# Prime Analysis in Binary

Brandynne Cho
Saint Mary's College of California

September 17th, 2012

> "The best number is 73. [...] 73 is the 21st prime number. Its mirror, 37, is the 12th, and its mirror, 21, is the product of multiplying - hang on to your hats - 7 and 3. [...] In binary, 73 is a palindrome: 1001001, which backwards is 1001001."
> -*Sheldon Cooper, The Big Bang Theory*

## 1 Introduction

Working with prime numbers in binary is of great ease as binary is already the system of counting numbers that computers use. The binary system includes two numbers, 0 and 1. To transition from counting in binary, one simply can add the correct sum of a decimal number, a number in base 10, in the correct place holders of the base 2 system. For example, to create the number 5 in binary, one must place a 1 in the $4 = 2^2$ place holder, a 0 in the $2 = 2^1$ place holder, and a 1 in the $1 = 2^0$ place holder. So $5 = 101_2$.

A relationship is formed when one of the binary digits of a prime can be changed to create another prime [2]. For instance, changing the second digit of $17 = 10001_2$ forms the prime number $19 = 10011_2$. These primes now have a relationship or are "neighbors". ((ADD a string of three or four here – 17 to 19 to XX to YY to ZZ. These primes form a neighborhood or "cluster"of at least five primes.)). It seems that most primes have neighbors, and lie in clusters. Some primes, such as $11 = 1011_2$ and $127 = 1111111_2$, a one digit change cannot be made to form another prime. We call these "isolated"primes.

For the primes that are not isolated, their relationships can be tracked on a graph. This graph, Figure 1, does not include all primes, but many of them. The graph is created through connections of a prime's neighbors and the neighbors that the primes neighbors have. For instance, 5, which is the minimum (i.e. smallest) prime for this cluster [1], has neighbors 7 and 13. 7 has no other neighbors besides 5, so we must move on to 13. 13 has another neighbor, 29, and another connection is made. Next, 29's other neighbors are 13, 31, 61. We've already seen 29, so next 61 is looked at. This process continues and forms our very large main graph.

Figure 1: The beginning of the main graph.

# 2 Does the graph contain all primes?

The graph in Figure 1 does not contain all of the primes. Why was described first by [[REFERENCE]]. For a prime $p$ suppose that its binary representation is $\sum_{i=0}^{n} e_i 2^i$, where the $e_i$'s are 0 or 1. Because $2 \equiv -1 \pmod 3$, we have

$$p \equiv \sum_{i=0}^{n} e_i (-1)^i \pmod 3.$$

Because $p > 3$ is prime it is not divisible by 3, and so this value is not 0. This suggest examining the value $\delta(p) = \sum_{i=0}^{n} e_i (-1)^i$, which is not a multiple of 3.

Now suppose $q$ is a neighbor of $p$. Then $q = p \pm 2^j$ for some $j$, and hence $\delta(q) = \delta(p) \pm (-1)^j$. Because neither value is divisible by 3, we see there is some $k$ so that $\delta(p), \delta(q) = 3k+1, 3k+2$ in some order.

Therefore, write $D(p)$ for the value $k$ such that $\delta(p) = 3k+1$ or $3k+2$. We have shown that if $p$ and $q$ are neighbors, then $D(p) = D(q)$. Conversely, if $D(p) \neq D(q)$ it is impossible for $p$ and $q$ to be neighbors.

As an example, consider $p = 37$ and $q = 41$. Since $37 = 100101_2$, $\delta(37) = 1 = 0 * 3 + 1$, so $D(37) = 0$. Thus 37 can only be neighbors with primes of $D$-value 0. On the other hand, because $41 = 101001_2$, $\delta(41) = -1 = (-1) * 3 + 2$, so $D(41) = -1$, which means that 41 cannot be neighbors with any primes of $D$-value 0. In particular, it is impossible to wander through the cluster of primes containing 37 and arrive at 41 – it must be in a separate cluster. [[GIVE A PRIME WITH D = 1, D=2 and D=-2]].

BREAK

If we think of a large prime $p$ in binary as consisting of leading and tailing 1's, with the other binary digits "randomly" distributed between 0's and 1's, then then there are $n-2$ choices of 0's and 1's, where $n$ is the number of digits in $p$, i.e., its "digit length". Looking at $\delta(p)$, on average we would expect the 1's to fall into spots contributing $+1$'s about as often as into those contributing $-1$. That is, we'd expect $\delta(p)$, and so $D(p)$ to be near 0. On the other hand, to have a prime with large $D$, there will have to be many 1's and few $-1$'s. This will be less common. Similarly it is less common to find primes with large negative $D$-values. We see this in the following chart. Notice $D$-values do appear to follow the normal distribution, however, there is a slight skew (Figure 2).

Figure 2: Percentages of primes per Delta (-2, 2) up to digit length 25.

# 3 Is the graph infinite?

[[Explain we are looking at the primes with $D$-value 0]]

When thinking of numbers in their infinite capacity, it is important to realize infinity's capacity, or lack thereof. As a child one is taught to think of number on a number line. In reality, the number of numbers is infinite. Numbers are much more like stars which stretch forever, an eternalness of light years away. As large as this seems, there are still multiple ways to view the infinite graph of prime numbers.

One way to think of an infinite graph is by means of a straight line up, aiming to find the biggest prime. This view of infinite prime numbers can move very quickly. Table 1 gives data showing the number of steps taken in the graph beginning at 5 until a prime larger than maximum $p = 2^b + 1$ is found. The number of steps travelled to reach the maximum is incredibly quick.

[[EXPLAIN THIS SUGGESTS TO US THAT THE GRAPH GOES ON FOREVER – IS INFINITE]]

| $b, p = 2^b + 1$ | base 10 equivalence | number of steps |
|---|---|---|
| 20 | 1 million | 77 |
| 30 | 1 billion | 233 |
| 50 | quadrillion (15 digits) | 607 |
| 100 | nonillion (30 digits) | 2843 |

Table 1: Total steps for the fastest path up to max prime, $p = 2^b + 1$.

### 3.1 What is the shape of the graph?

An alternate way to think of an infinite graph is by keeping track of all the connecting primes in a big, tall and wide tree-like graph. This is similar to scooping up or tallying the primes wanted.

The shape of the main graph appears to be more like the shape of a tree that expands the further it grows upward. The further the graph travels toward infinity, the more interconnected and complex it becomes.

[[PROBABLY INCLUDE A PICTURE??]]

### 3.2 Are there loops?

There are many loops in the graph. In the main graph. The first loop begins at 101 and splits to 103 and 229, where 71 and 199 follow 103 and 197 and 199 stem from 229. This loop, of six steps, is small, but the graph has many complex loops of greater sizes; as one loop is discovered, its discovery leads to more loops interconnecting with past loops. Figure 3 shows the loops in the graph for primes less than 1000. It is interesting to note that the steps taken to get around the loops will always be even. This is because for every step taken away from the base number a 0 is changed to a 1, or a 1 to a 0, in the binary form of a number. This step must eventually be reversed to us to get back. Since each step has an opposite, the total number of steps is even.

Figure 3: The main graph from 101 until 859 and 1847. DE denotes a Dead End where the graph continues but eventually ends without contribution to loops.

Outside of the main graph there are isolated primes, pairs, triples, quads, etc. Some of these excluded primes form loops themselves, but nonetheless do not contribute to loops in the main graph. Figure 4 samples a few isolated primes, a pair, and a quad.

Figure 4: The graph for $\delta = 0$ does include "Loaners" not in the main graph, consisting of isolated primes, pairs, quads, etc.

## 3.3 How many neighbors does each prime have?

When trying to determine whether the graph is finite or infinite, it is critical to consider the number of neighbors each prime has. In an infinite segment of the graph, there must be a series of primes all of whom have at least two neighbors each. So if most primes have fewer than two neighbors, then one would think the graph to be not finite, and so mostly consisting of isolated clusters. This, if the graph is to be infinite, one expects primes to have, on average, at least two neighbors.

Theoretically the number of neighbors suggests that the graph is infinite. The Prime Number Theorem [[REFERENCE]] states that the number of primes smaller than $n$ is approximately $\frac{n}{\ln n}$. So the fraction of numbers near $n$ that are prime is approximately $\frac{1}{\ln n}$. **Am I supposed to do all this arithmetic here?** [[YES]] Thus, the theoretical number of neighbors is approximately 2.16.

Contrastingly, we calculated the average number of neighbors for primes with $D$-value 0 for digit length 5 up to digit length 200. The data in Figure 5 shows the average number of neighbors declining from 2 at digit length 5 to about 1.90 at digit length 200. This data suggests that the graph is finite because, while the average number of neighbors is less than 2, the chances for dead ends increases.

Figure 5: The average number of neighbors per prime per digit length for primes with digit length 5 to 200.

It is possible that the error in the number of neighbors comes from the different deltas. Most data focused on is according to $D = 0$, therefore this inconsistency could be related to lack of inclusion of other deltas. However, it is more difficult to reach other deltas as their graphs begin at greater numbers, corresponding to the more extreme deltas away from 0.

### 3.4 What is the graph made up of?

The graph is made up of clusters that vary in size. The cluster sizes range from isolated primes, pairs, other smaller groups, up to a large mega-cluster. The isolated and pairs seem to make up about 20 percent of all primes; 15 percent and 5 percent, respectively. The mega-cluster is a very large portion of the graph and appears to make up at least 75 percent of primes. This is a significant chunk of numbers. Figure 2 depicts this data.

| Cluster Size and Percentage of Graph | | | |
|---|---|---|---|
| $b, p = 2^b + 1$ | 30 | 50 | 100 |
| 1 | 13.07 | 14.93 | 15.91 |
| 2 | 3.81 | 4.29 | 4.09 |
| 3 | 1.96 | 1.81 | 1.64 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 2500 | 78.96 | 76.96 | 76.36 |

Table 2: Cluster sizes and percentages for primes with digit length 30, 50, and 100 for $\delta = 0$

The distribution of the mega cluster is interesting to note. (See Figure 6.) Within the 75 percent of primes in the mega-cluster, 90 percent of primes have either one, two, or three neighbors. Specifically, 25.77 percent have one neighbor, 38.03 percent have two neighbors, and 24.74 percent have three neighbors. [[DOES THIS MEAN THE REAMAING 11% HAVE MORE THAN THREE NBRS??]] It is critical to have enough neighbors to avoid dead ends. In the mega-cluster, since 74.23 percent of primes have more than one neighbors, we expect the branching of the cluster to continue towards infinity.

The program for the cluster size and percentage of the graph was also run for other delta values -2 (Table 3), -1 (Table 4), 1 (Table 5), and 2 (Table 6). The data varied per digit length, but was similar to the $D$-value 0 case, with approximately 20 percent of the graph was made up of isolated

Figure 6: In the mega-cluster, the distribution of the number of neighbors per prime.

and pairs and approximately 75 percent of the graph was made up of the delta's mega-cluster.

| Cluster Size and Percentage of Graph | | | |
|---|---|---|---|
| $b, p = 2^b + 1$ | 30 | 50 | 100 |
| 1 | 20.24 | 17.52 | 15.58 |
| 2 | 3.42 | 3.92 | 4.02 |
| 3 | 2.9 | 2.04 | 1.86 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 1500 | 69.4 | 73.98 | 76.3 |

Table 3: $\delta = -2$

| Cluster Size and Percentage of Graph | | | |
|---|---|---|---|
| $b, p = 2^b + 1$ | 30 | 50 | 100 |
| 1 | 13.6 | 15.26 | 15.34 |
| 2 | 3.84 | 4.32 | 4.28 |
| 3 | 2.22 | 1.6 | 2.04 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 1500 | 77.6 | 76.76 | 76.26 |

Table 4: $\delta = -1$

| Cluster Size and Percentage of Graph | | | |
|---|---|---|---|
| $b, p = 2^b + 1$ | 30 | 50 | 100 |
| 1 | 17.44 | 17.02 | 17.06 |
| 2 | 3.22 | 4.1 | 4.38 |
| 3 | 1.66 | 1.76 | 2.02 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 1500 | 74.58 | 74.34 | 74.36 |

Table 5: $\delta = 1$

| Cluster Size and Percentage of Graph | | | |
|---|---|---|---|
| $b, p = 2^b + 1$ | 30 | 50 | 100 |
| 1 | 25.86 | 18.48 | 16.02 |
| 2 | 3.06 | 3.72 | 4.4 |
| 3 | 2.98 | 2.22 | 2.2 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 1500 | 62.94 | 72.92 | 75.2 |

Table 6: $\delta = 2$

# 4 Other Findings

There are a few other findings concerning the graph as a whole.

An approximate function to determine the number of isolated primes per digit length has been found. Equation 1 approximates the growth of the number of isolated primes by digit lenth $n$.

$$I(n) = \begin{cases} \frac{1.91^n}{122} & \text{while n is odd} \\ \frac{1.91^n}{103} & \text{while n is even} \end{cases} \quad (1)$$

Another function that determines the decay of the clusters of primes is noted. The decay begins with number of isolated primes and decreases until the largest group(s). (Generally there is only one large group.) If the number of isolated primes is known, Equation 2 will approximate the decay of the number of other groups, where k comes from the previous number of groups. However, for closer approximations this function changes based on the digit length. Figure 7 shows the values of k when the digit length is 23 until 15. [[why 23 down to 15??]]

$$D(k) = k^{0.9} \quad (2)$$

Figure 7: k values for digit lengths 23 until 15.

Tracking the relationships of the graph by putting the values into a matrix allows for finding the rank of the graph. Up to digit length 23 (about 8.4 million in base 10), the rank of the matrix is approximately 82 percent.

# 5 Concluding Thoughts

There are still more questions that arise regarding the clusters. Such as, exactly how big is the mega cluster? Is the mega cluster really multiple large clusters? **I'm not sure what else to conclude/how to go about this.**

# References


[1] Hartley, M. I. (2002). Partitions in the Prime Number Maze. *Acta Arithmetica, 105*(3), 227-38.

[2] Paulsen, W. (2000). The Prime Number Maze. *The Fibonacci Quarterly, 40*(3), 272-79.