

# Mapping the Discrete Logarithm

Joshua Holden

Joint work with Daniel Cloutier, Nathan Lindle  
(Senior Theses),  
Aaron Blumenfeld, Max Brugger, Christina Frederick,  
Matthew Friedrichsen, Andrew Hoffman, Brian Larson,  
Marcus Mace, and Emily McDowell (RHIT REU)

Rose-Hulman Institute of Technology  
<http://www.rose-hulman.edu/~holden>



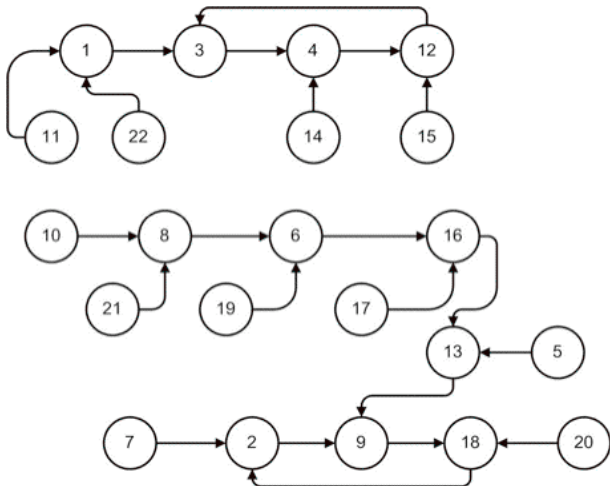
We are investigating the functional graph induced by the discrete exponentiation map  $x \mapsto g^x \pmod p$ .

**Definition** A *functional graph* is a directed graph such that each vertex must have exactly one edge directed out from it.

For each  $x$ , make an edge from the vertex  $x$  to the vertex  $f(x)$ .

**Question** How much does the discrete exponentiation functional graph (DEFG) look like a “random graph”?

We are investigating the functional graph induced by the discrete exponentiation map  $x \mapsto g^x \bmod p$ .



$$x \mapsto 3^x \bmod 23$$

The problem of inverting the discrete exponentiation map is called the **Discrete Logarithm Problem**.

$$x \mapsto y \equiv g^x \pmod{p}$$

Discrete exponentiation: easy.

$$x \leftarrow y \equiv g^x \pmod{p}$$

Discrete logarithm: thought to be hard.

The assumption that this problem is hard underlies the (assumed) security of several cryptographic protocols.

- ▶ Diffie-Hellman Key Agreement
- ▶ Blum-Micali Cryptographically Secure Pseudorandom Number Generator
- ▶ ElGamal Encryption
- ▶ ElGamal Digital Signature Scheme
- ▶ (Elliptic Curve DLP) Elliptic Curve Cryptography
- ▶ Etc.

Plus, discrete exponentiation makes pretty graphs.

# The structure of discrete exponentiation functional graphs is complicated enough to appear to be random.

We need to reconcile the additive structure with the multiplicative structure:

$$x \in \mathbb{Z}/p\mathbb{Z} \cong C_p$$

but

$$g^x \in (\mathbb{Z}/p\mathbb{Z})^* \cong C_{p-1}.$$

## Remarks

- ▶ All numbers are integers in  $\{1, 2, \dots, p-1\}$ .
- ▶  $p$  will be a prime  $\geq 3$  in all theorems.

## Our strategy will be:

1. Figure out what a “random graph” looks like.
2. Figure out what a discrete exponentiation functional graph (DEFG) looks like.
3. Compare.

## Our strategy will be:

1. Figure out what a “random graph” looks like.  
(Combinatorics, Analysis)
2. Figure out what a discrete exponentiation functional graph (DEFG) looks like.
3. Compare.

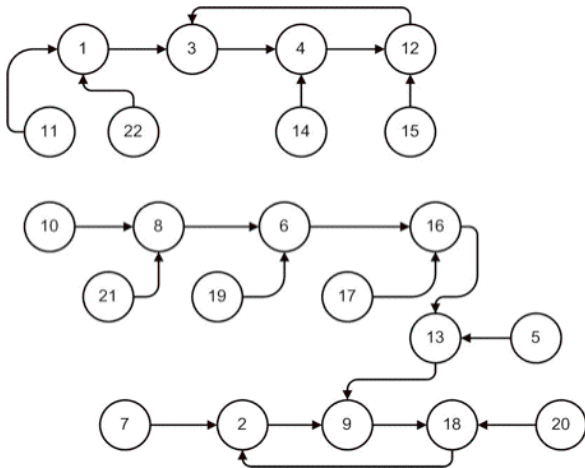
## Our strategy will be:

1. Figure out what a “random graph” looks like.  
(Combinatorics, Analysis)
2. Figure out what a discrete exponentiation functional graph (DEFG) looks like.  
(Number Theory, Computation)
3. Compare.

## Our strategy will be:

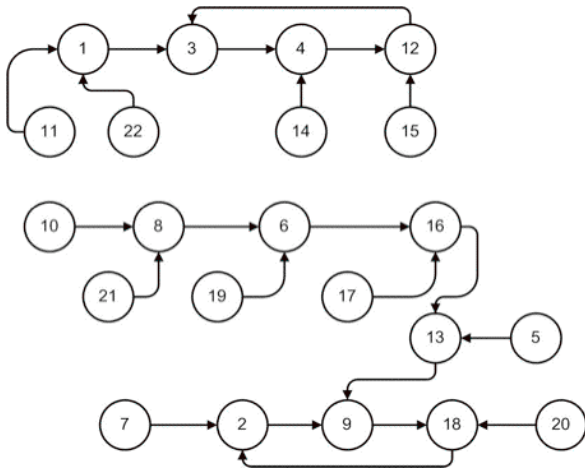
1. Figure out what a “random graph” looks like.  
(Combinatorics, Analysis)
2. Figure out what a discrete exponentiation functional graph (DEFG) looks like.  
(Number Theory, Computation)
3. Compare.  
(Statistics)

There are many graph measurements that we could examine.



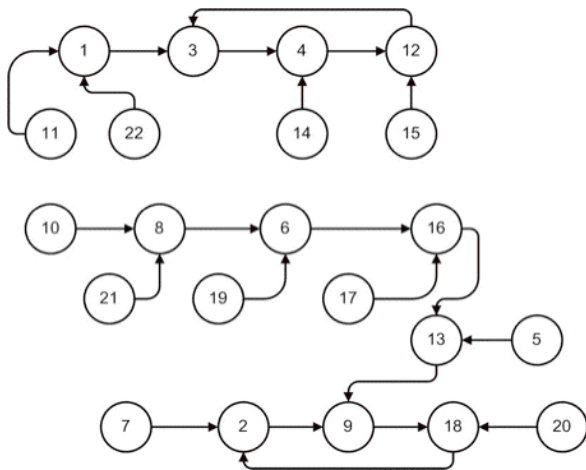
Number of connected components

There are many graph measurements that we could examine.



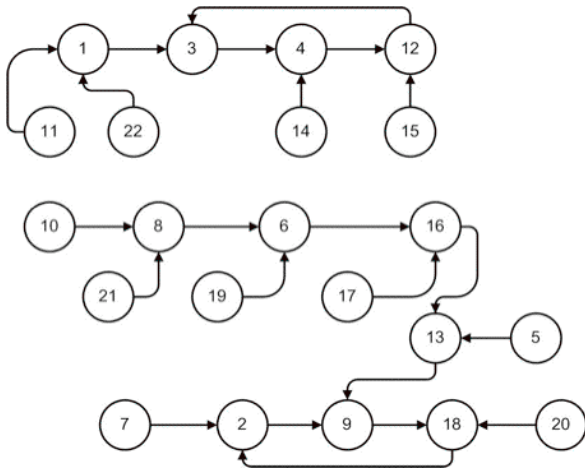
Number of cyclic nodes

There are many graph measurements that we could examine.



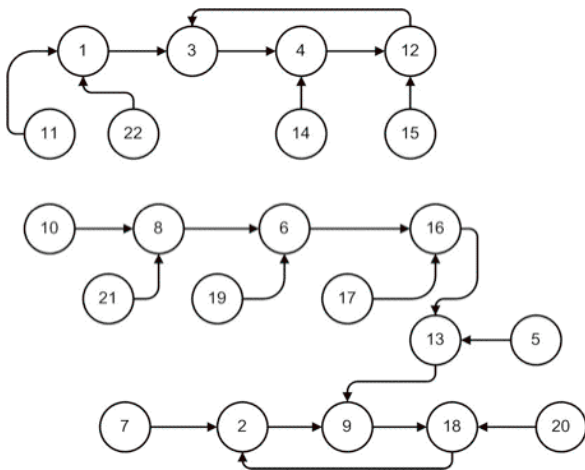
Number of image nodes

There are many graph measurements that we could examine.



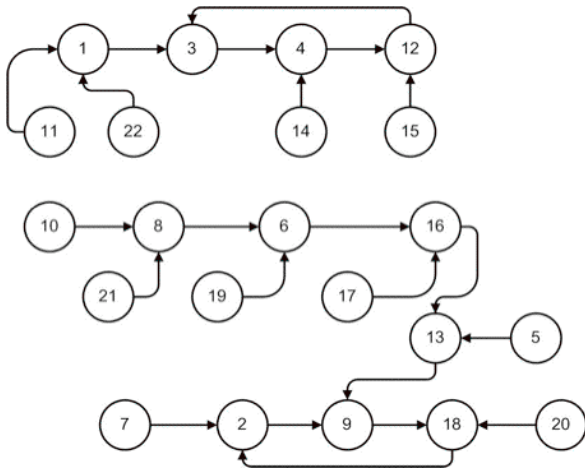
Average component size (as seen from a node)

There are many graph measurements that we could examine.



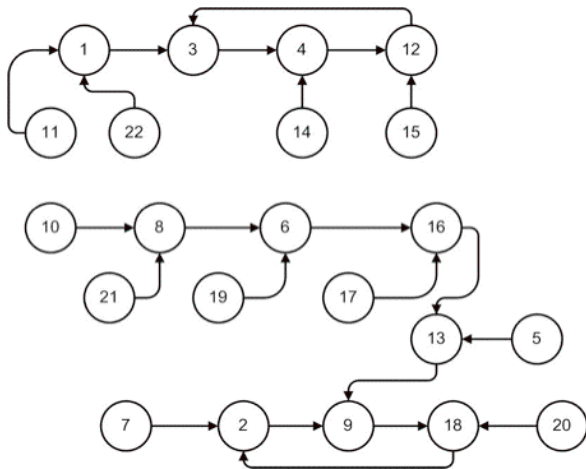
Average cycle length (as seen from a node)

There are many graph measurements that we could examine.



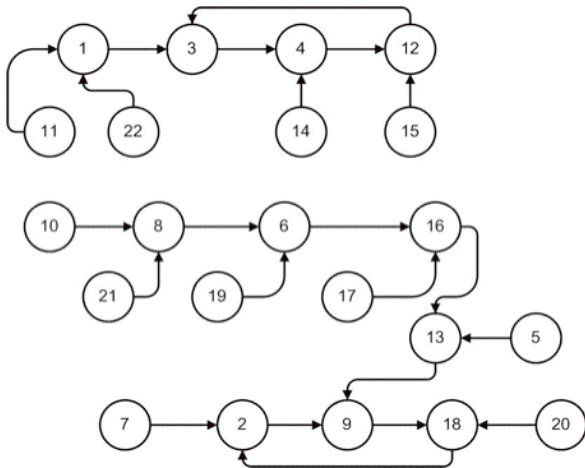
Average tail length (as seen from a node)

There are many graph measurements that we could examine.



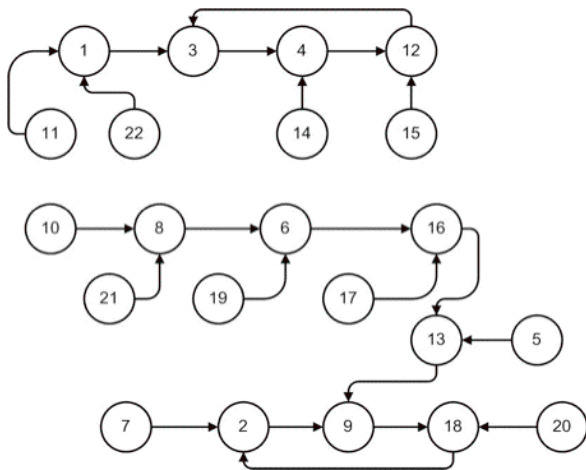
Maximum cycle length

There are many graph measurements that we could examine.



Maximum tail length

There are many graph measurements that we could examine.



Distribution of cycle sizes

## How do we measure a “random graph”?

(By “random graph” we mean a randomly chosen graph of a specified type on a specified number of nodes.)

**Answer** We count the total over all graphs of that type and divide by the number of graphs to get the expected value.

## How do we measure a “random graph”?

(By “random graph” we mean a randomly chosen graph of a specified type on a specified number of nodes.)

**Answer** We count the total over all graphs of that type and divide by the number of graphs to get the expected value.

Okay, wiseguy, so how do we do that?

## How do we measure a “random graph”?

(By “random graph” we mean a randomly chosen graph of a specified type on a specified number of nodes.)

**Answer** We count the total over all graphs of that type and divide by the number of graphs to get the expected value.

Okay, wiseguy, so how do we do that?

**Answer** Generating functions. In particular, exponential generating functions. (**Combinatorics**)

The expected mean values for the measurements of interest in a random functional graph of size  $n$  are:

$$\text{Number of components} \sim \frac{\ln(2n) + \gamma}{2}$$

$$\text{Number of cyclic nodes} \sim \sqrt{\pi n/2} - \frac{1}{3}$$

$$\text{Number of image nodes} \sim (1 - e^{-1})n$$

(Standard Results from Literature)

- ▶ All asymptotics are as  $n \rightarrow \infty$ .

The expected mean values for the measurements of interest in a random functional graph of size  $n$  are:

$$\textit{Average cycle length} \sim \sqrt{\pi n/8}$$

$$\textit{Average tail length} \sim \sqrt{\pi n/8}$$

$$\textit{Maximum cycle length} \sim c_1 \sqrt{\frac{\pi n}{2}} \approx 0.78248\sqrt{n}$$

$$c_1 = \int_0^\infty [1 - \exp(-\int_v^\infty e^{-u} \frac{du}{u})] dv$$

$$\textit{Maximum tail length} \sim \sqrt{2\pi n} \ln 2 \approx 1.73746\sqrt{n}$$

(Standard Results from Literature)

For each  $g = 1, 2, \dots, p - 1$  we build the graph; then we compute measurements over all graphs.

The first data set (Cloutier, 2006) was:

- ▶  $p = 100043 = (2)(50021) + 1$  (“safe prime”)
- ▶  $p = 100057 = (2^3)(3)(11)(379) + 1$
- ▶  $p = 106261 = (2^2)(3)(5)(7)(11)(23) + 1$

The original code was written in C++ by Daniel Cloutier (2006), using the RHIT parallel cluster. There were later revisions by Nathan Lindle (2008) and by Andrew Hoffman and Marc Mace (2009).

For each  $g = 1, 2, \dots, p - 1$  we build the graph; then we compute measurements over all graphs.

The first data set (Cloutier, 2006) was:

- ▶  $p = 100043 = (2)(50021) + 1$  (“safe prime”)
- ▶  $p = 100057 = (2^3)(3)(11)(379) + 1$
- ▶  $p = 106261 = (2^2)(3)(5)(7)(11)(23) + 1$

The original code was written in C++ by Daniel Cloutier (2006), using the RHIT parallel cluster. There were later revisions by Nathan Lindle (2008) and by Andrew Hoffman and Marc Mace (2009).

**Result** Arbitrary random graphs aren't a good model.

## So what is a better model?

We will use some **number theory** to understand why DEFG's aren't quite as "random" as the graphs we have analyzed so far.  
— We need to account for a little bit more structure.

## So what is a better model?

We will use some **number theory** to understand why DEFG's aren't quite as "random" as the graphs we have analyzed so far.  
— We need to account for a little bit more structure.

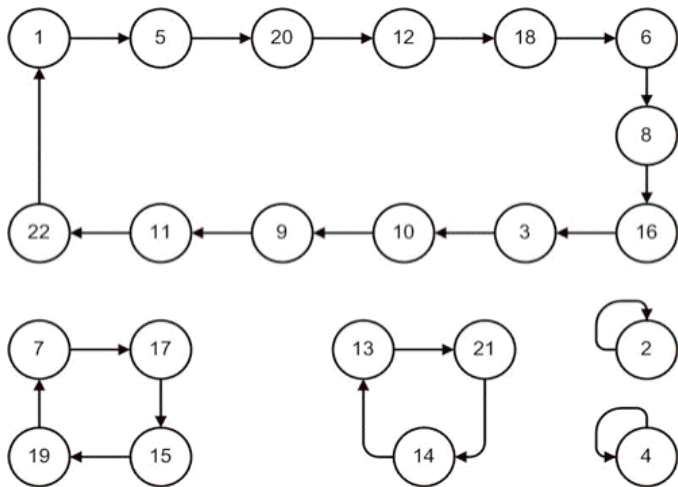
**Definition** An *m*-ary functional graph is a functional graph where each node has in-degree of exactly zero or *m*.

**Proposition** If  $m \mid (p - 1)$  then there are  $\phi\left(\frac{p-1}{m}\right)$  *m*-ary functional graphs produced by varying *g* for a given *p*.

Furthermore, the values of *g* that produce an *m*-ary graph are precisely those for which *g* is a strictly perfect *m*th power modulo *p*.

**Proposition** The number of image nodes in an *m*-ary functional graph is exactly  $\frac{p-1}{m}$ .

A unary functional graph is the same as a permutation.



$$x \mapsto 5^x \pmod{23}$$

The expected mean values for the measurements of interest in a random permutation of size  $n$  are:

$$\begin{aligned} \text{Number of components} &= \sum_{i=1}^n \frac{1}{i} = \Psi(n+1) + \gamma \\ &\sim \ln n + \gamma \end{aligned}$$

$$\text{Average cycle length} = \frac{n+1}{2}$$

$$\text{Maximum cycle length} \sim c_2 n \approx 0.62432965n$$

$$c_2 = \int_0^\infty \left[ 1 - \exp\left(-\int_v^\infty e^{-u} \frac{du}{u}\right) \right] dv$$

where  $\Psi(x) = \frac{d}{dx} \ln(\Gamma(x))$ .

(Standard Results from Literature)

The expected variances for the measurements of interest in a random permutation of size  $n$  are:

$$\begin{aligned} \text{Number of components} &= \Psi(n+1) + \Psi'(n+1) + \gamma - \frac{\pi^2}{6} \\ &\sim \ln n + \gamma - \frac{\pi^2}{6} \\ \text{Average cycle length} &= \frac{n^2 - 1}{12} \end{aligned}$$

(Standard Results from Literature)

Data was collected for three primes by Cloutier (2006), and 30 more primes by Hoffman (2009).

The values of  $g$  (primitive roots) which produced permutations were separated out.

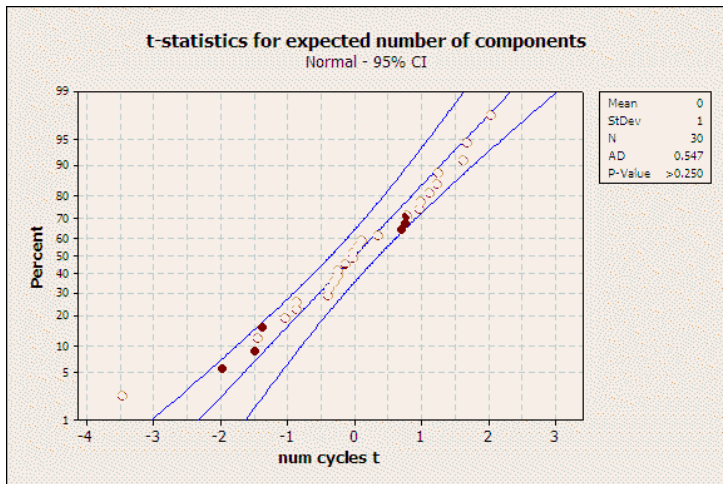
For each prime, we collected the data for all values of  $g$  and obtained a  $t$ -test statistic for the expected means and a  $P$ -value from a  $\chi$ -squared test for the expected variances.

These tests (indirectly) measure the probability that a randomly chosen set of graphs would behave like our graphs.

Random variation predicts the  $t$ -statistics should have a standard normal distribution and the  $P$ -values should have a uniform distribution.

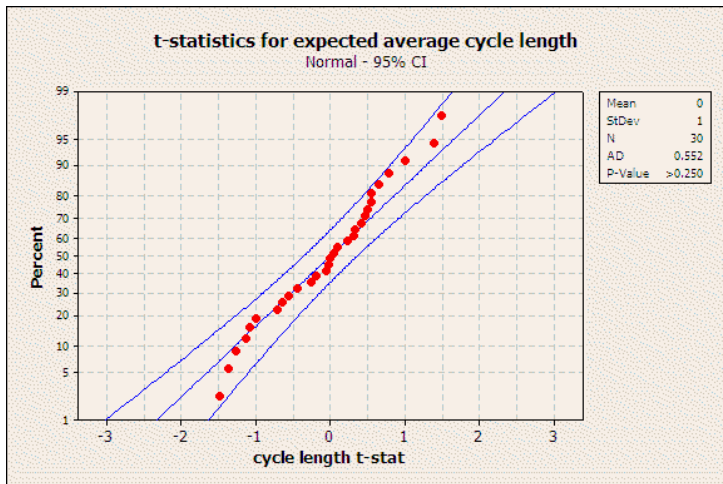
**Note:** In all situations low  $P$ -values are anomalous.

Some of the measurements behaved as predicted.



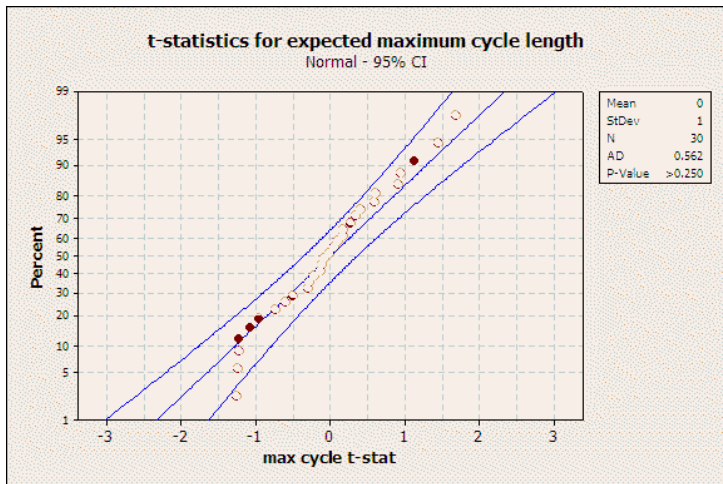
Number of components

Some of the measurements behaved as predicted.



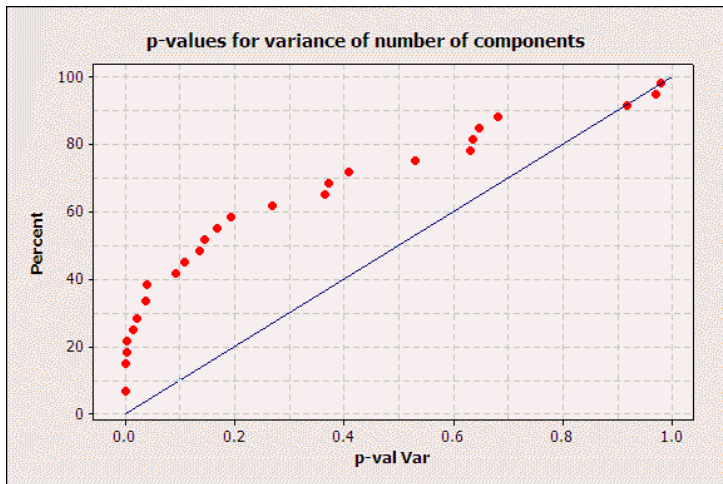
Average cycle length

Some of the measurements behaved as predicted.



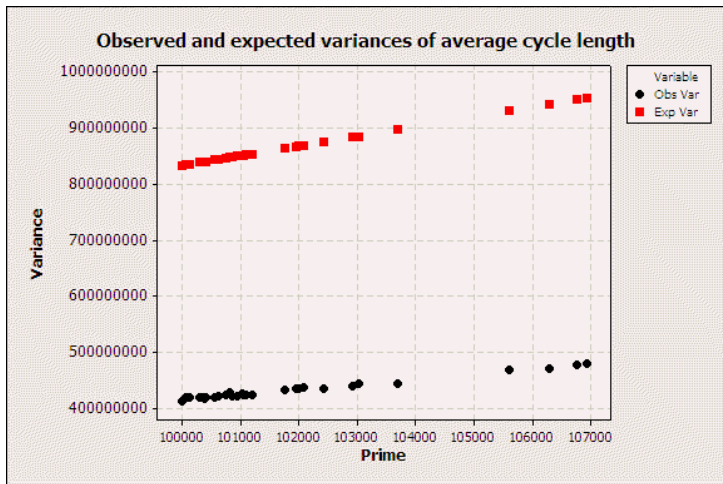
Maximum cycle length

Some of them looked okay at first but turned out to be significantly off.



Variance of number of components

And some of them were not even close.



Variance of average cycle length

For binary functional graphs, not all of the results we need seem to be in the literature.

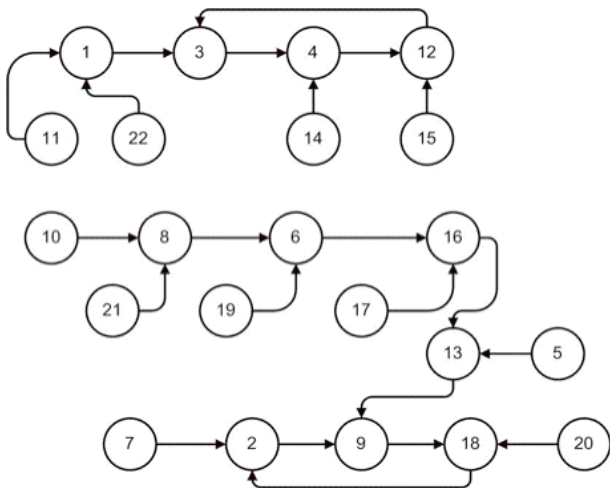
We used a general technique which was given by Flajolet and Odlyzko:

- ▶ Explicitly define the structure.
- ▶ Convert to exponential generating functions.
- ▶ “Mark” the structures of interest.
- ▶ Compute expected value generating functions.
- ▶ Perform “automatic” singularity analysis to get asymptotic form of coefficients.
- ▶ Normalize.

In some cases a more precise technique was also used:

- ▶ Explicitly define the structure.
- ▶ Convert to exponential generating functions.
- ▶ “Mark” the structures of interest.
- ▶ Compute expected value generating functions.
- ▶ Obtain DEs whose solutions are the generating functions.
- ▶ Use the DEs to find recursion relations.
- ▶ Solve the recursion relations.
- ▶ Normalize.

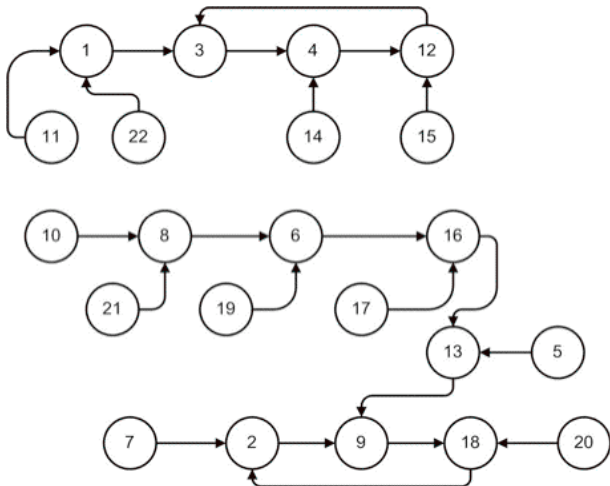
A binary functional graph (bfg) is a set of components.



$$x \mapsto 3^x \pmod{23}$$



A binary tree consists of a root node or a root node and two interchangeable binary tree branches.



$$x \mapsto 3^x \pmod{23}$$

We can express this structure in a shorthand notation.

BinFunGraph = set(Components)  
Component = cycle(Node\*BinaryTree)  
BinaryTree =  
    Node + Node\*set(BinaryTree, cardinality = 2)  
Node = Atomic Unit

Then we can convert the shorthand to generating functions.

$$f(z) = e^{c(z)} \quad (\text{Binary functional graphs})$$

$$c(z) = \ln \frac{1}{1 - zb(z)} \quad (\text{Connected components})$$

$$b(z) = z + \frac{1}{2}zb^2(z) \quad (\text{Binary trees})$$

And solve the generating functions.

$$f(z) = \frac{1}{\sqrt{1-2z^2}} \quad (\text{Binary functional graphs})$$

$$c(z) = \ln \frac{1}{\sqrt{1-2z^2}} \quad (\text{Connected components})$$

$$b(z) = \frac{1 - \sqrt{1-2z^2}}{z} \quad (\text{Binary trees})$$

One way to get information about the coefficients of generating functions is to use complex analysis.

**Notation** If  $f(z)$  is a generating function,  $[z^n]f(z)$  is the coefficient of  $z^n$  in  $f(z)$ .

**Theorem (Example of a “Transfer Theorem”)**

*Let  $\alpha \in \mathbb{R} \setminus \{0, 1, 2, 3, \dots\}$ . Then*

$$[z^n](1 - z)^\alpha \sim \frac{n^{-\alpha-1}}{\Gamma(-\alpha)}.$$

CAS's like Maple can apply these theorems automatically.

For example, we can expand  $f(z)$  around the singularity at  $z = 1/\sqrt{2}$ .

- ▶ First-order:  $[z^n]f(z) \sim \frac{2^{n/2}}{\sqrt{\pi n/2}}$ .
- ▶ Second-order:  $[z^n]f(z) \sim \frac{2^{n/2}}{\sqrt{\pi n/2}} - \frac{2^{n/2}}{4n\sqrt{\pi n/2}}$ .

We did “automatic analysis” using a Maple package. (Flajolet, Salvy, Zimmerman, et al.) It applies transfer theorems to convert the type of singularity to the asymptotics of the coefficients.

Or we can find a differential equation satisfied by  $f(z)$  and use it to derive a recurrence relation.

If  $f(z) = \sum_{n=0}^{\infty} a_n z^n$  then

$$a_{n+2} = \frac{2n+2}{n+2} a_n$$

$$a_0 = 1$$

$$a_1 = 0$$

The two methods agree: The total number of (labeled) binary functional graphs of size  $n$  is  $[z^n/n!]f(z)$ , or

$$\frac{n!^2}{\left(\frac{n}{2}\right)!^2 2^{n/2}} \sim n! \left( \frac{2^{n/2}}{\sqrt{\pi n/2}} - \frac{2^{n/2}}{4n\sqrt{\pi n/2}} \right).$$

## Now how can we find out what a “random” binary functional graph “looks like”?

We “mark” the structure of interest in  $f(z)$  with  $u$  and get  $\xi(u, z)$  such that

$$[\xi(u, z)]_{u=1} = f(z).$$

$$\xi_1(u, z) = e^{uc(z)} \quad (\text{Components})$$

$$\xi_2(u, z) = \frac{1}{1 - uz b(z)} \quad (\text{Cyclic Nodes})$$

$$\xi_3(u, z) = \frac{1}{\sqrt{1 - 2uz^2}} \quad (\text{Terminal Nodes})$$

Differentiating gives us generating functions for the total values of our measurements.

$$\Xi(z) = \left[ \frac{\partial}{\partial u} \xi(u, z) \right]_{u=1}$$

$$\Xi_1(z) = \frac{1}{1 - zb(z)} \log \left( \frac{1}{1 - zb(z)} \right) \quad (\text{Components})$$

$$\Xi_2(z) = \frac{zb(z)}{(1 - zb(z))^2} \quad (\text{Cyclic Nodes})$$

$$\Xi_3(z) = \frac{z^2}{(1 - 2z^2)^{3/2}} \quad (\text{Terminal Nodes})$$

Differentiating twice gives us generating functions for the sums of squares of our measurements.

$$\Xi^*(z) = \left[ \frac{\partial}{\partial u} u \left( \frac{\partial}{\partial u} \xi(u, z) \right) \right]_{u=1}$$

Then we use the formula  $\text{Var}(x) = \frac{1}{N} \left( \sum_{i=1}^N x_i^2 \right) - \bar{x}^2$ .

We can again expand these generating functions around the singularity at  $z = 1/\sqrt{2}$ .

$$[z^n]\Xi_1(z) \sim \frac{2^{n/2}(\log n + \gamma + \log 2)}{2\sqrt{2\pi n}} \quad (\text{Components})$$

$$[z^n]\Xi_2(z) \sim \frac{2^{n/2}(\sqrt{\pi n} - \sqrt{2})}{2\sqrt{\pi n}} \quad (\text{Cyclic Nodes})$$

$$[z^n]\Xi_3(z) \sim \frac{2^{n/2}(4n - 1)}{8\sqrt{2\pi n}} \quad (\text{Terminal Nodes})$$

Or we can derive the recurrence relations.

For example, if  $\Xi(z) = \sum_{n=0}^{\infty} a_n z^n$  (Average Cycle Length) then

$$a_{n+4} = 4a_{n+2} - 4a_n$$

$$a_0 = 0$$

$$a_1 = 0$$

$$a_2 = 2$$

$$a_3 = 0$$

The values of we get are totals over all graphs, so we divide by the number of graphs to normalize.

For example, the expected number of components is:

$$\frac{[z^n/n!]\Xi_1(z)}{[z^n/n!]f(z)} \sim \frac{n! \left( \frac{2^{n/2}(\log n + \gamma + \log 2)}{2\sqrt{2\pi n}} \right)}{n! \left( \frac{2^{n/2}}{\sqrt{\pi n/2}} \right)}$$
$$= \frac{\log n + \gamma + \log 2}{2}$$

The other measurements of interest are more complicated, but similar.

The expected mean values for the measurements of interest in a random bfg of size  $n$  are:

$$\begin{aligned} \text{Number of components} &= \binom{1}{2} \Psi \left( \frac{n}{2} + \frac{1}{2} \right) + \binom{1}{2} \gamma + \ln(2) \\ &\sim \frac{\ln(2n) + \gamma}{2} \end{aligned}$$

$$\begin{aligned} \text{Number of cyclic nodes} &= \frac{\sqrt{\pi} \Gamma(\frac{n}{2} + 1) - \Gamma(\frac{n}{2} + \frac{1}{2})}{\Gamma(\frac{n}{2} + \frac{1}{2})} \\ &\sim \sqrt{\pi n / 2} - 1 \end{aligned}$$

(Dan Cloutier, 2006 and Nathan Lindle, 2008)

The expected mean values for the measurements of interest in a random bfg of size  $n$  are:

$$\begin{aligned} \text{Average cycle length} &= \frac{\left(\frac{1}{2}\right)\sqrt{\pi}\Gamma\left(\frac{n}{2} + 1\right)}{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)} \\ &\sim \sqrt{\pi n/8} \end{aligned}$$

$$\begin{aligned} \text{Average tail length} &= \frac{\sqrt{\pi}\Gamma\left(2 + \frac{n}{2}\right) - n\Gamma\left(\frac{n}{2} + \frac{1}{2}\right) - \Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{n\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)} \\ &\sim \sqrt{\pi n/8} \end{aligned}$$

(Dan Cloutier, 2006 and Nathan Lindle, 2008)

The expected mean values for the measurements of interest in a random bfg of size  $n$  are:

(Cloutier, 2006)

$$\textit{Maximum cycle length} \quad \sim c_1 \sqrt{\frac{\pi n}{2}} \approx 0.78248 \sqrt{n}$$

$$c_1 = \int_0^\infty \left[ 1 - \exp \left( - \int_v^\infty e^{-u} \frac{du}{u} \right) \right] dv$$

(Holden, 2009)

$$\begin{aligned} \textit{Maximum tail length} \quad &\sim \sqrt{2\pi n} \ln 2 - \frac{\ln n}{2} - 2 - \frac{\gamma}{2} + \frac{3 \ln 2}{2} \\ &\approx 1.73746 \sqrt{n} - .5 \ln n - 1.24889 \end{aligned}$$

The expected variances for the measurements of interest in a random bfg of size  $n$  are:

*Number of components*

$$\begin{aligned}
 &= \frac{1}{2}\Psi\left(\frac{n}{2} + \frac{1}{2}\right) + \frac{\gamma}{2} + \ln(2) + \frac{\gamma^2}{4} + \gamma \ln(2) + \ln(2)^2 + \sum_{l=0}^{n/2-1} \frac{\Psi(l + \frac{1}{2})}{(2l+1)} - \frac{1}{4}\Psi\left(\frac{n}{2} + \frac{1}{2}\right)^2 \\
 &\sim \frac{2 \ln(n) - \ln(2)^2 - 2\gamma \ln(2) - \gamma^2}{4}
 \end{aligned}$$

*Number of cyclic nodes*

$$\begin{aligned}
 &= \frac{2\Gamma(\frac{n}{2} + \frac{1}{2})^2 + 4\frac{n}{2}\Gamma(\frac{n}{2} + \frac{1}{2})^2 - \sqrt{\pi}\Gamma(\frac{n}{2} + 1)\Gamma(\frac{n}{2} + \frac{1}{2}) - \pi\Gamma(\frac{n}{2} + 1)^2}{\Gamma(\frac{n}{2} + \frac{1}{2})^2} \\
 &\sim \left(2 - \frac{\pi}{2}\right)n - \frac{\sqrt{2\pi n}}{2} + 2 - \frac{\pi}{4}
 \end{aligned}$$

(Nathan Lindle, 2008)

The expected variances for the measurements of interest in a random bfg of size  $n$  are:

*Average cycle length*

$$\begin{aligned}
 &= \frac{1}{12} \frac{(-3\pi\Gamma(\frac{n}{2} + 1))^2 - 6\sqrt{\pi}\Gamma(\frac{n}{2} + 1)\Gamma(\frac{n}{2} + \frac{1}{2}) + 16\frac{n}{2}\Gamma(\frac{n}{2} + \frac{1}{2})^2 + 8\Gamma(\frac{n}{2} + \frac{1}{2})^2}{\Gamma(\frac{n}{2} + \frac{1}{2})^2} \\
 &\sim \left(\frac{2}{3} - \frac{\pi}{8}\right)n - \frac{\sqrt{2\pi n}}{4} + \frac{2}{3} - \frac{\pi}{16}
 \end{aligned}$$

*Average tail length*

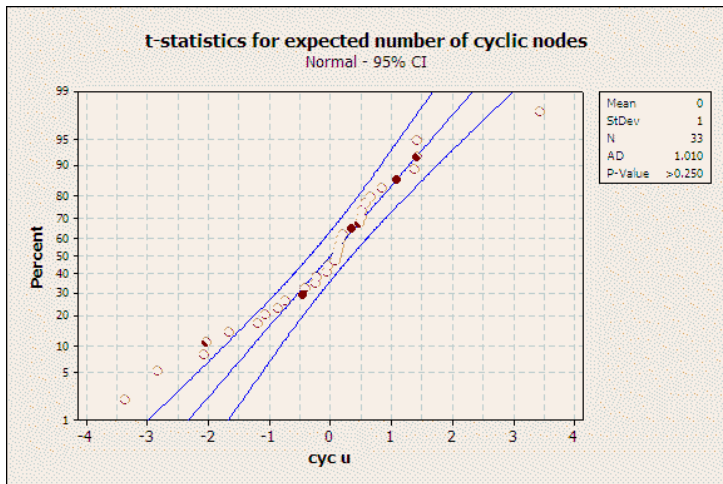
$$\begin{aligned}
 &= \frac{1}{6} \frac{(18\Gamma(\frac{n}{2} + \frac{3}{2}) + 8\Gamma(\frac{n}{2} + \frac{3}{2})\frac{n}{2} - 9\frac{n}{2}\sqrt{\pi}\Gamma(\frac{n}{2} + 1) - 9\sqrt{\pi}\Gamma(\frac{n}{2} + 1))}{(\frac{n}{2})\Gamma(\frac{n}{2} + \frac{1}{2})} \\
 &\quad - \frac{1}{4} \frac{(\sqrt{\pi}\Gamma(\frac{n}{2} + 1) + \frac{n}{2}\sqrt{\pi}\Gamma(\frac{n}{2} + 1) - 2\Gamma(\frac{n}{2} + \frac{3}{2}))^2}{(\frac{n}{2})^2\Gamma(\frac{n}{2} + \frac{1}{2})^2} \\
 &\sim \left(\frac{2}{3} - \frac{\pi}{8}\right)n - \frac{\sqrt{2\pi n}}{4} + \frac{8}{3} - \frac{9\pi}{16}
 \end{aligned}$$

BFG data was collected for three primes by Cloutier (2006), and 30 more primes by Lindle (2008).

Again, the values of  $g$  which produced binary functional graphs were separated out.

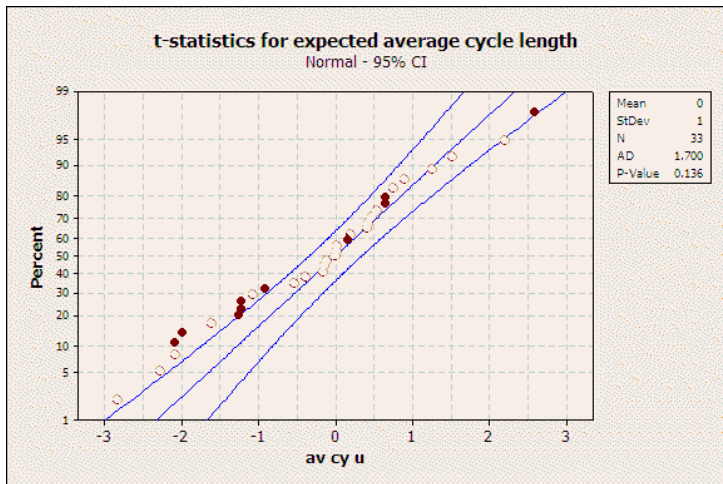
For each prime we collected the data for all values of  $g$  and obtained  $t$ -test statistics for the means and variances.

Again, some of the measurements behaved as predicted.



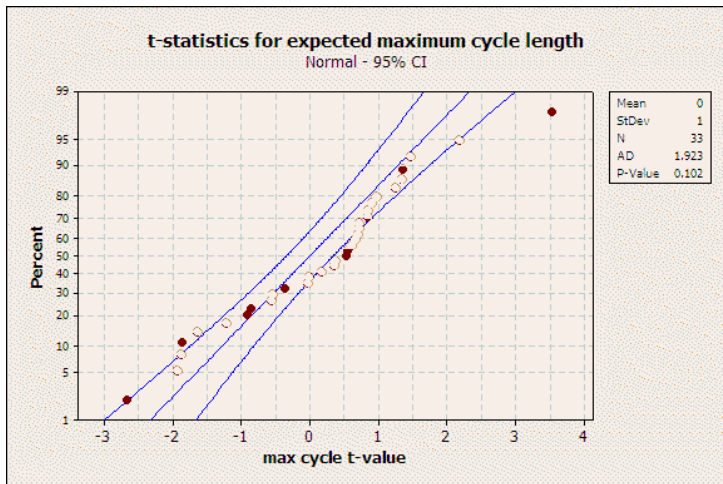
Number of cyclic nodes

Again, some of the measurements behaved as predicted.



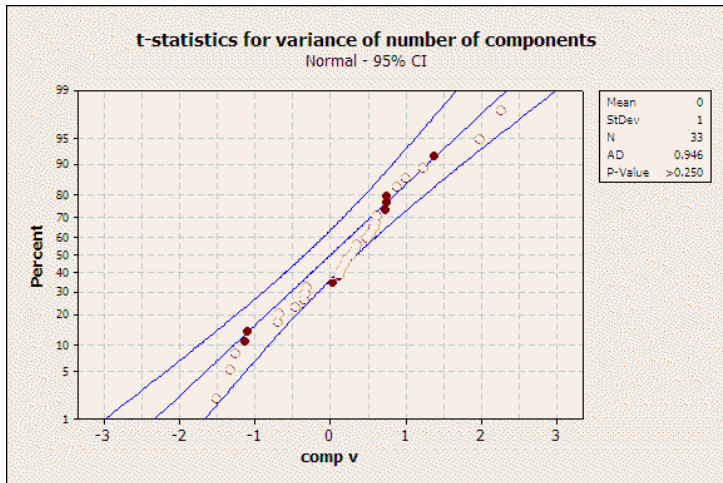
Average cycle length

Again, some of the measurements behaved as predicted.



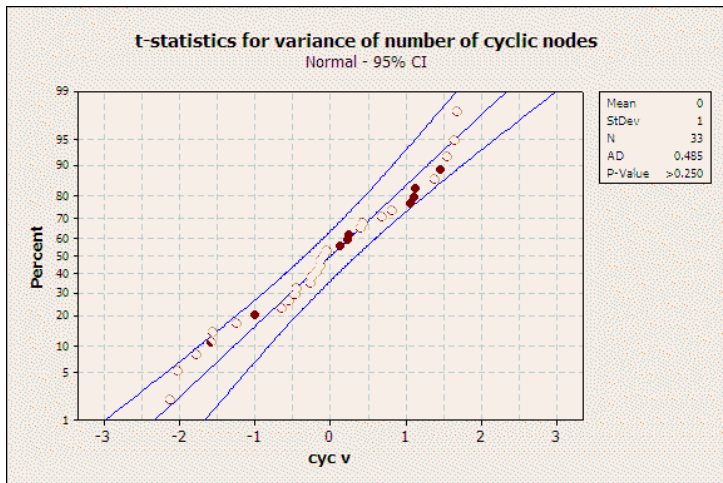
Maximum cycle length

Again, some of the measurements behaved as predicted.



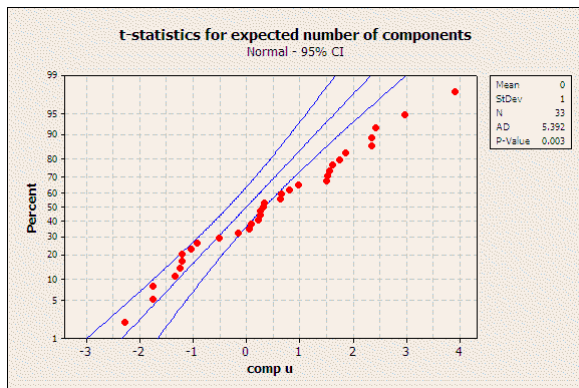
Variance of number of components

Again, some of the measurements behaved as predicted.



Variance of number of cyclic nodes

Some of them were a little odd.

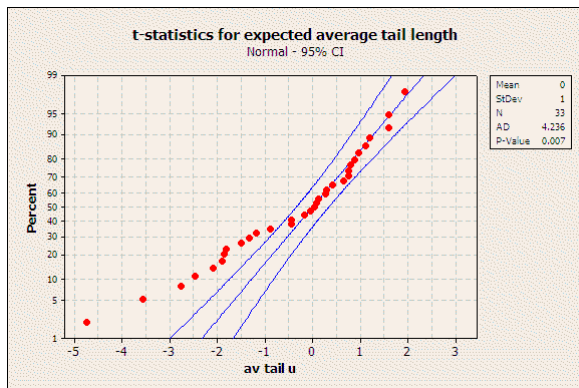


Test of  $\mu = 0$  vs not = 0

Variable	N	Mean	StDev	SE Mean	T	P
comp u	33	0.457	1.508	0.263	1.74	0.091

Number of components

Some of them were a little odd.

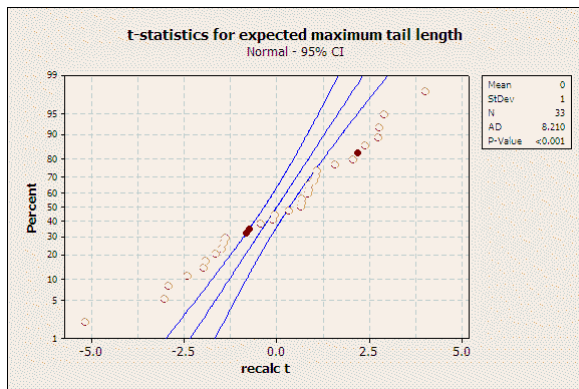


Test of  $\mu = 0$  vs not = 0

Variable	N	Mean	StDev	SE Mean	T	P
av tail u	33	-0.414	1.585	0.276	-1.50	0.143

Average tail length

Some of them were a little odd.

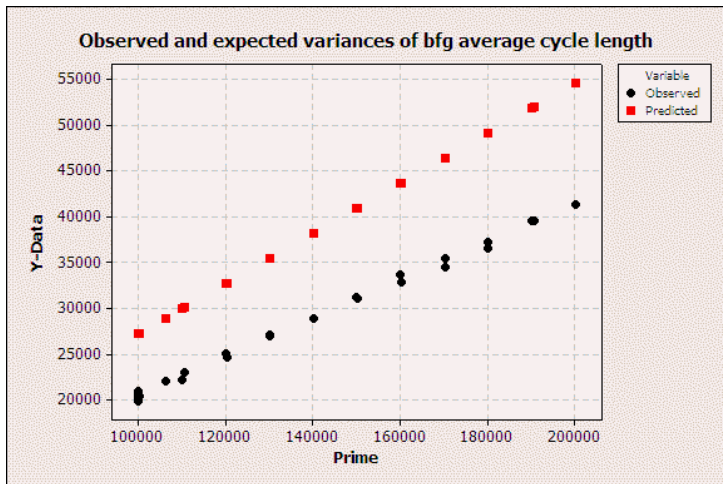


Test of  $\mu = 0$  vs not = 0

Variable	N	Mean	StDev	SE Mean	T	P
recalc t	33	0.083	2.026	0.353	0.23	0.816

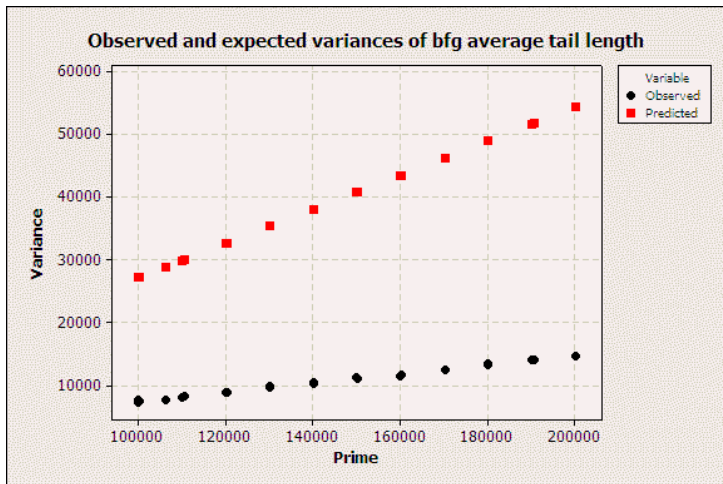
Maximum tail length

And some of them weren't even close.



Variance of average cycle length

And some of them weren't even close.

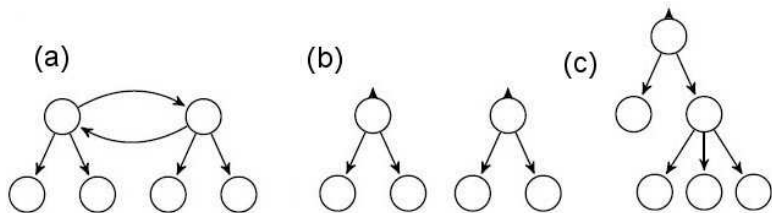


Variance of average tail length

For ternary functional graphs, a slightly different technique was used:

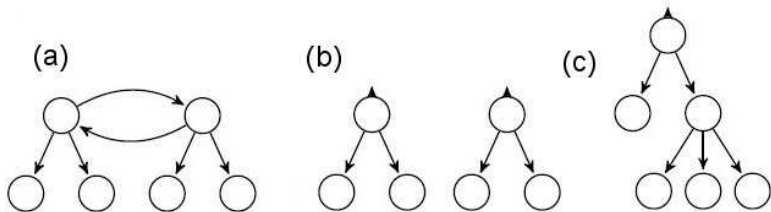
- ▶ Explicitly define the structure.
- ▶ Convert to exponential generating functions.
- ▶ “Mark” the structures of interest.
- ▶ Compute expected value generating functions.
- ▶ Obtain DEs whose solutions are the generating functions.
- ▶ Use the DEs to find recursion relations.
- ▶ Recursively compute expected values.
- ▶ Normalize.

A ternary functional graph (tfg) is a set of components.



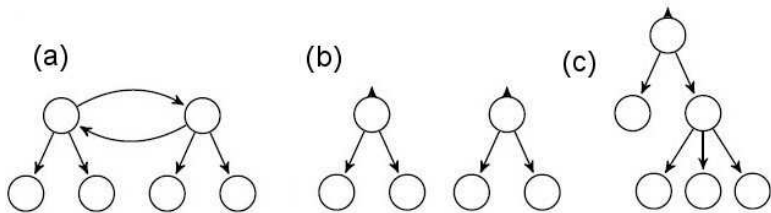
The ternary functional graphs of size 6.

Each component is a cycle of nodes, each of which is connected to two ternary trees.



The ternary functional graphs of size 6.

A ternary tree consists of a root node or a root node and three interchangeable ternary tree branches.



The ternary functional graphs of size 6.

We can again express this structure in a shorthand notation.

TernFunGraph = set(Components)  
Component = cycle(Node\*set(TernTree, cardinality=2)  
TernTree =  
Node + Node\*set(TernTree, cardinality = 3)  
Node = Atomic Unit

And convert the shorthand to generating functions.

$$f(z) = e^{c(z)} \quad (\text{Ternary functional graphs})$$

$$c(z) = \ln \frac{1}{1 - \frac{z}{2}t^2(z)} \quad (\text{Connected components})$$

$$t(z) = z + \frac{1}{3!}zt^3(z) \quad (\text{Ternary trees})$$

Note the increasing difficulty in solving explicitly for the tree function.

The expected mean value generating functions for the measurements of interest in a random tfg of size  $n$  are:

*Number of components:*  $\left[ \frac{\partial}{\partial u} e^{uc(z)} \right]_{u=1}$

*Number of cyclic nodes:*  $\left[ \frac{\partial}{\partial u} e^{\ln\left(\frac{1}{1-\frac{1}{2}uzt^2(z)}\right)} \right]_{u=1}$

(Brugger, 2008)

*Average tail length:*  $\left[ \frac{\partial}{\partial u} \frac{zut(z)}{\left(1 - \frac{1}{2}zt^2(z)\right)^2 \left(1 - \frac{1}{2}uzt^2(z)\right)} \right]_{u=1}$

(Max Brugger and Christina Frederick, 2007)

For the following mean value measurements, a “double marking” technique was used:

*Average component size:*

$$\left[ \frac{\partial^2}{\partial u \partial w} e^{c(z)} \ln \left( \frac{1}{1 - \frac{1}{2} u w z t^2 (u w z)} \right) \right]_{u=1, w=1}$$

*Average cycle length:*

$$\left[ \frac{\partial^2}{\partial u \partial w} e^{c(z)} \ln \left( \frac{1}{1 - \frac{1}{2} u w z t^2 (w z)} \right) \right]_{u=1, w=1}$$

(Brugger and Frederick, 2007)

Generating functions for maximum measurements have not yet been derived in the ternary case.

Data was collected on three measurements for nine primes by Brugger and Frederick (2007).

- ▶ number of components
- ▶ average number of cyclic nodes
- ▶ average cycle length

Again, the values of  $g$  which produced ternary functional graphs were separated out.

Results so far look similar to predictions but statistical tests have not yet been done.

We can also look at the expected distribution of cycle lengths in our graphs.

The expected proportion of random permutations having  $k$  cycles of length  $j$  approaches

$$\frac{1}{e^{(1/j)} j^k k!}$$

as the size  $n$  of the permutation goes to  $\infty$ .

(Standard Result from Literature)

## Data was collected for 30 primes by Hoffman (2009).

Frequencies were recorded for cycles of lengths 1, 2, 3, 5, 7, 10, and 20.

Example (not typical):  $p = 102061$

No. of 2-cyc.	Observed	Predicted
0	14139	14149
1	7082	7075
2	1772	1769
3	293	295
> 3	42	41

$P$ -value = 1.000

For each prime and each cycle length we collected the data for all values of  $g$  and obtained a  $P$ -value from a  $\chi$ -squared test for the expected distribution.

Anderson-Darling tests were conducted to determine if the distributions of the  $P$ -values were uniform.

Cycle Length	Test Statistic	Reject Uniformity?
1	31.74	Yes
2	3.28	Yes
3	1.35	No
5	1.10	No
7	0.93	No
10	0.57	No
20	1.08	No

**Result** Not always, especially for small cycles.

This is perhaps not that surprising, given the extra structure associated with 1-cycles and 2-cycles.

For example:

Theorem (Holden and Moree, 2004)

*There are  $\gg x/\ln x$  primes  $p \leq x$  such that the number of 1-cycles (fixed points) in the map  $x \mapsto g^x \pmod p$  averaged over primitive roots  $g$  is*

$$1 \pm \frac{p^{0.8313} d(p-1)^3 (2 + \ln p)}{\phi(p-1)}.$$

This suggests that the mean number of 1-cycles is predicted correctly but that the distribution may depend on the factorization of  $p - 1$ .

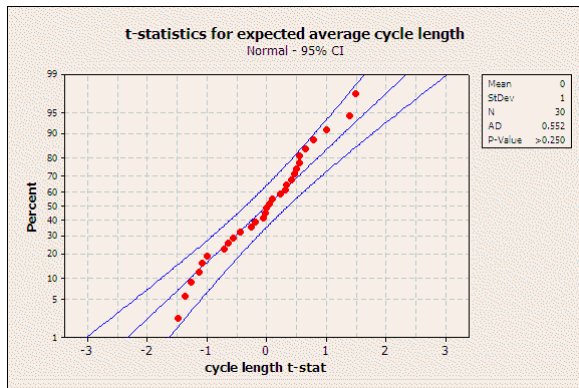
This has not been detected in the data so far, however.

So far we have predictions for many, but not all, of the measurements we have thought about.

	<b>Permutations</b> (Literature)	<b>BFG's</b> (new)	<b>TFG's</b> (new)
Number of components	Mean, Var	Mean, Var	Mean
Number of cyclic nodes	N/A	Mean, Var	Mean
Average component size	N/A		Mean
Average cycle length	Mean, Var	Mean, Var	Mean
Average tail length	N/A	Mean, Var	Mean
Maximum cycle length	Mean	Mean	
Maximum tail length	N/A	Mean	
Distribution of cycle sizes	Freqs		

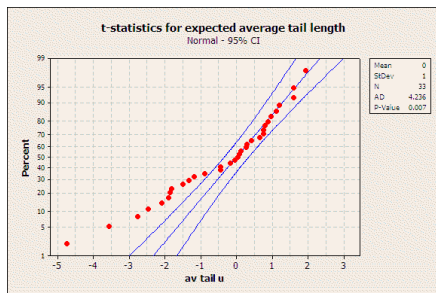
(Predictions in gray have not yet been compared to observations.)

We also have some conclusions.



- ▶ Random  $m$ -ary graphs provide the best model for the prime case so far.

## We also have some conclusions.

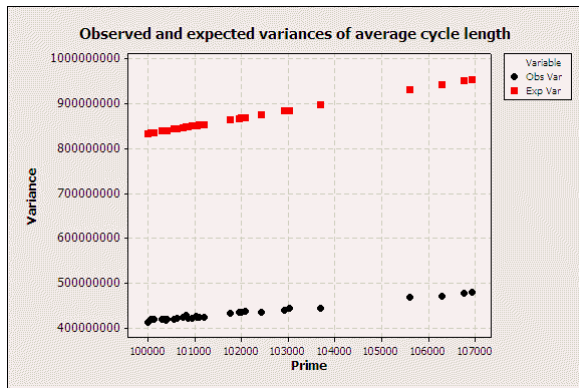


Test of  $\mu = 0$  vs  $\text{not} = 0$

Variable	N	Mean	StDev	SE Mean	T	P
av tail u	33	-0.414	1.585	0.276	-1.50	0.143

- ▶ The mean values of the collected measurements agree with predictions for randomly chosen  $m$ -ary graphs, although there is sometimes an extra source of variation in the means (*between* different primes).

We also have some conclusions.



- ▶ The variances (*within* each prime) of the collected measurements in many cases do **not** agree with the predictions. In general they are smaller, suggesting that our samples do not behave independently.

We also have some conclusions.

Cycle Length	Test Statistic	Reject Uniformity?
1	31.74	Yes
2	3.28	Yes
3	1.35	No
5	1.10	No
7	0.93	No
10	0.57	No
20	1.08	No

- ▶ The distribution of cycles of various (small) sizes does not always agree with predictions, perhaps due to the structure of the factorization of  $p - 1$ .

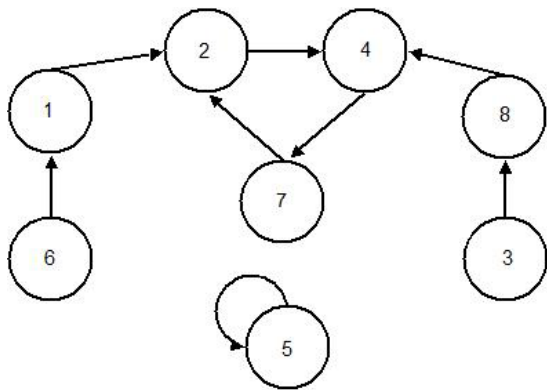
The original problem can be also be asked about other variations of the discrete logarithm problem.

For example:

- ▶ Composite moduli
- ▶ Finite fields
- ▶ Elliptic curves

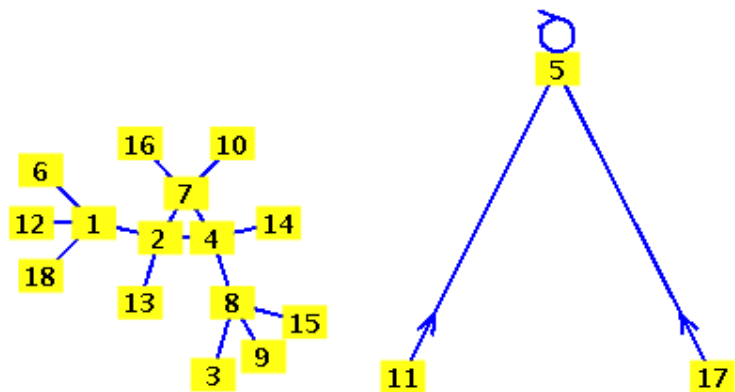
We have started investigation of the prime power case (Marc Mace, 2009) and the elliptic curve case (Aaron Blumenfeld, 2010).

The most obvious prime-power version of our map does not in general produce an  $m$ -ary graph.



$$x \mapsto 2^x \pmod{9}$$
$$\{1, \dots, 8\} \rightarrow \{1, \dots, 8\}$$

However, it does produce such a graph if taken on  $\{1, \dots, p^r(p-1)\} \rightarrow \{1, \dots, p^r(p-1)\}$ .



$$x \mapsto 2^x \pmod{9}$$

$$\{1, \dots, 18\} \rightarrow \{1, \dots, 18\}$$

And once again we can predict what the graph looks like based on the base and the modulus.

### Proposition

*Let  $n = p^r$ . If  $m \mid \phi(n)$  then there are  $\phi\left(\frac{\phi(n)}{m}\right)$   $pm$ -ary functional graphs produced by varying  $g$  for a given  $n$ .*

*Furthermore, the values of  $g$  that produce a  $pm$ -ary graph are precisely those for which  $g$  is a strictly perfect  $m$ th power modulo  $n$ .*

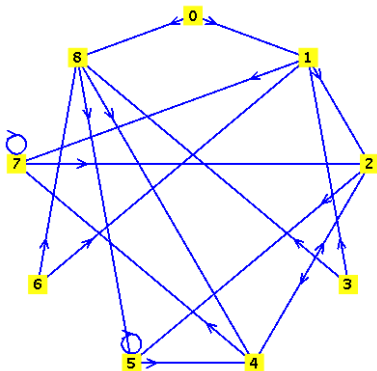
Data was collected on four measurements for modulus  $3^r$ ,  $r = 6, 7, 8, 9, 10$  by Mace (2009).

- ▶ number of components
- ▶ average number of cyclic nodes
- ▶ average cycle length
- ▶ and average tail length

The values of  $g$  which produced ternary functional graphs (primitive roots) were separated out.

The observed values clearly do **not** agree with what would be expected from randomly chosen ternary graphs. There is clearly some other structure here which needs to be taken into account.

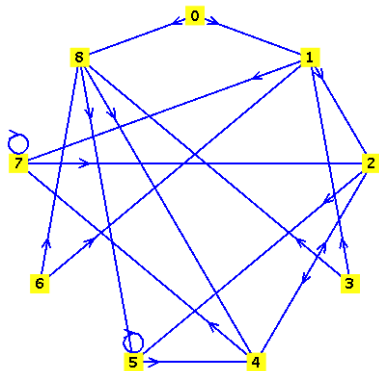
Another possibility is to construct the graph of a “multivalued function”.



$$x \bmod 9 \mapsto 2^x \bmod 9$$

Consider  $x \in \{1, \dots, p^r(p-1)\}$  but draw an edge from  $x \bmod p^r$  to  $g^x \bmod p^r$  for every such  $x$ .

Another possibility is to construct the graph of a “multivalued function”.

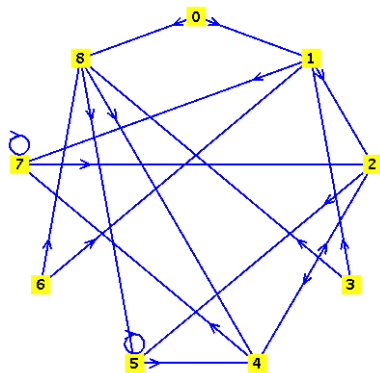


$$x \bmod 9 \mapsto 2^x \bmod 9$$

This results in a “multivalued function”

$$\{0, \dots, p^r - 1\} \Rightarrow \{0, \dots, p^r - 1\}.$$

Another possibility is to construct the graph of a “multivalued function”.



$$x \bmod 9 \mapsto 2^x \bmod 9$$

The resulting graph will not be functional, but every node will have the same out-degree (possibly with multiplicity) and will be  $m$ -ary (possibly with multiplicity) in regard to in-degree.

These graphs seem to have some very well-determined structure.

For example:

**Theorem (Holden and Robinson, 2010)**

*For any  $1 \leq g \leq p^r$ ,  $p \nmid g$ , the number of 1-cycles (fixed points) in the multi-map  $x \bmod p^r \mapsto g^x \bmod p^r$  as  $x$  ranges from 1 to  $p^r(p-1)$  is exactly  $p-1$  (counting multiplicity).*



## Data on elliptic curve DEFG's was collected by Blumenfeld (2010).

- ▶ comprehensively for three values of  $N$
- ▶ using random sampling for three more ( $\approx 1,150,000$  graphs total)

$E$ ,  $p$ , and  $B$  were chosen so that the functional graphs were binary.

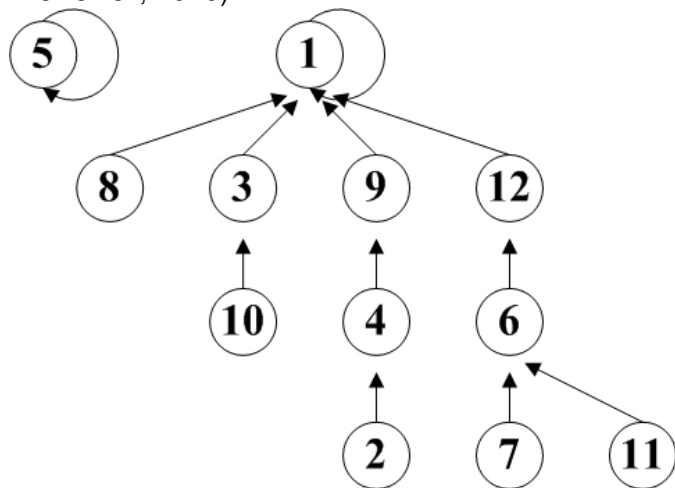
Results so far look similar to predictions in many, but not all, cases. Statistical tests have not yet been done.

## A similar investigation can be done of other maps:

- ▶ Square Discrete Exponentiation:  $x \mapsto g^{x^2} \bmod p$
- ▶ Discrete Lambert:  $x \mapsto xg^x \bmod p$  (related to ElGamal DS)
- ▶ Self-Power:  $x \mapsto x^x \bmod p$  (related to a variant of ElGamal DS)

# We have started investigation of the self-power map.

(Matthew Friedrichsen, Brian Larson, and Emily McDowell, 2010)

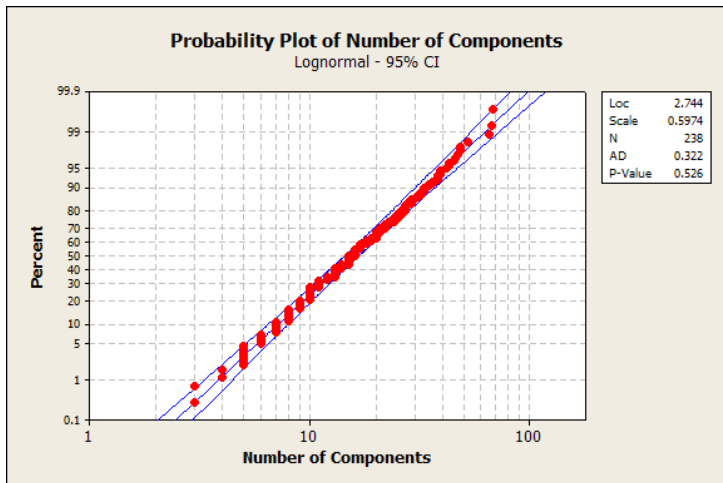


$$x \mapsto x^x \pmod{13}$$

Data was collected for 1090 primes by Friedrichsen, Larson, and McDowell (2010).

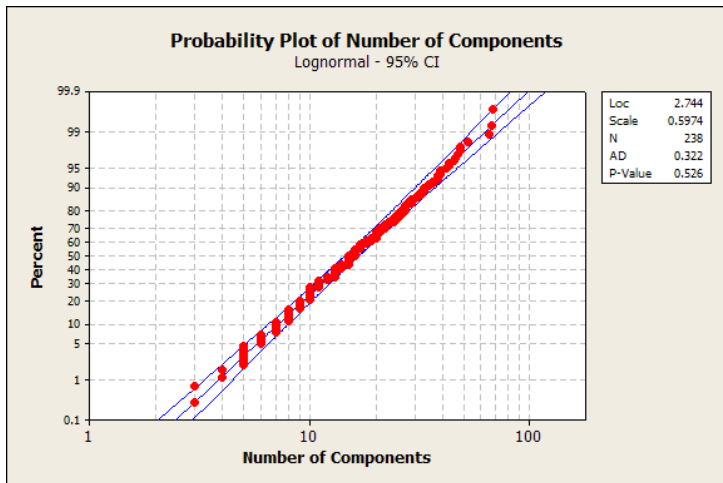
Primes were grouped into two size ranges and also by “safe primes” or not. Statistical tests were done on each category.

Data was collected for 1090 primes by Friedrichsen, Larson, and McDowell (2010).



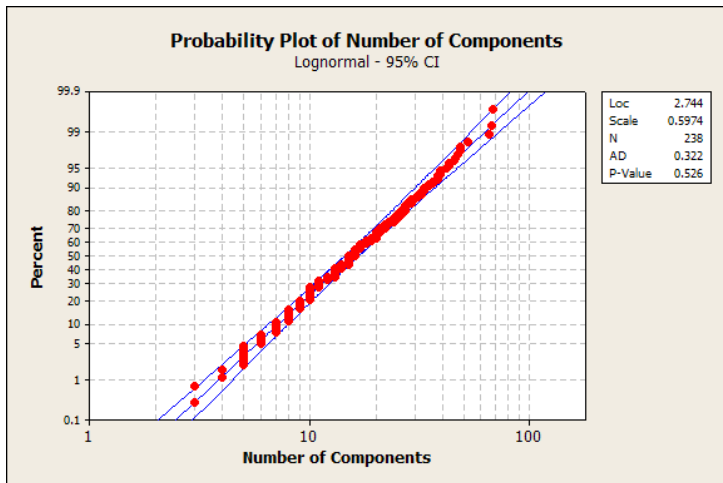
First of several surprises: measurements were not normally distributed — maybe lognormally?

Data was collected for 1090 primes by Friedrichsen, Larson, and McDowell (2010).



Much structure can be demonstrated theoretically; how is it influencing statistics?

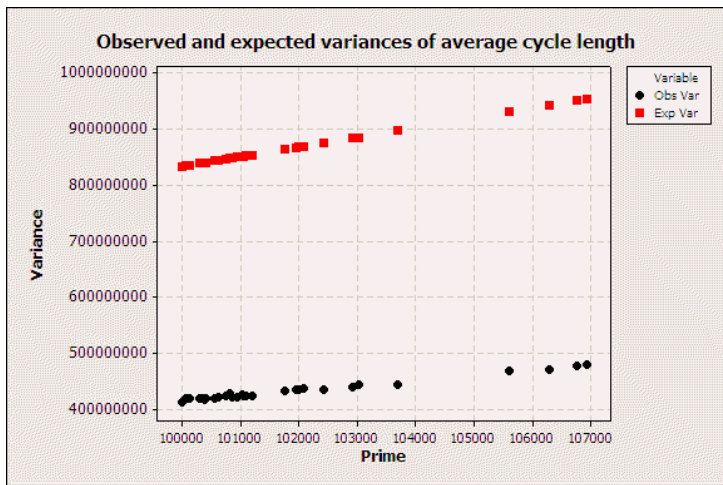
# Number of Components



## There's a lot of possibilities for future work.

- ▶ Missing measurements, especially variances
- ▶ More distributions in more cases
- ▶ Statistical tests for the ternary case and the elliptic curve case
- ▶ The quaternary case (and beyond)
- ▶ Data collection for new measurements, e.g. average component size as seen from a node, maximum component size
- ▶ More on composite moduli, e.g. prime powers, RSA numbers
- ▶ More on self-power graphs
- ▶ Graphs of “multivalued functions”
- ▶ Discrete Lambert graphs
- ▶ Finite fields, other groups

## And two **Big Questions** we'd like to answer!



- ▶ Why are the variances sometimes smaller than expected?
- ▶ Can we exploit this to attack the Discrete Log Problem?

# Thanks!

More information at:

<http://www.rose-hulman.edu/~holden/REU>

