7-29-2011

Structure and Randomness of the Discrete Lambert Map

Jing Jing Chen
Pomona College

Mark Lotts
Randolph-Macon College

Structure and Randomness of the Discrete Lambert Map

J.Chen and M. Lotts

Adviser: Joshua Holden

Mathematical Sciences Technical Report Series
MSTR 11-02

July 29, 2011

Department of Mathematics
Rose-Hulman Institute of Technology
http://www.rose-hulman.edu/math

Fax (812)-877-8333 Phone (812)-877-8193

Structure and Randomness of the Discrete Lambert Map

JingJing Chen
Pomona College

Mark Lotts
Randolph-Macon College

Advisor: Joshua Holden
Rose-Hulman Institute of Technology

29 July 2011

Produced at the Rose-Hulman Institute of Technology Mathematics REU,
sponsored by NSF grant DMS-1003924

Abstract

We investigate the structure and cryptographic applications of the Discrete Lambert Map (DLM), the mapping $x \mapsto xg^x \bmod p$, for p a prime and some fixed $g \in (\mathbb{Z}/p\mathbb{Z})^*$. The mapping is closely related to the Discrete Log Problem, but has received far less attention since it is considered to be a more complicated map that is likely even harder to invert. However, this mapping is quite important because it underlies the security of the ElGamal Digital Signature Scheme. Using functional graphs induced by this mapping, we were able to find non-random properties that could potentially be used to exploit the ElGamal DSS.

1 Introduction

In addition to encrypting and decrypting sensitive information, cryptography can also be used to help a message's recipient verify the identity of the sender. These protocols are known as digital signature schemes. Much like other cryptosystems, the security of these digital signature schemes relies on the difficulty of exploiting their underlying mathematical structure. Thus, problems generally considered to be computationally intractable, such as integer factorization and the Discrete Logarithm Problem (DLP), often serve as the basis for such schemes.

1.1 Motivation

One such scheme that is particularly important to our topic is the ElGamal Digital Signature Scheme (DSS). Suppose Alice needs to send a message M to Bob.

In order for Bob to be sure that Alice was indeed the sender of the message, Alice must sign the message in such a way that Bob can easily verify her identity. To accomplish this using the ElGamal DSS, Alice starts by choosing a large prime p and a secret signing key $x \in \mathbb{Z}$, selected randomly from $\{0, \ldots, p-2\}$. Alice then computes α a primitive root mod p, a generator of the cyclic group $(\mathbb{Z}/p\mathbb{Z})^*$, and releases the public key (p, α, y), where $y \equiv \alpha^x \pmod{p}$.

To actually sign M, Alice selects a nonce k from $\{0, \ldots, p-1\}$ where $\gcd(k, p-1) = 1$. Alice's signature (r, s) is then computed such that $r \equiv \alpha^k \pmod{p-1}$ and $s \equiv k^{-1}(M - xr) \pmod{p}$.

Bob then receives M from Alice, and wishes to verify her identity based on both the message's signature and Alice's public key. Bob starts the verification process by computing $v_1 \equiv y^r r^s \pmod{p}$ and $v_2 \equiv \alpha^M \pmod{p}$. If $v_1 \equiv v_2 \pmod{p}$, then Bob concludes that Alice was the sender of the message.

In order to forge Alice's signature, Frank must be able to find some v_1 and v_2 such that $v_1 \equiv y^r r^s \equiv \alpha^M \pmod{p}$, where M is the message on which Frank wants to forge Alice signature. Frank knows Alice's public key (p, α, y), but without Alice's secret signing key x, he cannot compute a valid s. This leaves Frank with a few options. The first option is to fix r and rearrange the equation in order to solve for s. This gives the equation

$$r^s \equiv (y^r)^{-1}\alpha^M \pmod{p}.$$

However, solving this equation for s would involve calculating discrete logarithms. Thus, this attack is not feasible since the DLP is considered to be a sufficiently hard problem. Another variation of the ElGamal DSS involves fixing s and solving for r. This gives the equation

$$y^r r^s \equiv \alpha^M \pmod{p}.$$

Although solving this equation is similar to solving the DLP, it is actually a slightly different problem. Whereas the DLP is based on the difficulty of inverting the map

$$x \mapsto g^x \bmod p,$$

for a fixed g in $\{1, \ldots, p-1\}$ and a prime p, the security of this variation of the ElGamal DSS is based on the difficulty of inverting the map

$$x \mapsto xg^x \bmod p.$$

Due to this map's resemblance to the Lambert W function [2], we will refer to this map as the Discrete Lambert Map (DLM). Although the DLP has been studied at great lengths, the DLM has received virtually no attention. This lack of previous work might be due to the fact that many people consider inverting the DLM to be more difficult than the DLP, but because of the implications that the DLM has for the security of the ElGamal DSS, we believe that it is important to study and analyze its behavior. As a result of our graph-theoretic and statistical methods for analyzing the DLM, we discovered various non-random structures in the functional graphs produced by the mapping. For example, we fully understand fixed points and how power residues determine which nodes can map to one another. However, it remains to be determined whether these structures can be exploited to break the ElGamal DSS, or whether they are simply patterns that occur frequently in random graphs.

1.2 Previous Work

While previous work on the discrete logarithm problem is abundant, the discrete Lambert problem has not seen much investigation. Since the function $x \mapsto xg^x \pmod{p}$ takes the form of a more embellished version of discrete exponentiation, it is assumed to be a more difficult problem to invert. However, its presence in the ElGamal DSS and the relevance of methods used to study similar maps using functional graphs make it a promising object of exploration.

Much analysis has been done on the study of mapping the discrete logarithm using functional graphs. The first to examine the graphs statistically was Lindle [6], who was later followed by Hoffman on statistical parameters and comparisons with random functional graphs [5]. Hoffman's code for generating relevant data for statistical analyses of permutations has been adopted and modified for use in our statistical investigations.

Friedrichsen, Larson and McDowell [4] studied the structure of the self-power map, $x \mapsto x^x \pmod{p}$, which also appears in a version of the ElGamal DSS, using the methods set forth by Hoffman, Cloutier and Holden [1]. Their findings provided inspiration and served as a model for our own proceedings.

2 Background and Methods

2.1 Functional Graphs

Functional graph. A *functional graph* is a directed graph in which each vertex, or *node*, has exactly one edge directed out from it. A functional graph can therefore be realized as a function mapping its domain onto itself.

The functional graph of the Discrete Lambert Map $f(x) = xg^x \pmod{p}$ consists of nodes $\{1, ..., p-1\}$ and directed edges from x to $f(x)$.

By studying the Discrete Lambert Map in functional graph form, we can more readily observe the basic behavior of the function through graph theoretic properties of the visual mapping. Some characteristics of interest regard the number and size of connected components, properties of cycles and fixed points, as well as terminal and image nodes.

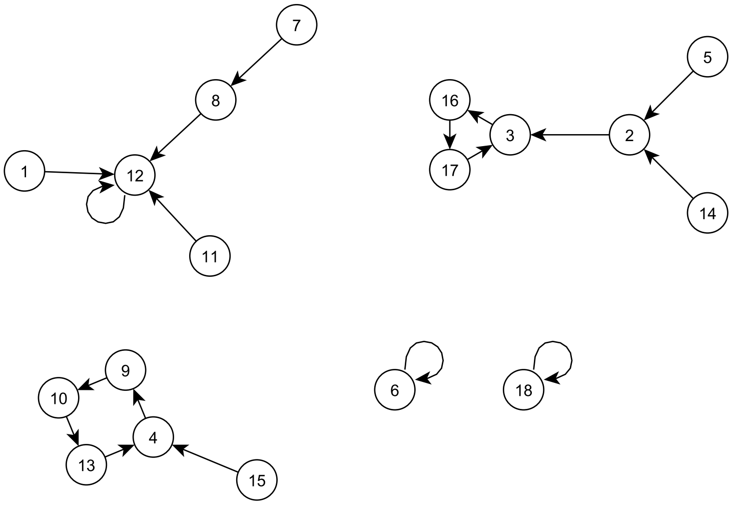


Figure 1: $x \mapsto x12^x \pmod{19}$

Node. An *image node* is a node x such that $x = f(y)$ for some node y. A *terminal node* is a node z for which there does not exist a node w where $f(w) = z$. In graph theory terms, image nodes have arrows pointing in to them, while terminal nodes do not.

Connected component. A *connected component* is a set of nodes connected by edges. Components are disjoint and partition the set of all nodes.

Cycle. An *n-cycle* is a set of n nodes $\{x_0, x_1 = f(x_0), ..., x_{n-1} = f(x_{n-2})\}$ such that $x_n = f(x_{n-1}) = x_0$. A 1-cycle is also known as a *fixed point*, i.e. a node that maps to itself.

In Figure 1, nodes $4, 9, 10, 13$ form a 4-cycle, and nodes $3, 16, 17$ form a 3-cycle. The fixed points are $6, 12, 18$.

Tail. A *tail* is a set of nodes whose directed path leads into a cycle.

m-ary graph. An *m-ary graph* is a functional graph where, for a fixed m, all image nodes in the graph have in-degree m.

2.2 Number Theory and Group Theory

The domain of the Discrete Lambert Map is the set of integers $\{1, ..., p-1\}$, closed under multiplication modulo p, where p is prime, also known as the algebraic group $(\mathbb{Z}/p\mathbb{Z})^*$. This group is cyclic, which means there exists an element $g \in (\mathbb{Z}/p\mathbb{Z})^*$ such that $\{g, g^2, ..., g^{p-1}\} = \{1, 2, ..., p-1\}$. g is known as a generator, or a *primitive root.*

Theorem 1. *Let ϕ denote the Euler totient function. If p is prime, then there exist $\phi(p-1)$ primitive roots modulo p.*

Proof. See Theorem 2.36 of [7]. □

Order. The *order* of an element $g \in (\mathbb{Z}/p\mathbb{Z})^*$, denoted $ord_p(g)$, is the smallest positive integer $n \in \{1, ..., p-1\}$ such that $g^n \equiv 1 \pmod{p}$. The order of g divides $p-1$, the order (size) of the group $(\mathbb{Z}/p\mathbb{Z})^*$. Primitive roots must have order $p-1$.

Power Residue. An element g is an n^{th} *power residue* if there exists $a \in (\mathbb{Z}/p\mathbb{Z})^*$ such that $g \equiv a^n \pmod{p}$. When $n = 2$, we call g a quadratic residue. There are $\frac{p-1}{2}$ quadratic residues in $(\mathbb{Z}/p\mathbb{Z})^*$.

Subgroup. A *subgroup* H of a group G is a subset of G that itself satisfies the group properties:

- *Closure*: For all $g, h \in G$, $gh \in G$.
- *Identity*: For all $g \in G$, there exists an identity element $e \in G$ such that $eg = ge = g$.
- *Inverse*: For all $g \in G$, there exists $h = g^{-1} \in G$ such that $gh = e$.
- *Associativity*: For all $x, y, z \in G$, $(xy)z = x(yz)$.

The order of subgroup divides the order of the group.

Coset. Let H be a subgroup of a group G. For $x \in G$, xH is the set of elements obtained by left multiplication of every element of H by x, known as a *left coset* of H. Similarly, Hx is a right coset obtained by right multiplication by x. (Here $(\mathbb{Z}/p\mathbb{Z})^*$ is a commutative group under multiplication, so xH=Hx.)

We will utilize the notion of cosets in the formulation of a result about the connected components of the Discrete Lambert Map.

2.3 Statistics

After determining the basic behavior of the functional graphs induced by the DLM, we use statistical methods to compare characteristics of the DLM graphs to the expected characteristics of a random functional graph. In this regard, the paper by Flajolet and Odlyzko was extraordinarily useful in helping us determine which graph characteristics would be worthy of examination. The following are the characteristics that we deemed important to analyzing the behavior of the graphs.

Total Sums

Number of Connected Components. The number of connected components in a functional graph.

Number of Cyclic Nodes. The number of nodes that are in a cycle of any length.

Number of Image Nodes. The number of nodes that have preimages.

Number of Terminal Nodes. The number of nodes that have no preimages.

Number of Fixed Points. The number of nodes that map to themselves.

Total Sums As Seen From a Node

Total Cycle Length. For each cycle, multiply the length of the cycle by the number of nodes that reach the cycle by a connected path. Add the results of the multiplications from each cycle.

Total Distance to Cycle. For each node, count the number of edges that must be crossed before reaching a cyclic node. Add the results of the additions for all nodes in the graph.

Maximal Values

Maximum Cycle Length. The number of nodes in the largest cycle.

Maximum Tail Length. The number of nodes in the longest tail.

After identifying the most relevant characteristics to examine, we chose twenty primes for which to gather information. These twenty primes were chosen based on the factorization of $p-1$ for each prime p.

Safe Prime. A prime p is a *safe prime* if $\frac{p-1}{2}$ is also prime.

The twenty primes chosen for examination were the first twenty safe primes greater than 40,000. Primes around 40,000 were chosen based on the success of previous work by Hoffman, who used a similar number of functional graphs in his analysis [5]. After finding the candidate primes, we modified Hoffman's code to generate and gather data on the functional graphs induced by the DLM for values of g from 2 to $p-2$ for each prime. Since we already know the exact structure of the DLM when $g = 1$ and $g = p-1$, we excluded those values of g from our analysis. We considered graphs induced by values of g with different orders separately since the order of g greatly influences the structure of the functional graphs. Thus, safe primes were ideal since, excluding $g = 1$ and $g = p - 1$, g can only have one of two orders, $\frac{p-1}{2}$ and $p - 1$, and the number of graphs with those orders is equal. If we used prime that were not safe, there could potentially be many, many orders, all of which might occur in different amounts from $g = 2$ to $g = p - 2$. This would unnecessarily complicate the statistical process.

After generating and gathering date for the DLM graphs, we used the asymptotic formulas in Flajolet and Odlyzko's paper to calculate the expected values of these characteristics for a random functional graph [3]. However, instead of generating the expected values using $p - 1$ nodes, we used our knowledge of the structure, size, and minimum number of connected components to make our expectations for a subgraph consisting of at least one connected component which we could then multiply to get the total expected value for each of the summed graph characteristics. A full description of this process can be found in the results section.

The last step of the statistical methods was to import both the observed and expected values for the graph characteristics into Minitab to process the data and perform statistical tests. The three most important tests we will use to process the data are the probability plot, the t-test, and tests for normality.

3 Results

3.1 Basic behavior

There are a few basic properties of the functional graph of $x \mapsto xg^x \pmod{p}$ that are evident upon close inspection:

1. For every value of g, $1 \mapsto g$.
2. For any prime p and every value of g, $p - 1$ is a fixed point.
3. When $g = 1$, every $x \in (\mathbb{Z}/p\mathbb{Z})^*$ is a fixed point.
4. In general, the graphs are not m-ary.

In addition to the values of 1 and $p - 1$, we can determine the images of other nodes based on the properties of g.

3.2 Images of $p - 2$ and $\frac{p-1}{2}$

Proposition 1. *For any prime p, when $g = 2$, then $(p - 2) \mapsto (p - 1) \pmod{p}$.*

Proof. Setting $g = 2$, we see that $(p-2) \mapsto (p-2)(2)^{(p-2)} \pmod{p}$. Simplifying the right hand side of this equation, we see that

$$\begin{aligned}
(p-2)(2)^{(p-2)} &\equiv (p-2)(2)^{((p-1)-1)} \\
&\equiv (p-2)(2)^{-1} \\
&\equiv (-2)(2^{-1}) \\
&\equiv -1 \equiv p-1 \pmod{p}.
\end{aligned}$$

□

Proposition 2. *For any prime p, when $g = p-2$, then $(p-2) \mapsto 1 \pmod{p}$.*

Proof. Setting $g = p-2$, we see that $(p-2) \mapsto (p-2)(p-2)^{(p-2)} \pmod{p}$. We see that the right hand side can be written as follows:

$$(p-2)(p-2)^{(p-2)} \equiv (p-2)^{(p-1)} \equiv 1 \pmod{p}.$$

□

Proposition 3. *If g is a quadratic residue, then $\frac{p-1}{2} \mapsto \frac{p-1}{2} \pmod{p}$.*

Proof. By Euler's criterion (see Theorem 11.3 of [8]), we know that $g^{\frac{(p-1)}{2}} \equiv 1 \pmod{p}$ if and only if g is a quadratic residue mod p. When g is a quadratic residue mod p, we have:

$$\frac{(p-1)}{2} \mapsto \frac{(p-1)}{2}(g)^{\frac{(p-1)}{2}} \equiv \frac{(p-1)}{2}(1) \equiv \frac{(p-1)}{2} \pmod{p}.$$

□

Proposition 4. *If g is not a quadratic residue, then $\frac{p-1}{2} \mapsto \frac{p+1}{2} \pmod{p}$.*

Proof. Similar to the preceding proof, by a corollary to Euler's criterion, if g is not a quadratic residue (also called a quadratic non-residue) mod p, then $g^{\frac{(p-1)}{2}} \equiv -1 \pmod{p}$. Thus, examining the discrete Lambert map, when g is not a quadratic residue mod p, we have:

$$\frac{(p-1)}{2} \mapsto \frac{(p-1)}{2}(g)^{\frac{(p-1)}{2}} \equiv \frac{(p-1)}{2}(-1) \equiv \frac{(p+1)}{2} \pmod{p}.$$

□

3.3 Fixed Points

We investigate the occurrence of fixed points in the functional graphs of the Discrete Lambert Map. Given a value of g for a known prime p, we can determine precisely the nodes that are fixed points.

Lemma 1. *Given such a functional graph, x is a fixed point if and only if $g^x \equiv 1 \pmod{p}$.*

Proof.
($\Rightarrow$)

Assume x is a fixed point. Thus, $x \equiv xg^x \pmod{p}$. Multiplying on the left by the multiplicative inverse of x, we have:

$$\begin{aligned} x^{-1}x &\equiv x^{-1}xg^x \pmod{p} \\ 1 &\equiv g^x \pmod{p} \end{aligned}$$

($\Leftarrow$)

Assume $g^x \equiv 1 \pmod{p}$. This gives us $xg^x \equiv x(1) \equiv x \pmod{p}$. Thus, the map $x \mapsto xg^x \bmod p$ maps x to itself since we have shown that $x \equiv xg^x \pmod{p}$ when $g^x \equiv 1 \pmod{p}$. Therefore, x is a fixed point. □

Proposition 5. *Given a functional graph of $x \mapsto xg^x \pmod{p}$, the fixed points are precisely the multiples of the order of g.*

Proof. Suppose $x \in (\mathbb{Z}/p\mathbb{Z})^*$ is a fixed point. Then by our Lemma, $g^x \equiv 1 \pmod{p}$. Since the order of g is the smallest integer n such that $g^n \equiv 1 \pmod{p}$, it must be that the order of g divides x. Thus, x must be a multiple of n, the multiplicative order of g. □

We observe that the multiples of n are also the logarithms of the n^{th} power residues mod p, where the base of the logarithm is a primitive root of $(\mathbb{Z}/p\mathbb{Z})^*$.

Corollary 1. *The number of fixed points of a functional graph of $x \mapsto xg^x \pmod{p}$ is the number of n^{th} power residues* mod p*, where n is the multiplicative order of g.*

Proof. Since the fixed points are precisely the logarithms of the n^{th} power residues, they must be equinumerous. □

Corollary 2. *If g is a primitive root, the only fixed point is $p-1$.*

Proof. We know that the fixed points are the logarithms of the n^{th} power residues mod p, where n is the multiplicative order of g. Since g is a primitive root, we know that its multiplicative order is $p-1$. We also know that 1 is the only $(p-1)^{st}$ power residue modulo p since $g^{p-1} \equiv 1 \pmod{p}$, along with all the multiples of $p-1$ since the exponents are computed modulo $p-1$. Thus, since 1 is the only n^{th} power residue when g is a primitive root, and the logarithm of 1 is $p-1$, the only fixed point of the graph is $p-1$. □

3.4 The Functional Graph of $p-1$

Similar to when $g = 1$, the functional graph of $x \mapsto x(p-1)^x \pmod{p}$ exhibits an entirely predictable and organized structure, as visualized in Figure 2.

Proposition 6. *Let f denote the Discrete Lambert Map of p. Let $g = p-1$. If x is odd, then $f(x) = p - x$, the additive inverse of $x \pmod{p}$. If x is even, then $f(x) = x$ is a fixed point.*

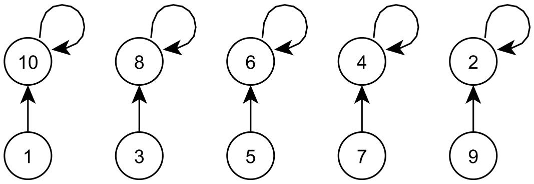


Figure 2: $x \mapsto x10^x \pmod{11}$

Proof. Suppose x is odd. Write $x = 2k+1, k \in \mathbb{Z}$. Since $(p-1)^2 \equiv 1 \pmod{p}$, we have:

$$\begin{aligned} f(x) &\equiv xg^x \\ &\equiv (2k+1)(p-1)^{2k+1} \\ &\equiv (2k+1)((p-1)^2)^k(p-1) \\ &\equiv (2k+1)(p-1) \\ &\equiv 2kp - 2k + p - 1 \\ &\equiv -2k-1 \equiv -x \equiv p - x \pmod{p}. \end{aligned}$$

Suppose x is even. Write $x = 2k, k \in \mathbb{Z}$. Then

$$f(x) \equiv xg^x \equiv (2k)(p-1)^{2k} \equiv (2k)((p-1)^2)^k \equiv 2k \equiv x \pmod{p}.$$

Thus, x is a fixed point. □

Proposition 7. *Let $g = p-1$. Then the functional graph of $f : x \mapsto xg^x \pmod{p}$ is a binary graph that has exactly $\frac{p-1}{2}$ connected components. Furthermore, each connected component consists of precisely one odd terminal node mapped to one even node, which is a fixed point.*

Proof. By previous proofs, we have shown that if x is odd, then its image is $p-x$, and if x is even, then it is a fixed point. This results in precisely the configuration described above. Since each connected component consists of exactly two nodes, there must be $\frac{p-1}{2}$ connected components. □

3.5 Investigations of Power Residues

Proposition 8. *Let $g \in (\mathbb{Z}/p\mathbb{Z})^*$ be an n^{th} power residue. Then, $x \in (\mathbb{Z}/p\mathbb{Z})^*$ is an n^{th} power residue if and only if $xg^x \pmod{p}$ is also an n^{th} power residue.*

Proof. ($\Rightarrow$) Suppose $x, g \in (\mathbb{Z}/p\mathbb{Z})^*$ are nth power residues. Then there exist $a, b \in (\mathbb{Z}/p\mathbb{Z})^*$ such that $a^n \equiv g \pmod{p}$ and $b^n \equiv x \pmod{p}$. Then

$$xg^x \equiv b^n(a^n)^{b^n} \equiv b^n(a^{b^n})^n \equiv (ba^{b^n})^n \pmod{p}.$$

Hence, xg^x is also an nth power residue mod p.

($\Leftarrow$) Suppose $xg^x \pmod p$ is an nth power residue. Then, we can write $xg^x \equiv z^n \pmod p$ for some $z \in (\mathbb{Z}/p\mathbb{Z})^*$. Since we also know that g is an nth power residue, we can write $g \equiv y^n \pmod p$ for some $y \in (\mathbb{Z}/p\mathbb{Z})^*$. Then we have as follows:

$$\begin{aligned} xg^x &\equiv z^n \\ xg^x(g^{-x}) &\equiv z^n(g^{-x}) \\ x &\equiv z^n(y^n)^{-x} \\ &\equiv z^n(y^{-x})^n \\ &\equiv (zy^{-x})^n \pmod p. \end{aligned}$$

Thus, x is also an n^{th} power residue mod p. □

Theorem 2. *If p, n are positive integers, and $\gcd(g,p) = 1$, then g is an n^{th} power residue modulo p iff $g^{\frac{p-1}{d}} \equiv 1 \pmod p$, where $d = \gcd(n, p-1)$. Furthermore, there are exactly d incongruent residues modulo p.*

Proof. See Proposition 9.17 of [8]. □

Let $g \in (\mathbb{Z}/p\mathbb{Z})^*$ have multiplicative order n. Since $\gcd(g,p) = 1$ and $\gcd(\frac{p-1}{n}, p-1) = \frac{p-1}{n}$, then $g^{(p-1)/\frac{(p-1)}{n}} = g^n \equiv 1 \pmod p$ implies that g is an $\frac{p-1}{n}^{th}$ power residue.

The combination of these two properties demonstrate that in the functional graph of the Discrete Lambert Map $x \mapsto xg^x \pmod p$, where g has multiplicative order n, all the $\frac{p-1}{n}^{th}$ power residues are mapped to (and from) each other.

3.6 Properties of Connected Components

Our previous observations about the behavior of certain power residues lead us to findings about the properties of connected components in a given functional graph. Here we give an upper bound on the number of nodes in a connected component of a graph, as well as characteristics of the nodes in a connected component. For the following results, let n denote the multiplicative order of $g \in (\mathbb{Z}/p\mathbb{Z})^*$.

Proposition 9. *Given any g, n is an upper bound on the number of nodes in a connected component containing a $\frac{p-1}{n}^{th}$ power residue.*

Proof. This follows almost immediately from the fact that all $\frac{p-1}{n}^{th}$ power residues are mapped to each other. There exist n of these residues, and at the most extreme, they are all in a single connected component. Thus n is the maximum number of nodes in a connected component which contains a $\frac{p-1}{n}^{th}$ power residue. □

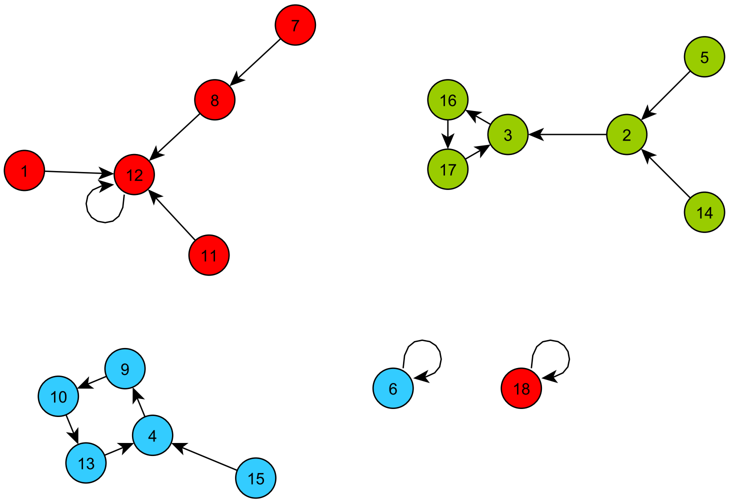
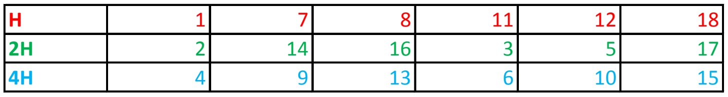


Figure 3: $x \mapsto x12^x \pmod{19}$



H	1	7	8	11	12	18
2H	2	14	16	3	5	17
4H	4	9	13	6	10	15

If g has order n, then it is a $\frac{p-1}{n}^{th}$ power residue. There exist n of these residues and they are all powers of g. Thus, all $\frac{p-1}{n}^{th}$ power residues form the multiplicative subgroup of $(\mathbb{Z}/p\mathbb{Z})^*$ generated by g.

Lemma 2. *Let H denote the multiplicative subgroup generated by g, and let $x \in (\mathbb{Z}/p\mathbb{Z})^*$. Then $y \in xH$ if and only if $yg^y \in xH$.*

Proof. ($\Rightarrow$) If $y \in xH$, then $y \equiv xg^k \pmod{p}$ for some $1 \leq k \leq n$. Thus $yg^y \equiv xg^k g^y \equiv xg^{k+y} \pmod{p} \in xH$, since $g^{k+y} \in H$.

($\Leftarrow$) Suppose $yg^y \in xH$. Then $yg^y \equiv xg^l \pmod{p}$ for some $1 \leq l \leq n$. This implies that $y \equiv xg^l g^{-y} \equiv xg^{l-y} \pmod{p}$, where $g^{l-y} \in H$. □

Since all $\frac{p-1}{n}^{th}$ power residues map to each other, and all elements of the same coset of the subgroup of these residues map to each other, each connected component of functional graph of $x \mapsto xg^x \pmod{p}$ must consist entirely of elements of the subgroup H generated by g (precisely the $\frac{p-1}{n}^{th}$ power residues) or elements of a coset xH for some $x \notin H$. Furthermore, these cosets partition the entire set of nodes $\{1, ..., p-1\}$. This is illustrated in Figure 3, where the disjoint cosets are represented by different colorings.

Proposition 10. *Given any g, n is an upper bound on the number of nodes in any given connected component in the functional graph of $x \mapsto xg^x \pmod{p}$.*

Proof. Let $g \in (\mathbb{Z}/p\mathbb{Z})^*$ have multiplicative order n. Let H denote the multiplicative subgroup generated by g which contains all $\frac{p-1}{n}^{th}$ power residues modulo p. We have already proved that the maximum number of nodes in a connected component containing an element of H is n.

Consider $x \in (\mathbb{Z}/p\mathbb{Z})^*$ such that x is not a $\frac{p-1}{n}^{th}$ power residue, and thus not in a connected component with elements of H. Then $x = x \cdot 1$ is an element of the left coset xH, and by previous proof, all the elements of the left coset xH must map to each other. Since the size of xH is equal to the size of H, any connected component containing elements of the coset xH for some $x \in (\mathbb{Z}/p\mathbb{Z})^*$ can have at most n nodes. Thus n is the maximum number of nodes in any connected component of the functional graph. □

Corollary 3. *If the functional graph of the given form has only one connected component, g must be a primitive root.*

Proof. A functional graph which consists of exactly one connected component must have all $p-1$ elements in the same connected component. Given that the order of g is an upper bound on the size of the component, and g can have order at most $p-1$, the order of g must be $p-1$, which implies that g is a primitive root modulo p. □

Corollary 4. *Given any g and any prime p, $\frac{p-1}{n}$ is a lower bound for the number of connected components in the functional graph $x \mapsto xg^x \pmod{p}$.*

Proof. We know from Proposition 11 that n is an upper bound on the number of nodes in any connected component of the functional graph $x \mapsto xg^x \pmod{p}$. Consider the case where each connected component of the functional graph contains exactly n nodes, the maximum amount. In this scenario, the number of connected components is precisely $\frac{p-1}{n}$, which is the minimum amount possible since if any connected component contained fewer than n nodes, the remaining nodes would have to be contained in one or more additional connected components. □

3.7 Cycles

Although cycles in the functional graphs of the Discrete Lambert Map are seemingly random in occurence and size, there is some pattern evident in their nodes when they do appear.

Lemma 3. *Let $f^{(n)}(x)$ denote the function $f(x) \equiv xg^x \pmod{p}$ applied n times (e.g. $f^{(2)}(x) = f(f(x))$). Then $f^{(n)}(x) \equiv xg^{x+f(x)+f^{(2)}(x)+\ldots+f^{(n-1)}(x)} \pmod{p}$.*

Proof. We prove this by induction. Let $f^{(0)}(x) = x$, and $f^{(1)}(x) = f(x) = xg^x$.

Suppose $f^{(k)}(x) \equiv xg^{x+f(x)+f^{(2)}(x)+\ldots+f^{(k-1)}(x)} \pmod{p}$. Then

$$\begin{aligned} f^{(k+1)}(x) &= f(f^{(k)}(x)) \\ &= f^{(k)}(x)g^{f^{(k)}(x)} \\ &\equiv xg^{x+f(x)+f^{(2)}(x)+\ldots+f^{(k-1)}(x)}g^{f^{(k)}(x)} \\ &\equiv xg^{x+f(x)+f^{(2)}(x)+\ldots+f^{(k-1)}(x)+f^{(k)}(x)} \pmod{p}. \end{aligned}$$

□

Proposition 11. *If the functional graph of $x \mapsto xg^x$ contains an n-cycle, then the sum of the nodes in the n-cycle is divisible by the order of g.*

Proof. Suppose $x, f(x), \ldots f^{(n-1)}(x)$ are the n nodes of an n-cycle. Then

$$x = f^{(n)}(x) \equiv xg^{x+f(x)+f^{(2)}(x)+\ldots+f^{(n-1)}(x)} \pmod{p}.$$

This implies that

$$1 \equiv g^{x+f(x)+f^{(2)}(x)+\ldots+f^{(n-1)}(x)} \pmod{p}.$$

Thus the order of g must divide $x + f(x) + f^{(2)}(x) + \ldots + f^{(n-1)}(x)$, the sum of the nodes in the n-cycle. □

Corollary 5. *If g is a primitive root, and the functional graph of $x \mapsto xg^x$ contains a 2-cycle, then the sum of the nodes in the 2-cycle is $p-1$.*

Proof. Suppose $x, f(x)$ compose a 2-cycle. By our previous theorem, $x + f(x) \mid p-1$, the order of g. However, since $p-1$ is always a fixed point, $x, f(x) \neq p-1$ and so $x, f(x) < p-1$. This implies that $x+f(x)$ cannot be a multiple of $p-1$, therefore $x + f(x) = p-1$. □

Proposition 12. *If the order of g divides $g+1$, then $g \mapsto 1$, and g and 1 form a 2-cycle.*

Proof. We know that 1 always maps to g. Let n denote the order of g. Suppose n divides $g+1$, and write $g+1 = nk$ for some $k \in \mathbb{Z}$. Then $f(g) = gg^g \equiv g^{g+1} \equiv g^{nk} \equiv (g^n)^k \equiv 1 \pmod{p}$. Thus g also maps to 1, and they form a 2-cycle. □

4 Statistical Analysis

After analyzing the structure of the Discrete Lambert Map, we used statistical methods to compare the Discrete Lambert Map-induced graphs to random functional graphs. As stated in the introduction, we began this process by first selecting the twenty safe primes we would use and determining which graph characteristics were most important to examine based on previous work and results from literature [5, 3]. The next step was to generate data for all the graphs for $g = 2$ through $g = p-2$ for each of the forty primes; however, within each prime, we wanted to average the data collected for graphs that were produced by values of g with similar orders. Since each prime is a safe prime, and since we are excluding $g = 1$ and $g = p - 1$, we are left with only quadratic residues and primitive roots, which have order $\frac{p-1}{2}$ and $p - 1$, respectively. Although Cloutier, Hoffman, and Lindle's code provided a good starting point for our own data collection program, we had to modify it quite a bit so that the right values of g were used and so that the overall totals and observed averages could be broken down based on the order of g. After our program was up and running, we gathered the observed averages for the graph characteristics for each order of each prime.

Our code also calculated the expected means for each of the graph characteristics we selected for each order or each prime. The paper by Flajolet and Odlyzko contains asymptotic approximations for all of our characteristics of interest. However, since we already know much of the structure of the DLM-induced graphs, we did not simply plug $p - 1$ into these approximations since that would give us the expected values for any functional graph on $p - 1$ nodes. Instead, since we know that n, the order of g, is an upper bound on the number of nodes in any connected component of the graph, we plugged n into the approximations and then multiplied the result by $\frac{p-1}{n}$, the minimum number of connected components. Essentially, we were taking advantage of the fact that a DLM-induced functional graph on $p - 1$ nodes acts more like $\frac{p-1}{n}$ functional graphs on n nodes. This slight modification helped us to better predict the behavior of the DLM-induced graphs since the observed means would be compared to expected means that did not take into account configurations that simply could not exist in DLM-induced graphs. It is also important to note that although the Flajolet and Odlyzko paper did have asymptotic approximations for maximum cycle length and maximum tail length, we were not able to find a way to predict the observed average maxima, which would require a sampling distribution of the maximum. Thus, we excluded the maxima from our statistical analysis. Also, since we were only considering values of g with order $p - 1$ and order $\frac{p-1}{2}$, we know that the fixed points of the quadratic residues will always be $\frac{p-1}{2}$ and $p - 1$, while the primitive roots will have $p - 1$ as their only fixed point. With that said, our expectation for the number of fixed points is based on the number of cosets in the graph; therefore, our expected and observed values were always identical. As a result, we did not perfrom statistical tests on the average number of fixed points.

Once our expected values were calculated, we used Excel to find the standard deviation and variance for each characteristic for each order of each prime. Although previous iterations of the code calculated standard deviation and variance internally, the additional complication of breaking the data down by order made that portion of the code obsolete, so we opted to use Excel's built-in functiona to calculate those statistics.

After collecting the observed data, calculating the expected data, and finding the standard deviations and variances, we used the statistical software Minitab to compare our observed and expected means using t-tests. These tests produced t and p values that provide a measure of how statistically similar two data sets are based on the mean and standard deviation. For this paper, we will consider a p-value of 0.05 or less to be statistically significant. A p-value of 0.05 or less means that there is less than a 5% chance that the differences in the data set were due to chance. Although the individual t and p values are important, we were more concerned with the distribution, mean, and standard deviation of the sets of t-values for each graph characteristic for the primitive roots and the quadratic residues separately. By looking at these two categories separately, we can easily tell whether those orders make a difference in the statistics. The t-values themselves measure how many standard deviations away from the mean specific data points fall, and it is a well-known result that ideally, the mean of a set of t-values is 0 with a standard deviation that approaches 1 as the number of samples approches infinity. Since we are considering such large collections of graphs, the standard deviation should be very close to 1. To see data tables with all of the averaged information we collected, please see Appendix A. We used probability plots in Minitab to measure how close each set of t-values was to this ideal mean and standard deviation. Initially, our expected and observed values did not line up very well. In Figure 4, we see that the t-values for comparing the expected and observed total cycle lengths for primitive roots do not have the expected mean and standard deviation. The low p-value of less than 0.001 indicates that this discrepancy is likely not due to chance and that there is a statistical difference between the observed and expected behavior of the t-values. The same is true for the t-values of the quadratic residues, which can be seen in Figure 5.

We attributed this huge discrepancy between our observed and expected results to inccurate predictions. Since Flajolet and Odlyzko's paper listed only one term for the asymptotic forms for the average values of the graph characteristics, we decided to try computing a second term to improve our predictions [3]. For most of the characteristics, the paper listed the generating function that they used to create the asymptotic forms. In these instances, we used a special package in Maple that converted the normalized generating functions to their asymptotic forms and then added the second term to our approximation.

In the case of the average tail length, the paper did not list its generating function. Therefore, we used a result from another paper about the generating function used to count the average tail length for binary functional graphs [1]. Using the function β from that paper, we created the function τ, which marks the edges along one tree, where

$$\tau(z, u) = ze^{t(z)} + uz\tau(z, u)e^{t(z)},$$

where $t(z)$ is the number of trees. We then used Maple to solve this function for τ and plugged it into a larger generating function that would count the average

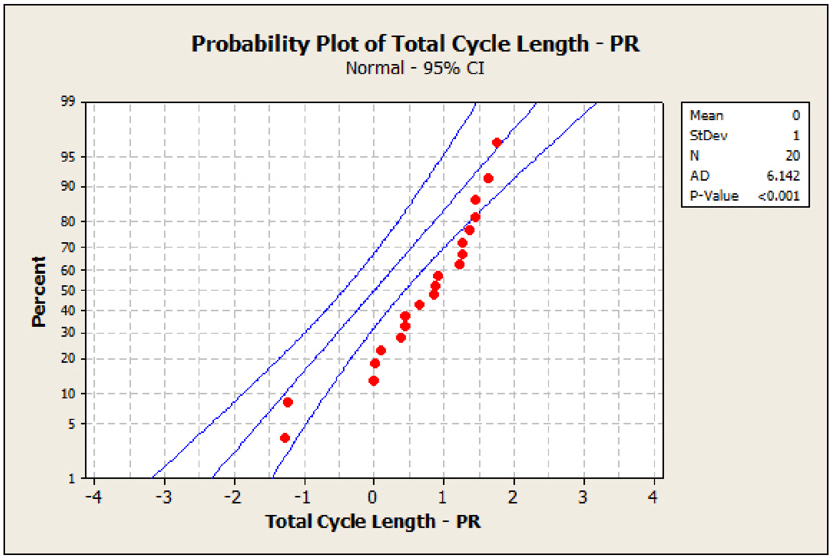


Figure 4: Plot of the t-values associated with comparing the observed and expected means of total cycle length for all primtive roots.

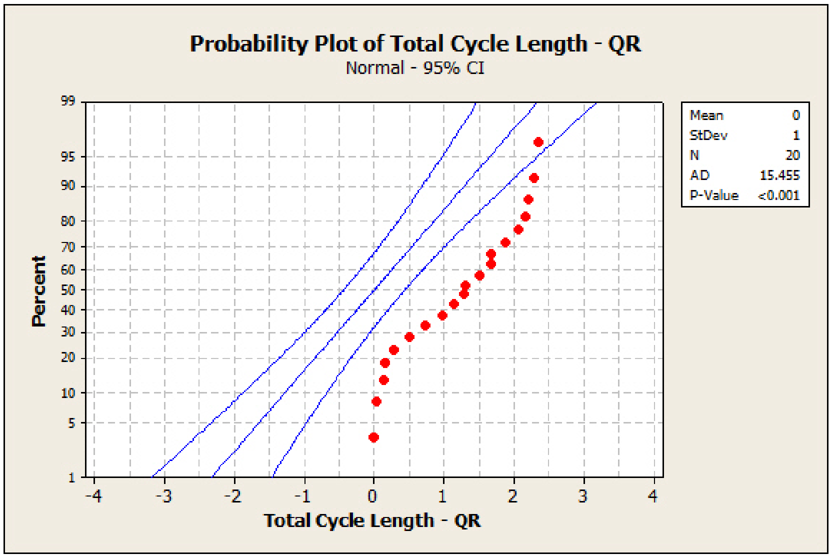


Figure 5: Plot of the t-values associated with comparing the observed and expected means of total cycle length for all quadratic residues.

tail length. This gave us

$$\xi(u,z) = e^{\log\frac{1}{1-t(z)}} \cdot \frac{1}{1-t(z)} \cdot \tau(u,z),$$

where the first term marks all the possible components except one, the second term marks all the possible trees in a component except one, and the last term marks the tree of interest. Differentiating this function with respect to u and then evaluating it at $u = 1$ yields the correct generating function

$$\Xi(z) = \frac{\mathrm{LambertW}(-z)^2}{(1+\mathrm{LambertW}(-z))^4},$$

where LambertW is the Lambert W function. After we found this generating function, we followed the same method that we used with the other generating functions, and used Maple to compute a second term for the asymptotic approximation of the average tail length. To see all of the two-term approximations, please see Appendix C.

After adding the second term to our expansion, virtually all of t-values improved and our expected and observed values were much closer. In Figure 6, we are now plotting the new t-values that were calculated with the improved expected means of the cycle length of primitive roots. As you can see from the sufficiently high p-value of 0.143, the mean and standard deviation of the t-values are not sufficiently different from the expected mean of 0 and standard deviation of 1. The same is true for quadratic residues, as is seen in Figure 7.

Unfortunately, this addition of the second term did not improve the distribution for the t-values of all the characteristics. For example, the t-values for the terminal and expected nodes are still off by a good amount. This might be due to the fact that there is a guaranteed fixed point, $p-1$, in each graph, which is obviously going to always be a terminal node, and the generating functions for the expected value of terminal and expected nodes do not take that into account. Excluding the image and terminal nodes t-distribution, every other t-distribution for both quadratic residues and primitive roots for each graph characteristic fell within the fail to reject region, with the one exception of the distribution of the t-values for comparing average total tail length for quadratic residues. It is not clear why this distribution is not statistically close to its expected mean and standard deviation, but it is possible that this is an example of non-randomness that is occuring in the DLM-induced graphs. We attempted to make our expected values more accurate by computing a second term for the normalizing factor of the generating functions, but that hardly changed the expected values. To see all the probability plots for each characteristic and for both quadratic residues and primitive roots, please see Appendix B.

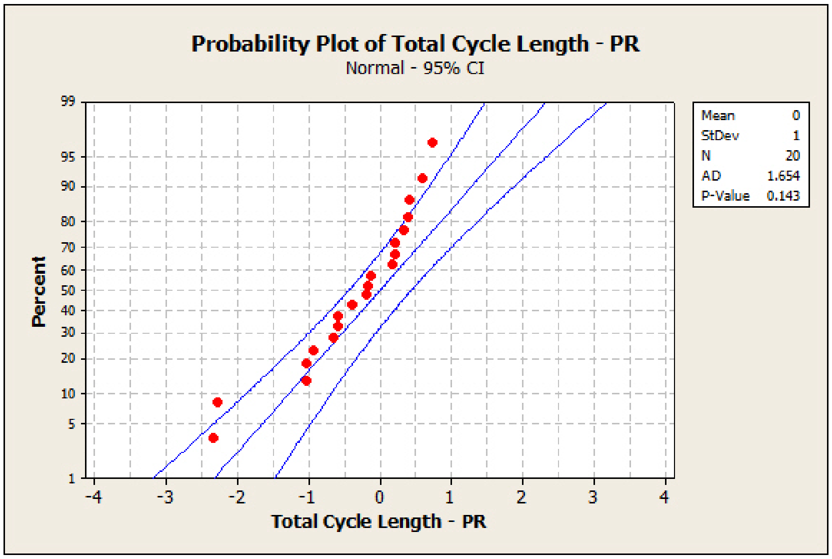


Figure 6: Plot of the t-values associated with comparing the observed and improved expected means of total cycle length for all primtive roots.

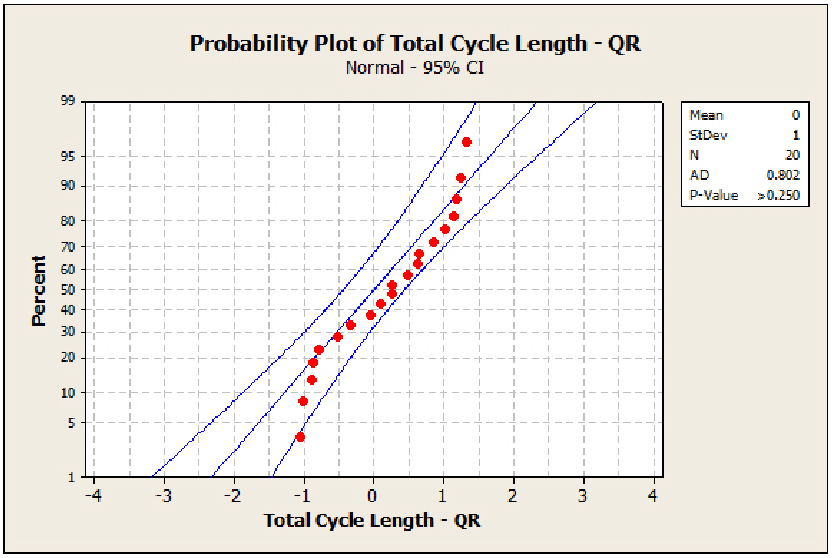


Figure 7: Plot of the t-values associated with comparing the observed and improved expected means of total cycle length for all quadratic residues.

5 Conclusion

The work presented in this paper is result of only initial investigations of the structure of the Discrete Lambert Map. Through studying the functional graphs of this map, we were able to determine fairly precisely the behavior of certain graph-theoretic characteristics based on information about the chosen value of g, which, for cryptographic purposes, will help us evaluate the presumed difficulty of inverting this function. We observed that for $g = 1$ and $g = p - 1$, the graphs are entirely predictable, which confirm that these are not good choices for a secure cryptosystem. For other values of g, we know the images of certain nodes, such as $\frac{p-1}{2}$ and $p - 2$. We also noted properties of the nodes in a cycle, as well as specific instances of 2-cycles.

The order of g is also crucial to our understanding of the behavior of the function, as it gives us the fixed points of the map as well the maximum size and minimum number of connected components in the functional graph. Knowing that a connected component consists entirely of elements of the same coset of $g \in (\mathbb{Z}p\mathbb{Z})^*$ allow us to view these graphs as compositions of smaller subgraphs, each of which corresponds to a coset. This proved relevant for our methods of statistical analysis, since we could then account for the known minimum number of connected components in comparing our DLM functional graphs to random functional graphs.

From the statistical side, t-tests performed on observed and expected average values for graph characteristics such as number of connected components, cyclic nodes, total cycle length and total tail length showed that differences between DLM functional graphs and random functional graphs in these characteristics were not statistically significant. This suggests that in these aspects, DLM-induced graphs appear similar to random functional graphs. For some parameters of interest, such as number of image nodes, terminal nodes, and tail length, our DLM graphs produced data which did not seem to fit expected values well. This may have been a result of non-random behavior of the DLM, but can more likely be attributed to inaccurate expectations of random graph data. Since our expected values came from literature about general random functional graphs, they did not account for the known fixed points that occur in DLM functional graphs. This may have produced the discrepency we saw in the statistical analysis.

Other aspects of the functional graphs are not easily explained, such as the in-degree of nodes and the occurrence and length of cycles. For example, with the exception of $g = 1$ and $g = p - 1$, the graphs do not appear to be m-ary in any way, but we have been unable to prove a general result. While we have a lower bound on the number of connected components, it is not evident when this minimum occurs, or how to construct an upper bound for the number of components. We have also observed other structural patterns for which we do not have a definitive explanation, and future work could involve further investigation and formalization of these phenomena:

1) Graphs generated by values of g with equal and small multiplicative orders mod p have structurally similar connected components.

2) For specific values of g, the functional graphs created by the Discrete Lambert Map contain cycles composed solely of primitive roots.

A considerable amount of future work lies in statistical analysis of the DLM. For example, the expected means for the number of image and terminal nodes need to be refined by taking into account the known fixed points. We observed an abnormal mean and standard deviation for the t-values associated with comparing those predictions to the observed data, which led us to believe that the asymptotic approximations in Flajolet and Odlyzko's paper were not sufficient in predicting the values for those graph characteristics. In addition, a formula for computing expected means for characteristics such as maximum tail length and maximum cycle length currently does not exist. Approximations for estimated variances are also lacking for each of our graph characteristics. There are some methods in literature that might prove useful in deriving these approximations, but there have yet to exist explicit formulations. These expected variances can then be statistically compared to the observed variations using ANOVA tests. The statistical analysis could also be expanded to include primes other than safe primes in order to see if the factorization of $p-1$ influences the characteristics of the induced graphs. Futher improvements can also be made on the code used to generate the data. Currently, the standard deviations need to be computed manually outside of the code, but with a few slight modifications, the code could itself produce that data.

6 Acknowledgments

We would like to thank Joshua Holden for his indispensable role in advising and guiding us on this project. We also recognize previous REU participants Friedrichsen, Larson, and McDowell, whose paper served as inspiration for our own work, and Cloutier, Lindle, and Hoffman, whose code helped us to collect the relevant statistical data for our investigations. Finally, we thank the NSF for the grant that made this REU program possible.

References

[1] Cloutier, D., Holden, J., *Mapping the discrete logarithm*, Involve **3**: 197-213, 2010.

[2] Corless, R. M.; Gonnet, G. H.; Hare, D. E. G.; Jeffrey, D. J.; Knuth, D. E. (1996). *On the Lambert W function*, Advances in Computational Mathematics **5**: 329-359.

[3] Flajolet, P., Odlyzko, A., *Random Mapping Statistics*, Advances in Cryptology–EUROCRYPT '89 (Houthalen, Belgium, 1989), Lecture Notes in Comput. Sci., vol. 434, Berlin: Springer, 1990, 329-354.

[4] Friedrichsen, M., Larson, B., McDowell, E., *Structure and Statistics of the Self-Power Map*, Rose-Hulman Undergraduate Mathematics Journal **11** (2010), no. 1.

[5] Hoffman, A., *Statistical investigation of structure in the discrete logarithm*, Rose-Hulman Undergraduate Mathematics Journal **10** (2009), no. 2.

[6] Lindle, N., *A statistical look at maps of the discrete logarithm*, Senior thesis, Rose-Hulman Institute of Technology, http://www.csse.rose-hulman.edu/images/docs/theses/NathanLindle2008.pdf, 2008.

[7] Montgomery, H.L., Niven, I., Zuckerman, H. S., *An Introduction to the Theory of Numbers*, New York: Wiley, 1991.

[8] Rosen, K.H., *Elementary Number Theory and Its Applications*, 5^{th} Ed., Boston: Pearson/Addison-Wesley, 2005.

A Data Tables

This section includes all of the observed and expected data for the graph characteristics, as well as the t and p values associated with testing the observed and expected means against one another. The expected values in these tables were computed using the two-term asymptotic expansion, with the exception of the terminal and image nodes expected values, which were left the same since the second term caused them to become much worse.

Number of Connected Components								
Prime	**Order**	**Graphs**	**Obs Avg**	**Exp Avg**	**St Dev**	**Var**	***t*-stat**	***p*-val**
40127	20063	20062	11.23273	11.177	3.495364	12.21757	2.26	0.024
40499	20249	20248	11.19553	11.18622	2.609219	6.808026	0.51	0.612
40739	20369	20368	11.17596	11.19213	2.590197	6.70912	-0.89	0.373
40787	20393	20392	11.17144	11.19331	2.61968	6.862723	-1.19	0.233
40823	20411	20410	11.19015	11.19419	2.58423	6.678247	-0.22	0.823
40883	20441	20440	11.20083	11.19566	2.601074	6.765586	0.28	0.776
41387	20693	20692	11.23777	11.20791	2.627239	6.902385	1.63	0.102
41507	20753	20752	11.23453	11.21081	2.61189	6.821971	1.31	0.191
41519	20759	20758	11.20561	11.2111	2.602855	6.774856	-0.3	0.761
41543	20771	20770	11.21767	11.21168	2.600053	6.760275	0.33	0.74
41579	20789	20788	11.20242	11.21254	2.610721	6.815865	-0.56	0.576
41759	20879	20878	11.20447	11.21686	2.58445	6.67938	-0.69	0.489
41843	20921	20920	11.1869	11.21887	2.609873	6.811435	-1.77	0.076
41879	20939	20938	11.25065	11.21973	2.611636	6.820643	1.71	0.087
41927	20963	20962	11.22488	11.22088	2.607007	6.796484	0.22	0.824
42023	21011	21010	11.19448	11.22316	2.610886	6.816723	-1.59	0.111
42179	21089	21088	11.2317	11.22687	2.621732	6.873479	0.27	0.789
42299	21149	21148	11.25506	11.22971	2.606954	6.796208	1.41	0.157
42359	21179	21178	11.28119	11.23113	2.61506	6.838536	2.79	0.005
42443	21221	21220	11.26956	11.23311	2.607718	6.800195	2.04	0.042
40127	40126	20062	5.949507	5.935071	1.916821	3.674202	1.07	0.286
40499	40498	20248	5.94918	5.939685	1.92016	3.687016	0.7	0.482
40739	40738	20368	5.981	5.942639	1.931664	3.731326	2.83	0.005
40787	40786	20392	5.931346	5.943228	1.924653	3.70429	-0.88	0.378
40823	40822	20410	5.941401	5.943669	1.943921	3.77883	-0.17	0.868
40883	40882	20440	5.947701	5.944404	1.91652	3.673048	0.25	0.806
41387	41386	20692	5.961483	5.95053	1.930109	3.725319	0.82	0.414
41507	41506	20752	5.942704	5.951978	1.914681	3.666002	-0.7	0.485
41519	41518	20758	5.956306	5.952122	1.930348	3.726244	0.31	0.755
41543	41542	20770	5.968127	5.952411	1.927563	3.715498	1.18	0.24
41579	41578	20788	5.94593	5.952844	1.930673	3.727498	-0.52	0.606
41759	41758	20878	5.948606	5.955004	1.938585	3.758112	-0.48	0.633
41843	41842	20920	5.944073	5.956009	1.932891	3.736069	-0.89	0.372
41879	41878	20938	5.966281	5.956439	1.939137	3.760254	0.73	0.463
41927	41926	20962	5.962933	5.957012	1.920272	3.687444	0.45	0.655
42023	42022	21010	5.964112	5.958155	1.940634	3.766061	0.44	0.656
42179	42178	21088	5.977049	5.960008	1.946931	3.790539	1.27	0.204
42299	42298	21148	5.973946	5.961429	1.933751	3.739391	0.94	0.347
42359	42358	21178	5.963925	5.962137	1.936406	3.749667	0.13	0.893
42443	42442	21220	5.959001	5.963128	1.937432	3.753644	-0.31	0.756

Number of Cyclic Nodes

Prime	Order	Graphs	Obs Avg	Exp Avg	St Dev	Var	*t*-stat	*p*-val
40127	20063	20062	355.5924	354.382	131.6946	17343.47	1.3	0.193
40127	40126	20062	249.6274	250.724	128.8308	16597.37	-1.21	0.228
40499	20249	20248	357.2966	356.024	130.533	17038.86	1.39	0.165
40499	40498	20248	252.6404	251.885	130.3355	16987.34	0.82	0.41
40739	20369	20368	356.8789	357.0793	131.0907	17184.77	-0.22	0.827
40739	40738	20368	253.3772	252.6313	130.6463	17068.46	0.81	0.415
40787	20393	20392	356.0126	357.29	131.2882	17236.59	-1.39	0.165
40787	40786	20392	253.2665	252.7803	132.3613	17519.51	0.52	0.6
40823	20411	20410	357.2524	357.448	131.1337	17196.06	-0.21	0.831
40823	40822	20410	254.2715	252.892	132.5435	17567.77	1.49	0.137
40883	20441	20440	357.9445	357.7111	132.7968	17634.99	0.25	0.802
40883	40882	20440	253.5366	253.078	131.3474	17252.13	0.5	0.618
41387	20693	20692	359.9552	359.9134	132.137	17460.18	0.05	0.964
41387	41386	20692	255.3635	254.6352	132.8781	17656.58	0.79	0.43
41507	20753	20752	360.7228	360.4357	134.3856	18059.48	0.31	0.758
41507	41506	20752	253.4441	255.0046	132.6159	17586.97	-1.7	0.09
41519	20759	20758	360.4336	360.4879	133.4573	17810.86	-0.06	0.953
41519	41518	20758	256.0999	255.0415	133.7712	17894.74	1.14	0.254
41543	20771	20770	360.5166	360.5923	132.4567	17544.79	-0.08	0.934
41543	41542	20770	255.5311	255.1153	132.6163	17587.07	0.45	0.651
41579	20789	20788	360.9208	360.7488	134.2905	18033.93	0.18	0.853
41579	41578	20788	253.586	255.226	132.6531	17596.84	-1.78	0.075
41759	20879	20878	360.4141	361.5303	133.3345	17778.08	-1.21	0.226
41759	41758	20878	256.1443	255.7786	133.6274	17856.29	0.4	0.693
41843	20921	20920	362.7241	361.8944	134.1375	17992.86	0.89	0.371
41843	41842	20920	256.1609	256.036	134.0547	17970.65	0.13	0.893
41879	20939	20938	361.7813	362.0503	133.8673	17920.45	-0.29	0.771
41879	41878	20938	256.0287	256.1463	133.6683	17867.2	-0.13	0.899
41927	20963	20962	363.6237	362.2581	133.6965	17874.76	1.48	0.139
41927	41926	20962	256.1655	256.2933	132.9362	17672.02	-0.14	0.889
42023	21011	21010	362.7877	362.6734	133.0626	17705.66	0.12	0.901
42023	42022	21010	256.4947	256.5869	133.6966	17874.78	-0.1	0.92
42179	21089	21088	363.2387	363.3472	133.793	17900.58	-0.12	0.906
42179	42178	21088	258.2918	257.0633	135.2315	18287.57	1.32	0.186
42299	21149	21148	363.2545	363.8647	132.5776	17576.82	-0.67	0.503
42299	42298	21148	256.957	257.4292	134.8165	18175.49	-0.51	0.61
42359	21179	21178	365.7958	364.1231	135.3907	18330.65	1.8	0.072
42359	42358	21178	257.8609	257.612	135.3907	18330.63	0.27	0.789
42443	21221	21220	365.6761	364.4846	135.3082	18308.31	1.28	0.2
42443	42442	21220	257.6276	257.8676	134.6595	18133.19	-0.26	0.795

Number of Terminal Nodes								
Prime	**Order**	**Graphs**	**Obs Avg**	**Exp Avg**	**St Dev**	**Var**	***t* -stat**	***p* -val**
40127	20063	20062	14762.43	14761.53	45.59554	2078.953	2.78	0.005
40127	40126	20062	14762.04	14761.53	45.66224	2085.04	1.57	0.116
40499	20249	20248	14898.28	14898.38	46.11516	2126.608	-0.31	0.757
40499	40498	20248	14899.66	14898.38	45.91716	2108.386	3.96	0
40739	20369	20368	14987.23	14986.67	45.98598	2114.71	1.74	0.082
40739	40738	20368	14989.44	14986.67	46.49471	2161.758	8.5	0
40787	20393	20392	15004.29	15004.33	46.36748	2149.943	-0.13	0.896
40787	40786	20392	15004.4	15004.33	46.19099	2133.608	0.22	0.829
40823	20411	20410	15017.02	15017.57	45.14228	2037.825	-1.77	0.077
40823	40822	20410	15018.53	15017.57	46.12181	2127.221	2.97	0.003
40883	20441	20440	15040.09	15039.65	45.88905	2105.805	1.38	0.168
40883	40882	20440	15040.5	15039.65	46.34672	2148.019	2.62	0.009
41387	20693	20692	15224.53	15225.06	45.7847	2096.239	-1.65	0.099
41387	41386	20692	15225.67	15225.06	46.39823	2152.796	1.88	0.06
41507	20753	20752	15270.01	15269.2	46.95264	2204.55	2.48	0.013
41507	41506	20752	15270.36	15269.2	46.84779	2194.715	3.57	0
41519	20759	20758	15274.89	15273.62	46.46534	2159.028	3.93	0
41519	41518	20758	15275.13	15273.62	46.67358	2178.423	4.65	0
41543	20771	20770	15281.99	15282.45	46.41985	2154.803	-1.42	0.155
41543	41542	20770	15283.73	15282.45	47.64238	2269.796	3.88	0
41579	20789	20788	15295.01	15295.69	46.83155	2193.194	-2.08	0.037
41579	41578	20788	15296.5	15295.69	46.69699	2180.609	2.48	0.013
41759	20879	20878	15361.92	15361.91	46.47234	2159.678	0.03	0.974
41759	41758	20878	15363.08	15361.91	46.89271	2198.926	3.62	0
41843	20921	20920	15391.49	15392.81	46.99637	2208.659	-4.07	0
41843	41842	20920	15392.5	15392.81	46.51539	2163.681	-0.96	0.339
41879	20939	20938	15406.52	15406.06	47.0134	2210.26	1.44	0.149
41879	41878	20938	15407.96	15406.06	47.20175	2228.005	5.84	0
41927	20963	20962	15423.63	15423.71	46.41091	2153.972	-0.27	0.783
41927	41926	20962	15424.04	15423.71	46.87373	2197.147	0.99	0.321
42023	21011	21010	15459.66	15459.03	47.14316	2222.477	1.93	0.054
42023	42022	21010	15459.25	15459.03	46.57621	2169.343	0.69	0.492
42179	21089	21088	15516.36	15516.42	47.16138	2224.196	-0.19	0.852
42179	42178	21088	15517.17	15516.42	46.29377	2143.113	2.36	0.018
42299	21149	21148	15561.63	15560.56	47.00392	2209.368	3.29	0.001
42299	42298	21148	15561.78	15560.56	47.09649	2218.08	3.74	0
42359	21179	21178	15583.21	15582.64	46.84488	2194.442	1.78	0.075
42359	42358	21178	15582.89	15582.64	47.37204	2244.11	0.77	0.443
42443	21221	21220	15613.61	15613.54	46.99878	2208.885	0.23	0.821
42443	42442	21220	15612.54	15613.54	46.71666	2182.446	-3.13	0.002

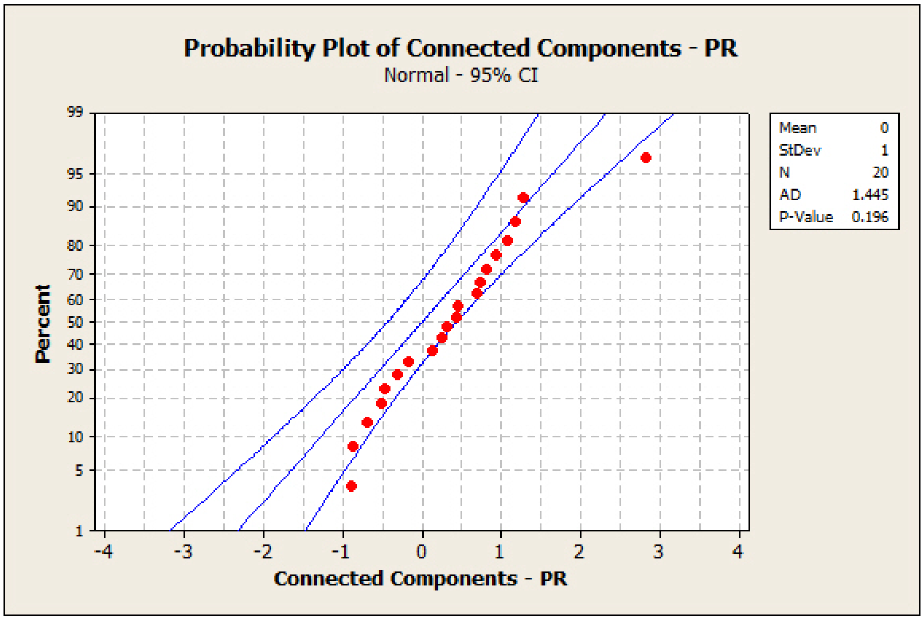
Number of Image Nodes								
Prime	**Order**	**Graphs**	**Obs Avg**	**Exp Avg**	**St Dev**	**Var**	***t* -stat**	***p* -val**
40127	20063	20062	25363.57	25364.47	45.59554	2078.953	-2.78	0.005
40127	40126	20062	25363.96	25364.47	45.66224	2085.04	-1.57	0.116
40499	20249	20248	25599.72	25599.62	46.11516	2126.608	0.31	0.757
40499	40498	20248	25598.34	25599.62	45.91716	2108.386	-3.96	0
40739	20369	20368	25750.77	25751.33	45.98598	2114.71	-1.74	0.082
40739	40738	20368	25748.56	25751.33	46.49471	2161.758	-8.5	0
40787	20393	20392	25781.71	25781.67	46.36748	2149.943	0.13	0.896
40787	40786	20392	25781.6	25781.67	46.19099	2133.608	-0.22	0.829
40823	20411	20410	25804.98	25804.43	45.14228	2037.825	1.77	0.077
40823	40822	20410	25803.47	25804.43	46.12181	2127.221	-2.97	0.003
40883	20441	20440	25841.91	25842.35	45.88905	2105.805	-1.38	0.168
40883	40882	20440	25841.5	25842.35	46.34672	2148.019	-2.62	0.009
41387	20693	20692	26161.47	26160.94	45.7847	2096.239	1.65	0.099
41387	41386	20692	26160.33	26160.94	46.39823	2152.796	-1.88	0.06
41507	20753	20752	26235.99	26236.8	46.95264	2204.55	-2.48	0.013
41507	41506	20752	26235.64	26236.8	46.84779	2194.715	-3.57	0
41519	20759	20758	26243.11	26244.38	46.46534	2159.028	-3.93	0
41519	41518	20758	26242.87	26244.38	46.67358	2178.423	-4.65	0
41543	20771	20770	26260.01	26259.55	46.41985	2154.803	1.42	0.155
41543	41542	20770	26258.27	26259.55	47.64238	2269.796	3.88	0
41579	20789	20788	26282.99	26282.31	46.83155	2193.194	2.08	0.037
41579	41578	20788	26281.5	26282.31	46.69699	2180.609	-2.48	0.013
41759	20879	20878	26396.08	26396.09	46.47234	2159.678	-0.03	0.974
41759	41758	20878	26394.92	26396.09	46.89271	2198.926	-3.62	0
41843	20921	20920	26450.51	26449.19	46.99637	2208.659	4.07	0
41843	41842	20920	26449.5	26449.19	46.51539	2163.681	0.96	0.339
41879	20939	20938	26471.48	26471.94	47.0134	2210.26	-1.44	0.149
41879	41878	20938	26470.04	26471.94	47.20175	2228.005	-5.84	0
41927	20963	20962	26502.37	26502.29	46.41091	2153.972	0.27	0.783
41927	41926	20962	26501.96	26502.29	46.87373	2197.147	-0.99	0.321
42023	21011	21010	26562.34	26562.97	47.14316	2222.477	-1.93	0.054
42023	42022	21010	26562.75	26562.97	46.57621	2169.343	-0.69	0.492
42179	21089	21088	26661.64	26661.58	47.16138	2224.196	0.19	0.852
42179	42178	21088	26660.83	26661.58	46.29377	2143.113	-2.36	0.018
42299	21149	21148	26736.37	26737.44	47.00392	2209.368	-3.29	0.001
42299	42298	21148	26736.22	26737.44	47.09649	2218.08	-3.74	0
42359	21179	21178	26774.79	26775.36	46.84488	2194.442	-1.78	0.075
42359	42358	21178	26775.11	26775.36	47.37204	2244.11	-0.77	0.443
42443	21221	21220	26828.39	26828.46	46.99878	2208.885	-0.23	0.821
42443	42442	21220	26829.46	26828.46	46.71666	2182.446	3.13	0.002

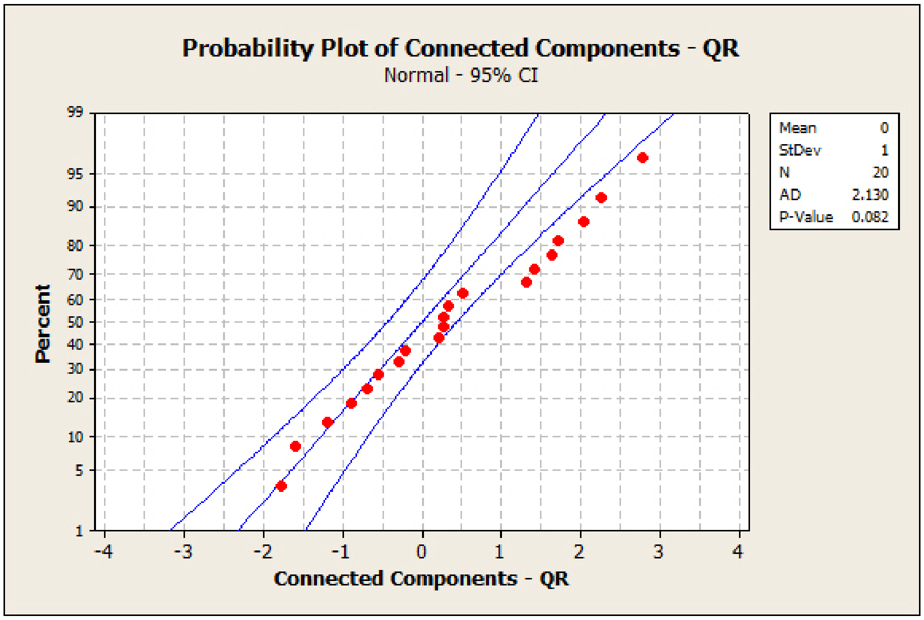
Total Cycle Length								
Prime	**Order**	**Graphs**	**Obs Avg**	**Exp Avg**	**St Dev**	**Var**	***t*-stat**	***p*-val**
40127	20063	20062	3588366	3575046	1843359	3.4E+12	1.02	0.306
40127	40126	20062	5039640	5063713	3636356	1.32E+13	-0.94	0.348
40499	20249	20248	3628179	3624814	1842761	3.4E+12	0.26	0.795
40499	40498	20248	5129644	5134169	3683992	1.36E+13	-0.17	0.861
40739	20369	20368	3646927	3657043	1854242	3.44E+12	-0.78	0.436
40739	40738	20368	5176393	5179795	3714295	1.38E+13	-0.13	0.896
40787	20393	20392	3664784	3663501	1863364	3.47E+12	0.1	0.922
40787	40786	20392	5194746	5188936	3747489	1.4E+13	0.22	0.825
40823	20411	20410	3671640	3668346	1880266	3.54E+12	0.25	0.802
40823	40822	20410	5206857	5195796	3774066	1.42E+13	0.42	0.675
40883	20441	20440	3684762	3676427	1875334	3.52E+12	0.64	0.525
40883	40882	20440	5212867	5207235	3744592	1.4E+13	0.22	0.83
41387	20693	20692	3737485	3744537	1902566	3.62E+12	-0.53	0.594
41387	41386	20692	5298354	5303655	3810383	1.45E+13	-0.2	0.839
41507	20753	20752	3748891	3760814	1934448	3.74E+12	-0.89	0.375
41507	41506	20752	5265121	5326699	3792792	1.44E+13	-2.34	0.019
41519	20759	20758	3768897	3762443	1936730	3.75E+12	0.48	0.631
41519	41518	20758	5318433	5329005	3870088	1.5E+13	-0.39	0.694
41543	20771	20770	3774139	3765702	1928669	3.72E+12	0.63	0.528
41543	41542	20770	5353179	5333619	3873061	1.5E+13	0.73	0.467
41579	20789	20788	3782008	3770592	1951088	3.81E+12	0.84	0.399
41579	41578	20788	5280380	5340542	3804112	1.45E+13	-2.28	0.023
41759	20879	20878	3783250	3795074	1931137	3.73E+12	-0.88	0.376
41759	41758	20878	5391318	5375199	3911840	1.53E+13	0.6	0.552
41843	20921	20920	3824283	3806517	1951283	3.81E+12	1.32	0.188
41843	41842	20920	5375391	5391399	3914893	1.53E+13	-0.59	0.554
41879	20939	20938	3797183	3811425	1937939	3.76E+12	-1.06	0.288
41879	41878	20938	5370662	5398346	3900073	1.52E+13	-1.03	0.304
41927	20963	20962	3833690	3817972	1943707	3.78E+12	1.17	0.242
41927	41926	20962	5416767	5407614	3926305	1.54E+13	0.34	0.736
42023	21011	21010	3847730	3831076	1942465	3.77E+12	1.24	0.214
42023	42022	21010	5409982	5426166	3906746	1.53E+13	-0.6	0.548
42179	21089	21088	3851687	3852404	1970213	3.88E+12	-0.05	0.958
42179	42178	21088	5438371	5456357	3934515	1.55E+13	-0.66	0.507
42299	21149	21148	3855273	3868836	1957198	3.83E+12	-1.01	0.314
42299	42298	21148	5451236	5479619	3984335	1.59E+13	-1.04	0.3
42359	21179	21178	3872637	3877061	1971914	3.89E+12	-0.33	0.744
42359	42358	21178	5496089	5491263	3938029	1.55E+13	0.18	0.858
42443	21221	21220	3904246	3888585	2002972	4.01E+12	1.14	0.255
42443	42442	21220	5518656	5507577	3991537	1.59E+13	0.4	0.686

Total Tail Length								
Prime	**Order**	**Graphs**	**Obs Avg**	**Exp Avg**	**St Dev**	**Var**	***t* -stat**	***p* -val**
40127	20063	20062	3538070	3534920	1103402	1.22E+12	0.4	0.686
40127	40126	20062	5008379	5010212	2179624	4.75E+12	-0.12	0.905
40499	20249	20248	3587373	3584316	1106108	1.22E+12	0.39	0.694
40499	40498	20248	5053873	5080171	2177074	4.74E+12	-1.72	0.086
40739	20369	20368	3622108	3616305	1117894	1.25E+12	0.74	0.459
40739	40738	20368	5099666	5125478	2220861	4.93E+12	-1.66	0.097
40787	20393	20392	3630983	3622715	1117001	1.25E+12	1.06	0.29
40787	40786	20392	5118274	5134555	2230210	4.97E+12	-1.04	0.297
40823	20411	20410	3625544	3627524	1113730	1.24E+12	-0.25	0.8
40823	40822	20410	5124993	5141367	2227101	4.96E+12	-1.05	0.294
40883	20441	20440	3631847	3635545	1129483	1.28E+12	-0.47	0.64
40883	40882	20440	5138259	5152726	2213059	4.9E+12	-0.93	0.35
41387	20693	20692	3709924	3703151	1137139	1.29E+12	0.86	0.392
41387	41386	20692	5257904	5248474	2318493	5.38E+12	0.59	0.559
41507	20753	20752	3712990	3719308	1142770	1.31E+12	-0.8	0.426
41507	41506	20752	5287264	5271358	2293700	5.26E+12	1	0.318
41519	20759	20758	3720874	3720925	1150644	1.32E+12	-0.01	0.995
41519	41518	20758	5267223	5273648	2296121	5.27E+12	-0.4	0.687
41543	20771	20770	3722466	3724160	1151156	1.33E+12	-0.21	0.832
41543	41542	20770	5270870	5278230	2291850	5.25E+12	-0.46	0.644
41579	20789	20788	3730686	3729014	1158719	1.34E+12	0.21	0.835
41579	41578	20788	5295373	5285104	2303448	5.31E+12	0.64	0.52
41759	20879	20878	3773622	3753316	1167715	1.36E+12	2.51	0.012
41759	41758	20878	5302561	5319522	2330791	5.43E+12	-1.05	0.293
41843	20921	20920	3757532	3764675	1160001	1.35E+12	-0.89	0.373
41843	41842	20920	5339714	5335609	2317687	5.37E+12	0.26	0.798
41879	20939	20938	3769686	3769547	1157981	1.34E+12	0.02	0.986
41879	41878	20938	5332805	5342509	2306934	5.32E+12	-0.61	0.543
41927	20963	20962	3766448	3776046	1164557	1.36E+12	-1.19	0.233
41927	41926	20962	5361303	5351713	2336719	5.46E+12	0.59	0.552
42023	21011	21010	3783271	3789054	1164996	1.36E+12	-0.72	0.472
42023	42022	21010	5361492	5370136	2317710	5.37E+12	-0.54	0.589
42179	21089	21088	3807929	3810226	1170797	1.37E+12	-0.28	0.776
42179	42178	21088	5375722	5400120	2348305	5.51E+12	-1.51	0.131
42299	21149	21148	3829275	3826538	1170531	1.37E+12	0.34	0.734
42299	42298	21148	5415993	5423222	2339281	5.47E+12	-0.45	0.653
42359	21179	21178	3825265	3834703	1178816	1.39E+12	-1.17	0.244
42359	42358	21178	5429767	5434785	2387286	5.7E+12	-0.31	0.76
42443	21221	21220	3844412	3846143	1185613	1.41E+12	-0.21	0.832
42443	42442	21220	5444164	5450988	2349245	5.52E+12	-0.42	0.672

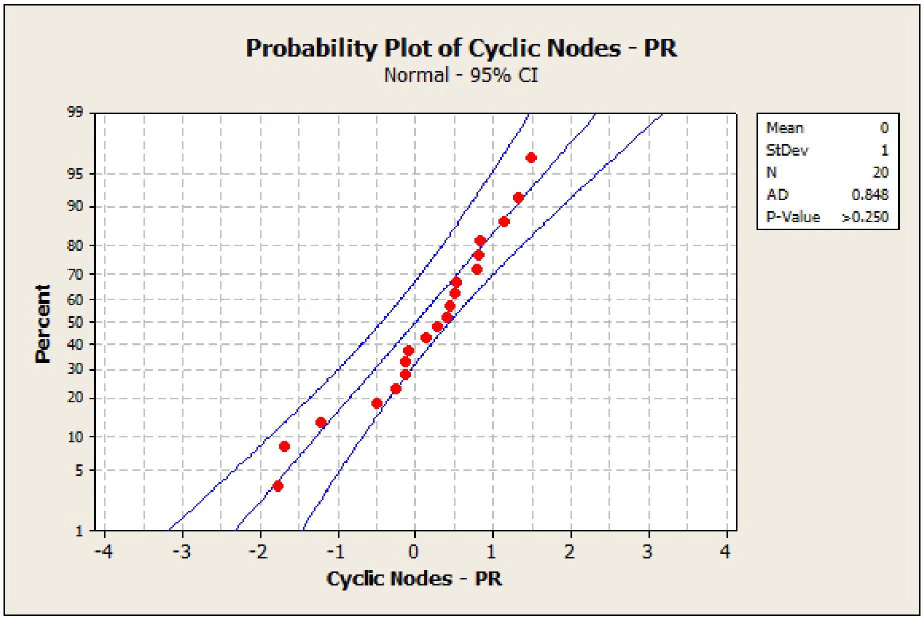
B Probability Plots

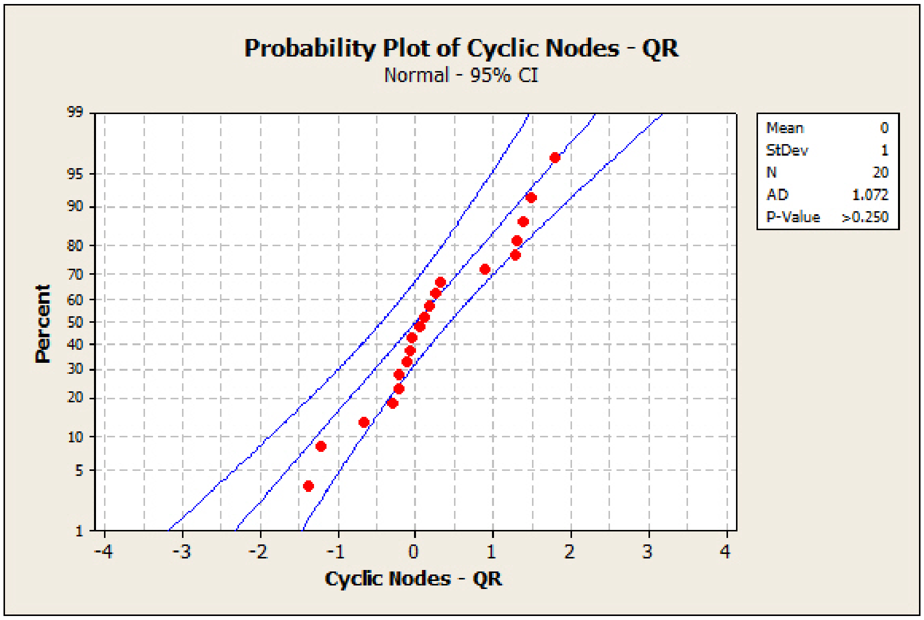
This section includes the probability plots for the t-value distribution for both primitive roots and quadratic residues for each graph characteristic. If the p-value indicates that a mean of 0 and a standard deviation of 1 is not a good fit for the distribution, we instead included the results of a normality test for t-values in question, along with the results of a t-test on the previously computed t-values to see if the mean for those values could indeed be 0. We also included this test on some of the more borderline cases.

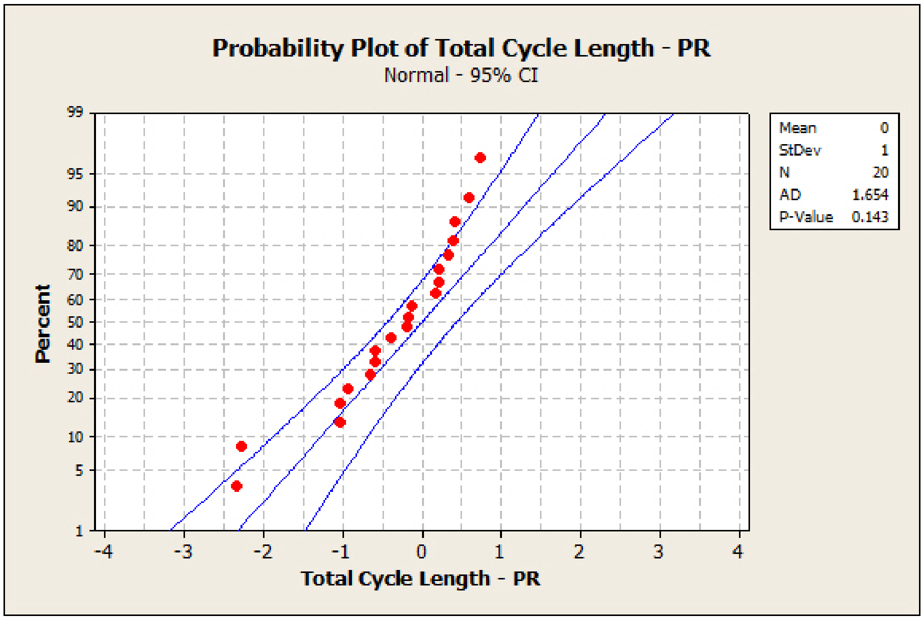


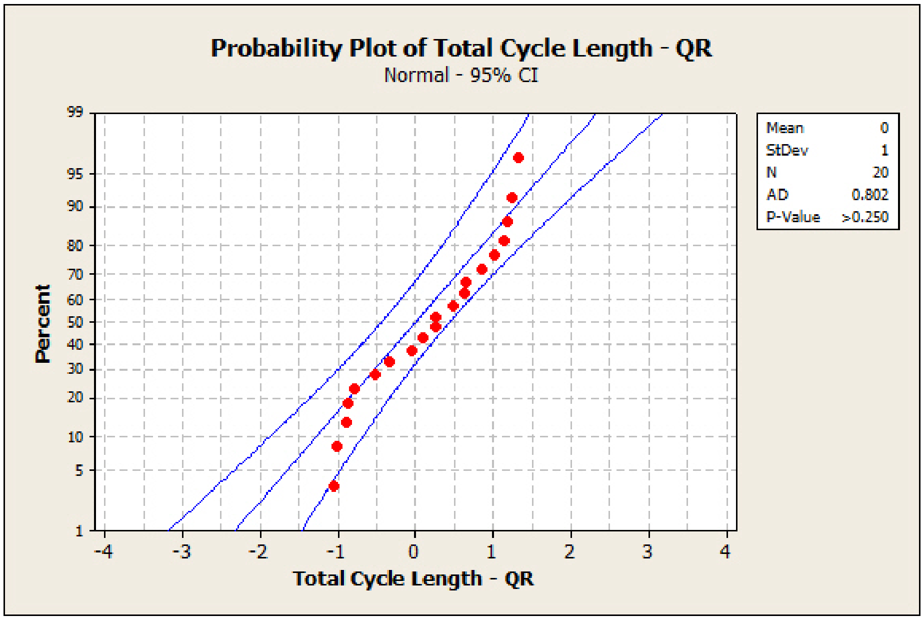


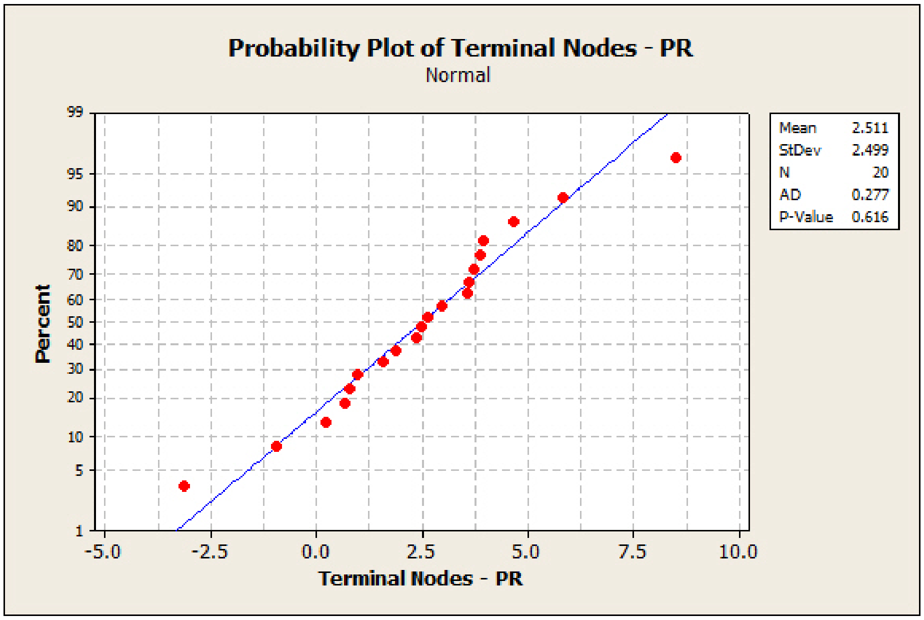
Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Connected Components - QR	20	0.378	1.315	0.294	(-0.238, 0.993)	1.28	0.215


Probability Plot of Cyclic Nodes - PR
Normal - 95% CI
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
-4
-3
-2
-1
0
1
2
3
4
Cyclic Nodes - PR
Mean 0
StDev 1
N 20
AD 0.848
P-Value >0.250

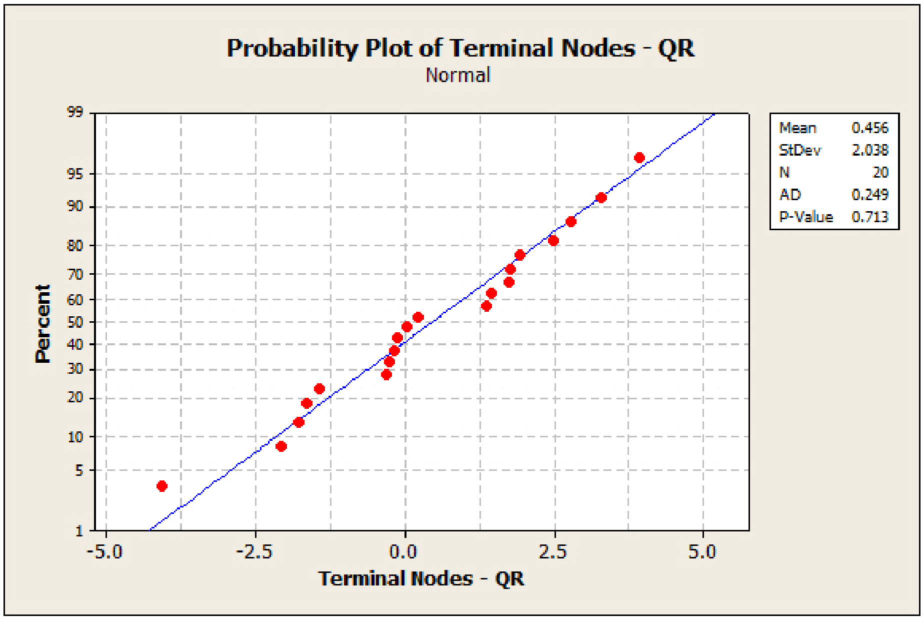

Probability Plot of Cyclic Nodes - QR
Normal - 95% CI
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
-4
-3
-2
-1
0
1
2
3
4
Cyclic Nodes - QR
Mean 0
StDev 1
N 20
AD 1.072
P-Value >0.250


Probability Plot of Total Cycle Length - PR
Normal - 95% CI
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
-4
-3
-2
-1
0
1
2
3
4
Total Cycle Length - PR
Mean 0
StDev 1
N 20
AD 1.654
P-Value 0.143

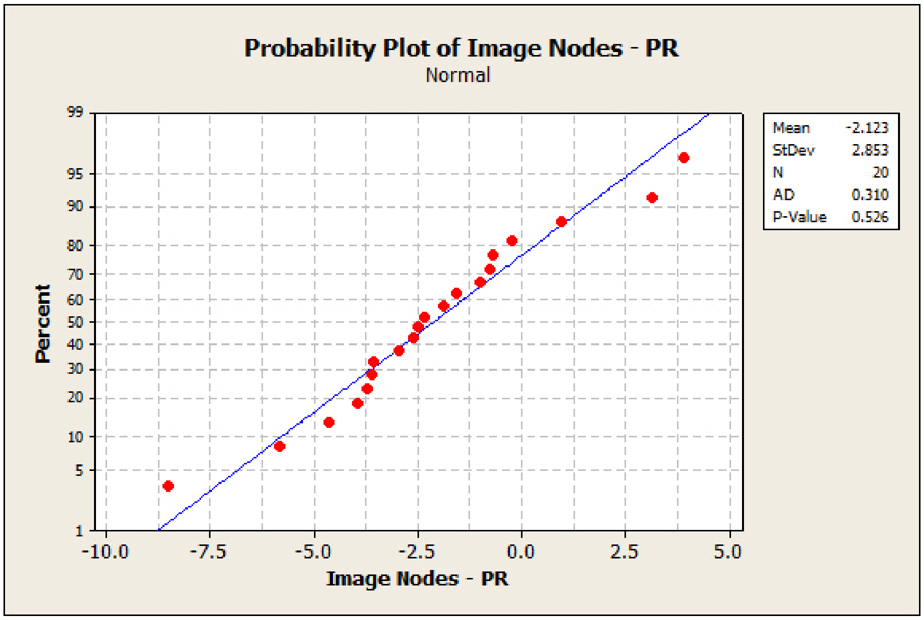

Probability Plot of Total Cycle Length - QR
Normal - 95% CI
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
-4
-3
-2
-1
0
1
2
3
4
Total Cycle Length - QR
Mean 0
StDev 1
N 20
AD 0.802
P-Value >0.250



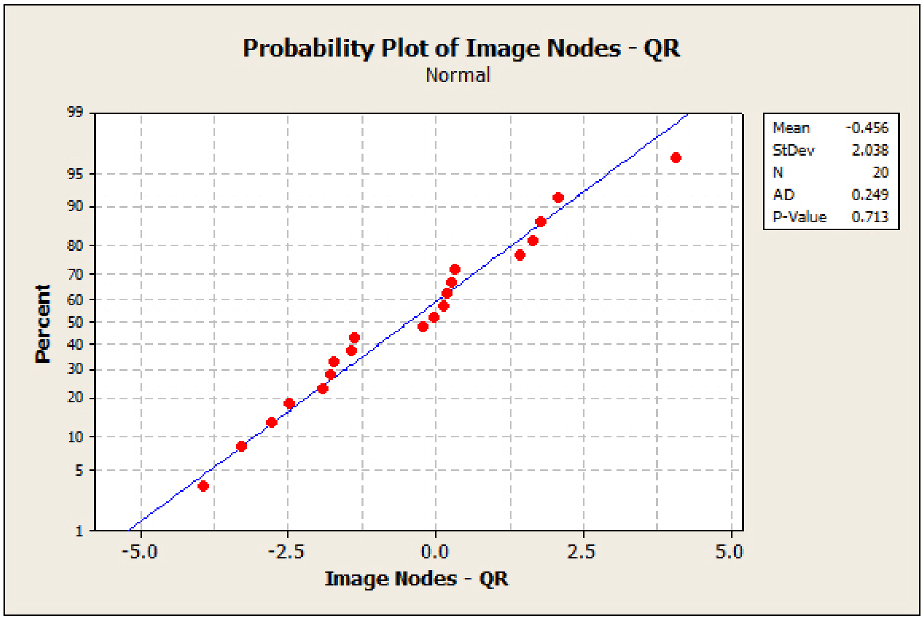
Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Terminal Nodes - PR	20	2.511	2.499	0.559	(1.341, 3.681)	4.49	0



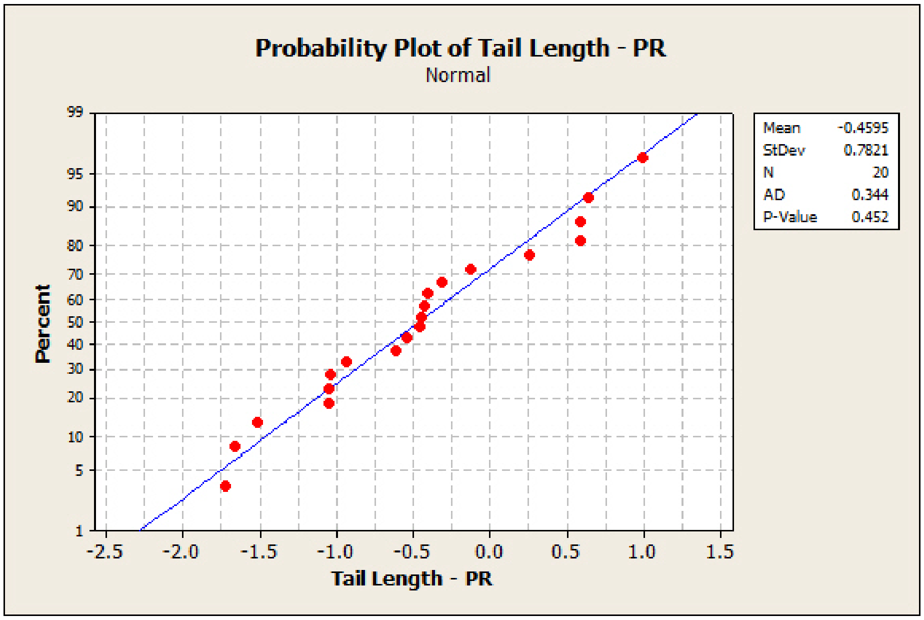
Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Terminal Nodes - QR	20	0.456	2.038	0.456	(-0.498, 1.410)	1	0.33



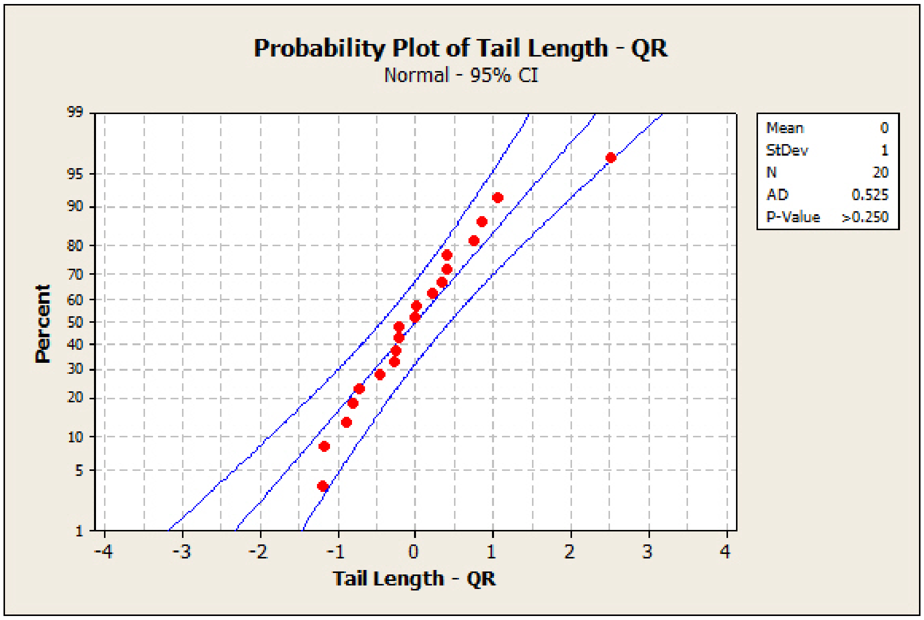
Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Image Nodes - PR	20	-2.123	2.853	0.638	(-3.458, -0.788)	-3.33	0.004



Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Image Nodes - QR	20	-0.456	2.038	0.456	(-1.410, 0.498)	-1	0.33



Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Tail Length - PR	20	-0.46	0.782	0.175	(-0.826, -0.093)	-2.63	0.017



C Asymptotic Approximations

This table shows the second terms that we calculated for each of the asymptotic approximations. The first terms can be found in Flajolet and Odlyzko's paper [3]. Note that the second terms for the Image Nodes and Terminal Nodes approximations were calculated using the expanded normalizing factor found on the bottom row of this table. We expanded the normalizing factor in hopes that it would compensate for the large second terms in the approximations of those characteristics, but it turned out that it made little difference. Thus, it was not used in the calculations for any of the other second terms.

First Two Terms for Asymptotic Approximations	
Connected Components	$\sim \frac{1}{2}\log n + \frac{1}{2}\gamma$, where $\gamma \approx 0.6351814227$
Cyclic Points	$\sim \sqrt{\frac{\pi n}{2}} - \frac{1}{3}$
Terminal Nodes	$\sim e^{-1}n + \frac{5.532822049n^{\frac{3}{2}}}{12n-1}$
Image Nodes	$\sim (1-e^{-1})n + \frac{9.506947588n^{\frac{3}{2}}}{12n-1}$
Cycle Length	$\sim \sqrt{\frac{\pi n}{8}} + \frac{n}{3}$
Tail Length	$\sim \sqrt{\frac{\pi n}{8}} - \frac{2n}{3}$
Normalizing Factor	$\sim \frac{e^n}{\sqrt{2\pi n}} - \frac{e^n}{12n\sqrt{2\pi n}}$