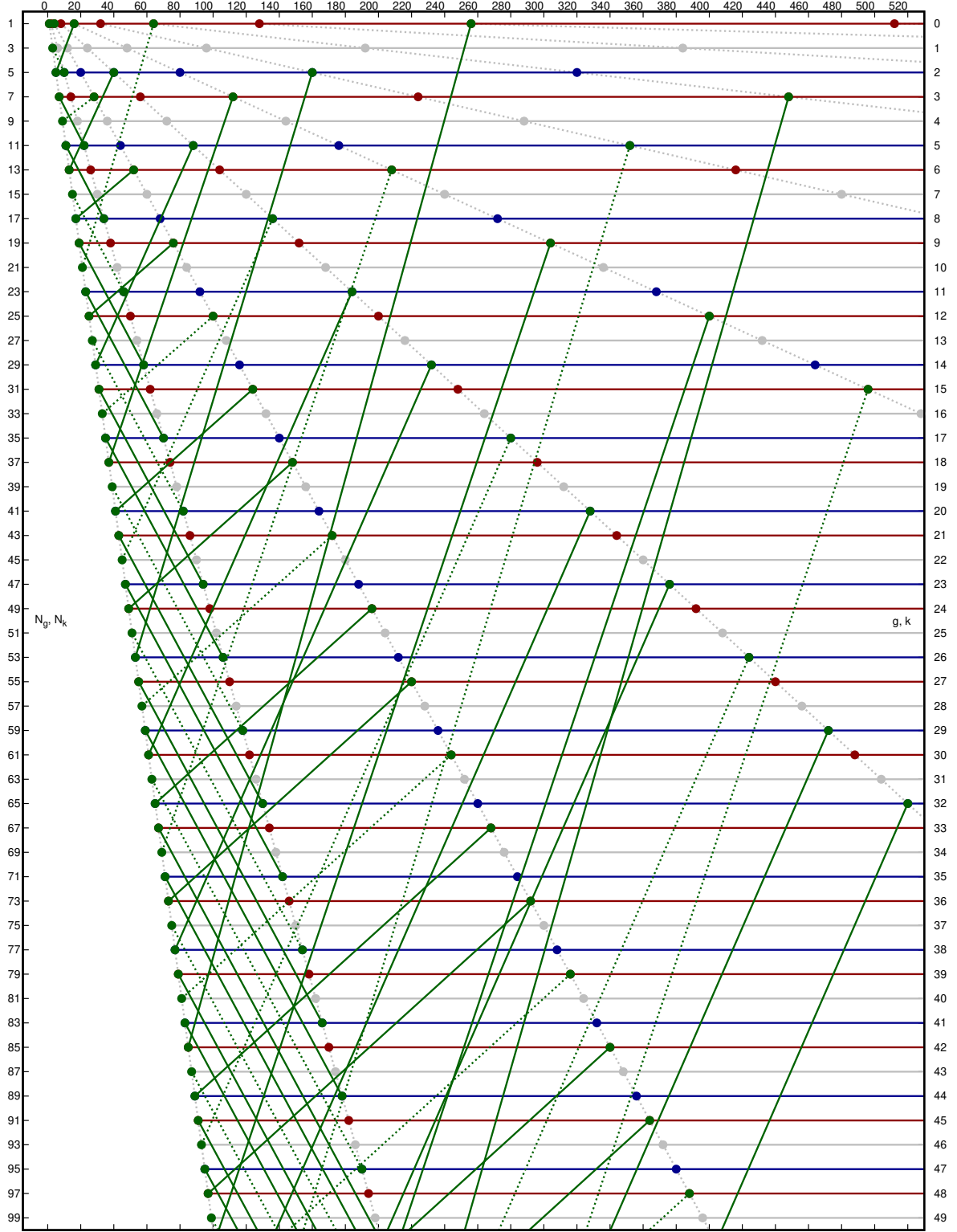


Figure 1: N_g - A_k - N_k relation, best viewed in landscape orientation

7.1 Proposed solution using graph theory

By further splitting and/or using the same base (2 or 3) for g and k in table 2 resp. Src and Dst in table 3 the mapping could be translated to some kind of adjacency matrix and/or be solved by iteration. The distinction between $S^{(n)}$ and $S^{(h)}$ seems to be reasonable in the way that within $n < 4$ the connections

Set	n	C-Map	Src	Dst
$S^{(1G)}$	1	$B_s \mapsto G_s$	$3s$	$2s$
$S^{(1R)}$	1	$B_s \mapsto R_s$	$3s + 1$	$2s + 1$
$S^{(1B)}$	1	$B_s \mapsto B_s$	$3s + 2$	$2s + 1$
$S^{(2R)}$	2	$R_s \mapsto R_s$	$3s$	$4s$
$S^{(2G)}$	2	$R_s \mapsto G_s$	$3s + 1$	$4s + 1$
$S^{(2B)}$	2	$R_s \mapsto B_s$	$3s + 2$	$4s + 2$
$S^{(3R)}$	3	$B_s \mapsto R_s$	$3s$	$8s + 2$
$S^{(3B)}$	3	$B_s \mapsto B_s$	$3s + 1$	$8s + 4$
$S^{(3G)}$	3	$B_s \mapsto G_s$	$3s + 2$	$8s + 7$
$S^{(hR)}$	≥ 4	$? \mapsto R_s$	$?$	$8s + 6$
$S^{(hG)}$	≥ 4	$? \mapsto G_s$	$?$	$8s + 3$
$S^{(hB)}$	≥ 4	$? \mapsto B_s$	$?$	$8s$

Table 3: Subsets of reverse path connections per destination color

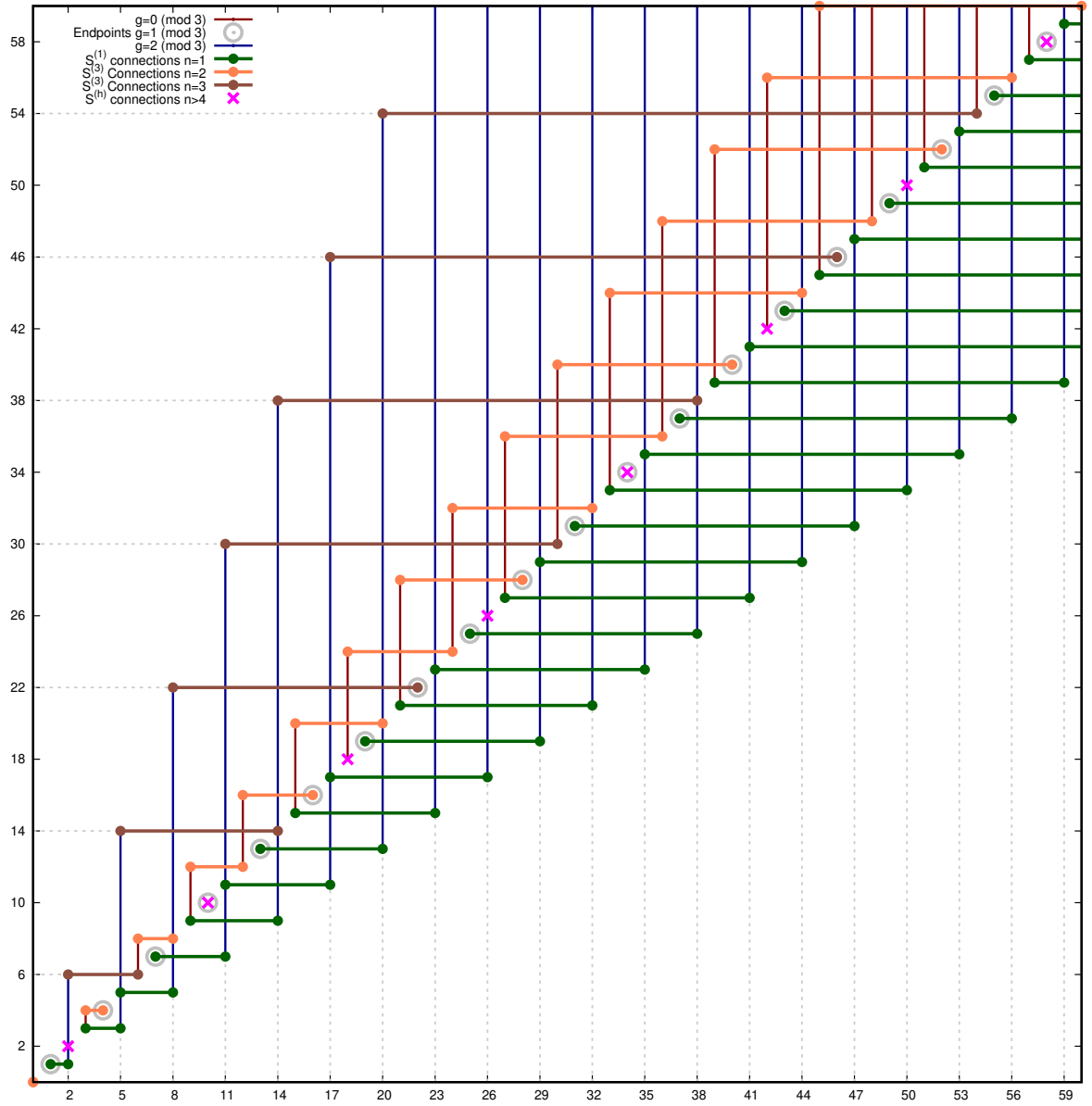


Figure 2: g-k connections

per N_g are limited to 2 in set R and 3 in set B. Also there is no mapping from 1 within that range (except the self-mapping at 1).

The paths to the endpoints could be shortened succesively to get a simpler graph.

7.2 Other possible solutions

- The equations in 2 can be grouped to vectors on which rotation and shift operations are applied so there might be a solution using linear algebra.

- Fig. 1 shows that A_k splits the set of odd numbers into periodic intervals.

- Some mapping to complex numbers could handle the rotation.

We suggest to handle $n = 1, 2, 3$ separately using graph theory, although in terms of intervals there is no such distinction.

8 Summary

It has been shown that the reverse tree started from 1 has only the trivial cycle. If there already exists a proof that all odd numbers are connected somehow, the provided lemmas show that the reverse tree is built without gaps or loops.

9 Acknowledgements

Thank you!

10 Sources

Additional material used for generating the graphs with gnuplot can be found on
<https://github.com/maaaax/collatz>