

# Mapping the Discrete Logarithm 3: Revenge of the Stats

Joshua Holden

Joint work with Daniel Cloutier, Nathan Lindle (Senior Theses),  
Max Brugger, Christina Frederick, Andrew Hoffman, and  
Marcus Mace (RHIT REU)

Rose-Hulman Institute of Technology

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

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



# The Question

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

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

We are investigating the functional graph induced by the discrete exponentiation map

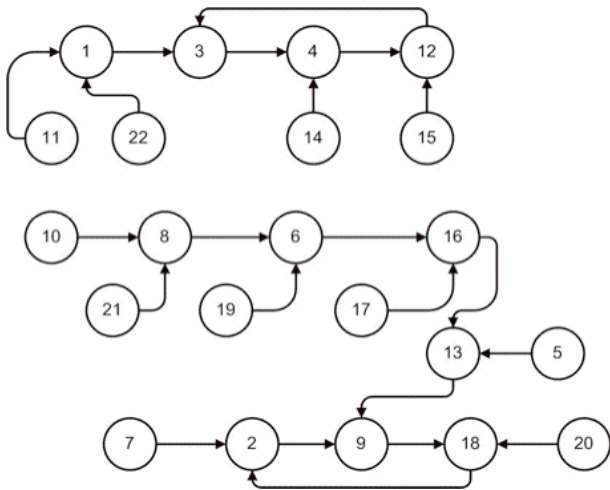
$$x \mapsto g^x \pmod{p}.$$

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

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



# Example: $x \mapsto 3^x \pmod{23}$



# Strategy

- 1 Figure out what a “random graph” looks like.  
(Combinatorics — exponential generating functions,  
Analysis — singularity analysis)
- 2 Figure out what a discrete exponentiation functional graph  
(DEFG) looks like.  
(Number Theory —  $m$ -ary-ness, Computation)
- 3 Compare.  
(Statistics)



# *m*-ary Functional Graphs

Let  $\phi(k) = \#\{1 \leq c \leq k \mid \gcd(c, k) = 1\}$  be the Euler phi function.

## Proposition

*If  $m \mid (p - 1)$  then there are  $\phi(\frac{p-1}{m})$  *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 with  $p - 1$  nodes is exactly  $\frac{p-1}{m}$ .*



# Making Predictions

Permutations:

- Results are in the literature.

Binary Functional Graphs:

(Dan Cloutier, 2006, Nathan Lindle, 2008, Holden, 2009)

- Use generating functions, and either:
- Singularity analysis, or
- Explicitly solve recurrence relations.

Ternary Functional Graphs:

(Max Brugger and Christina Frederick, 2007, Brugger, 2008)

- Use generating functions.
- Obtain recurrence relations.
- Compute desired values.



## Next Question

How do the actual discrete exponentiation functional graphs compare with this model?

— We will look at the collected data and some statistical analysis.



# Predictions for Some Measurements of Interest

	<b>Permutations</b>	<b>BFG's</b>	<b>TFG's</b>
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	



# Permutation Data

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

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

For each prime, we collected the data for all relevant 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.

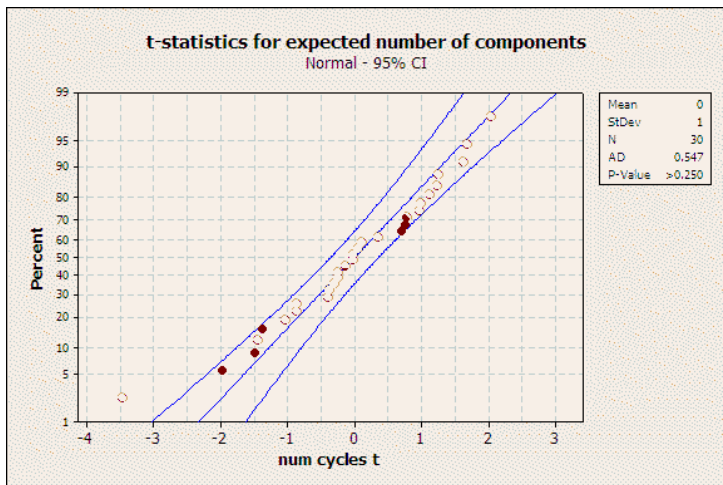
These tests (indirectly) measure the probability that a randomly chosen set of graphs would behave like our graphs.

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

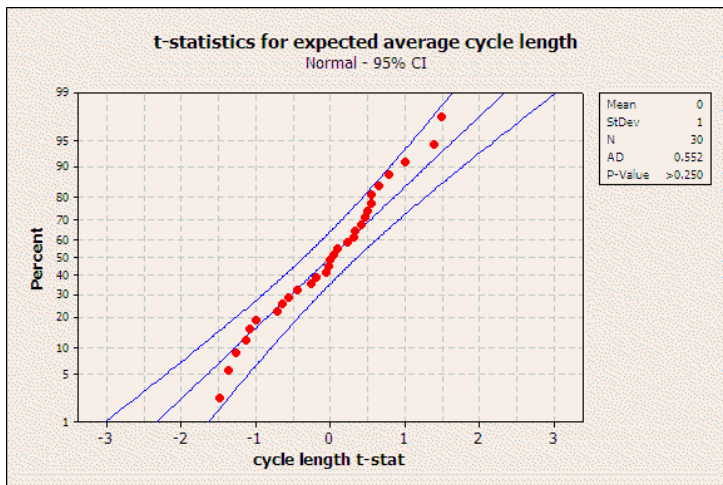
**Note:** In all situations low  $P$ -values are anomalous.



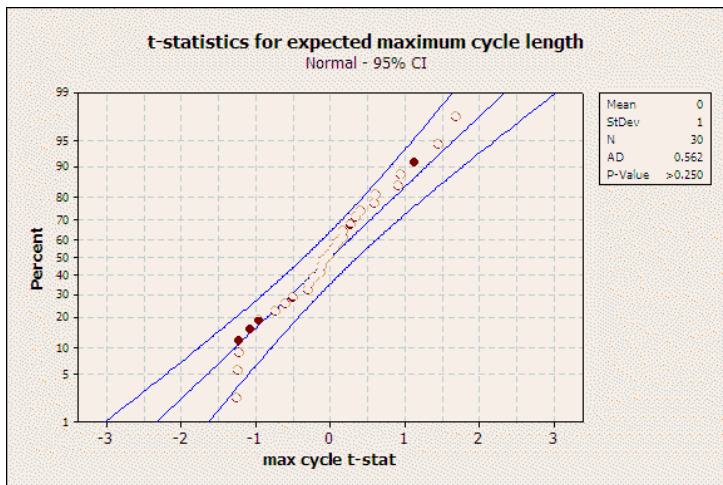
# Number of components



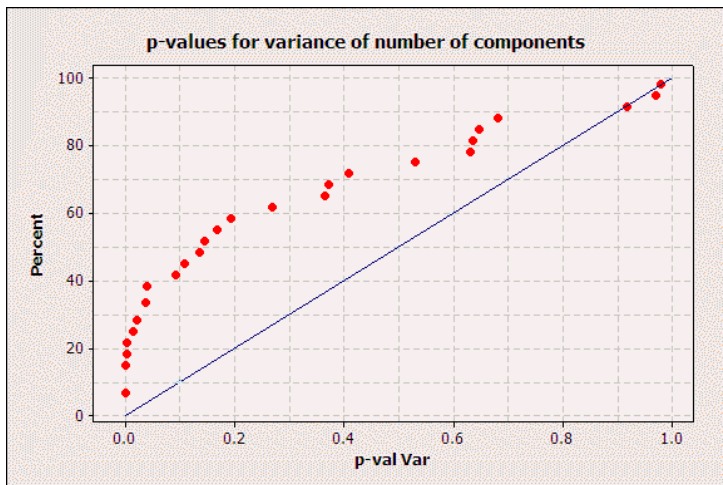
# Average cycle length



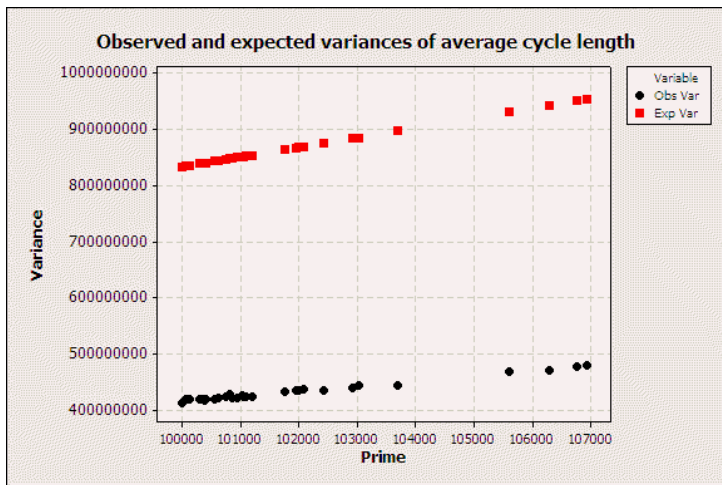
# Maximum cycle length



# Variance of number of components



# Variance of average cycle length



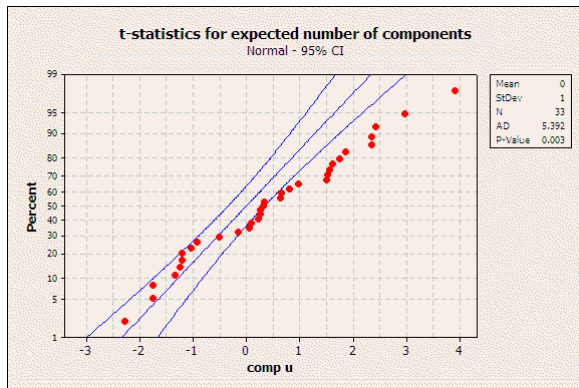
# Binary Functional Graph Data

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



# Number of components

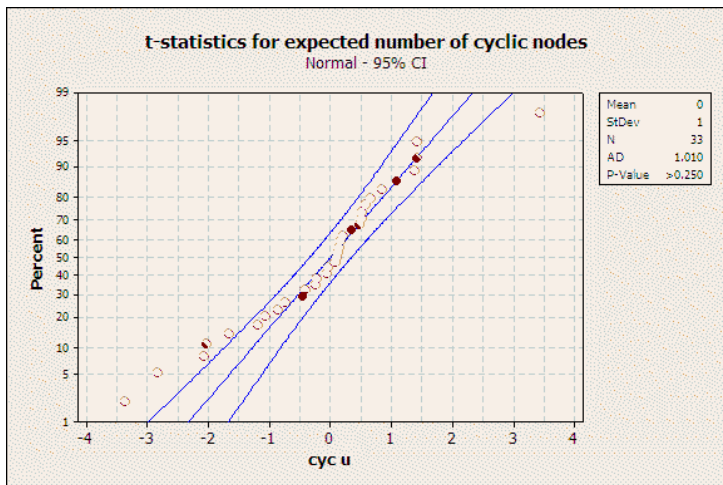


Test of  $\mu = 0$  vs  $\text{not} = 0$

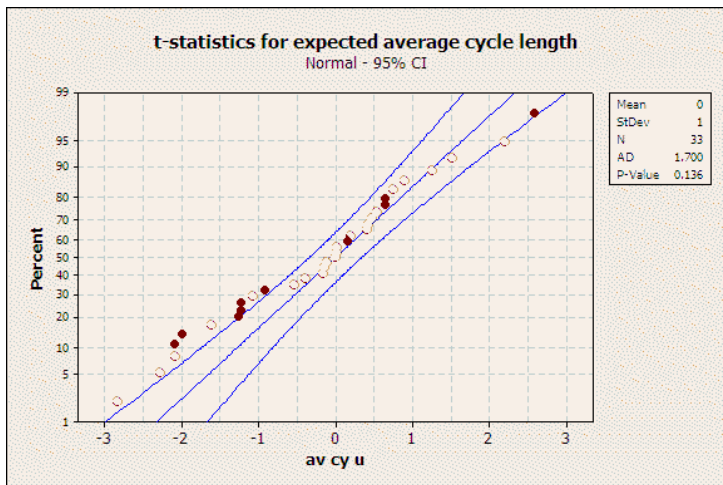
Variable	N	Mean	StDev	SE Mean	T	P
comp u	33	0.457	1.508	0.263	1.74	0.091



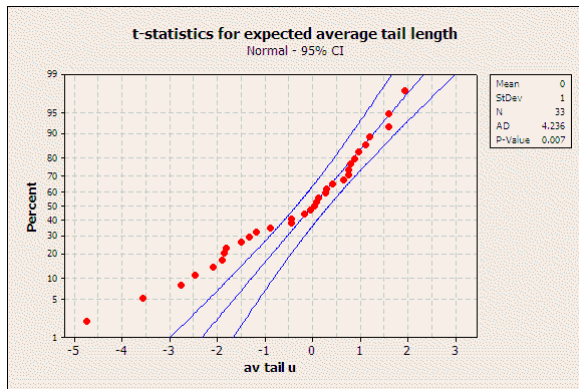
# Number of cyclic nodes



# Average cycle length



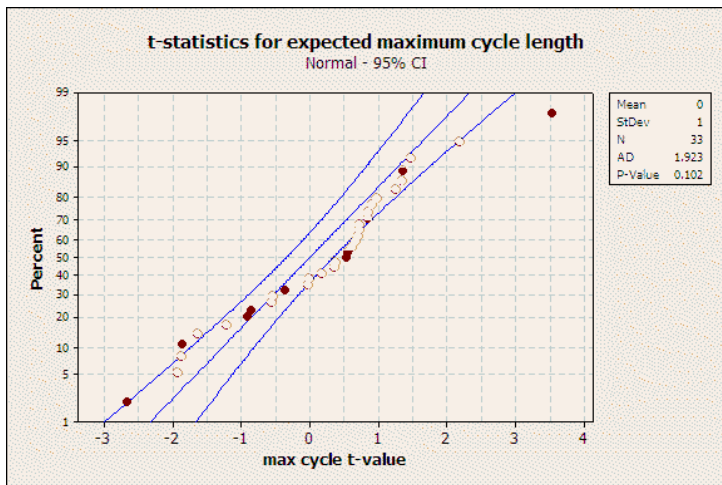
# Average tail length



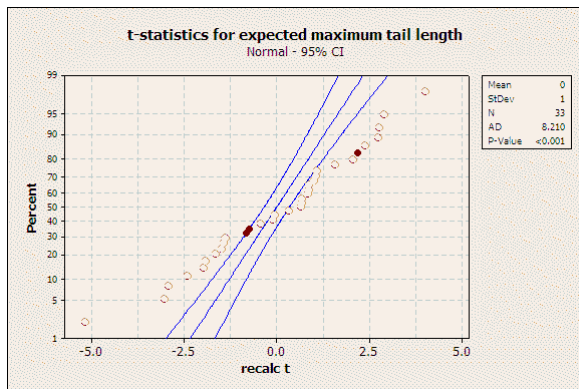
Test of  $\mu = 0$  vs  $\text{not} = 0$

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

# Maximum cycle length



# Maximum tail length

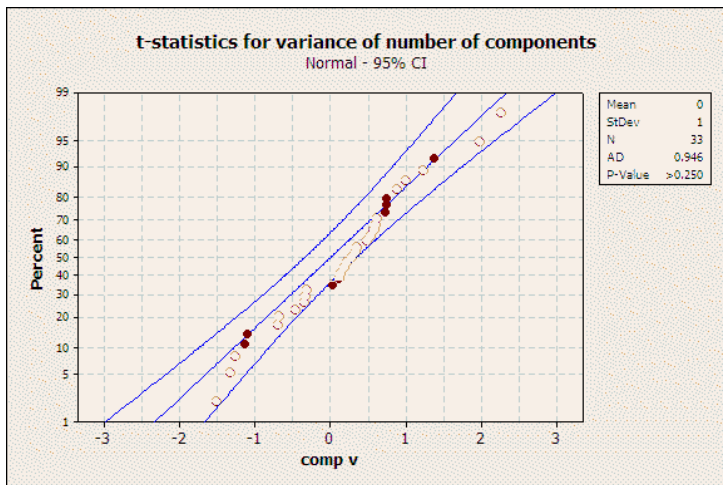


Test of  $\mu = 0$  vs  $\text{not} = 0$

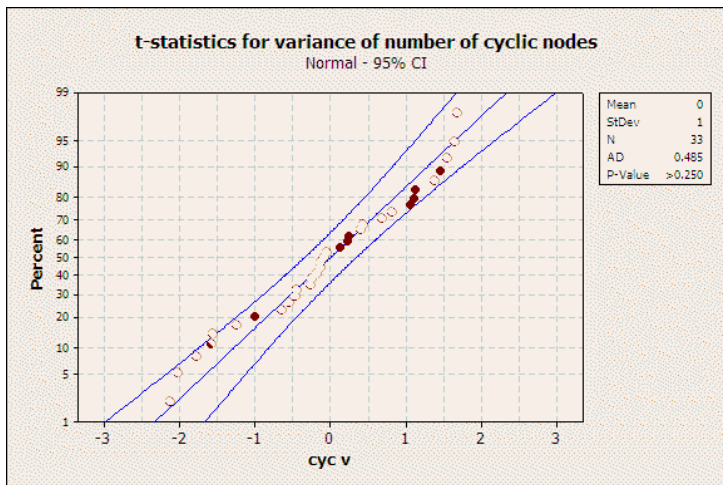
Variable	N	Mean	StDev	SE Mean	T	P
recalc t	33	0.083	2.026	0.353	0.23	0.816



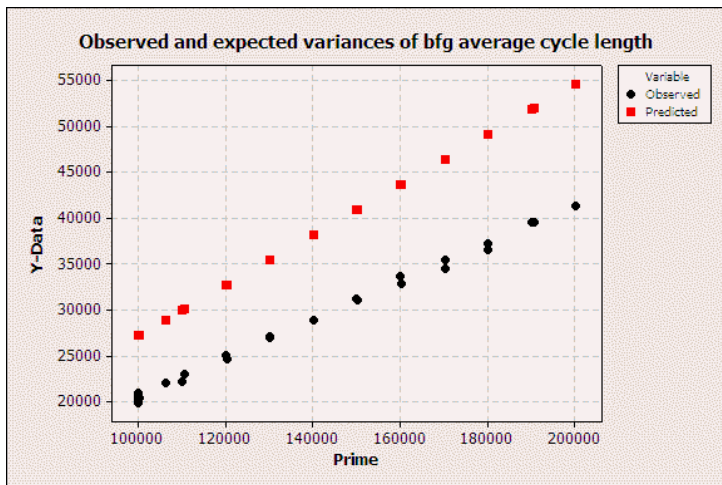
# Variance of number of components



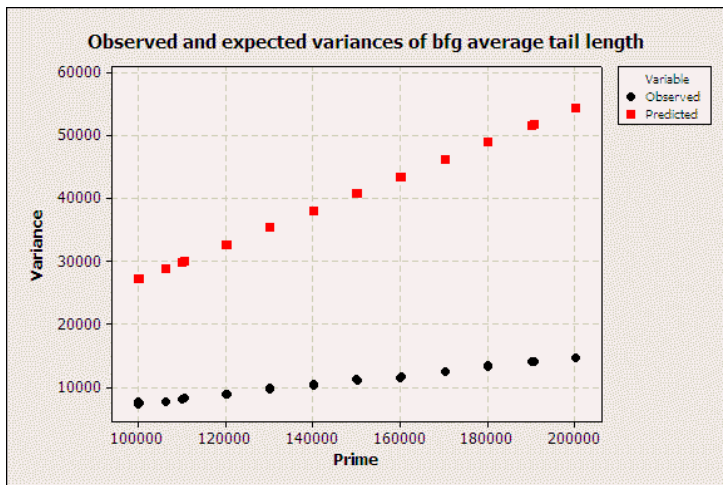
# Variance of number of cyclic nodes



# Variance of average cycle length



# Variance of average tail length



# Ternary Functional Graph Data

Data was collected on number of components, average number of cyclic nodes, and average cycle length for nine primes by Brugger and Frederick (2007).

Again, the values of  $g$  which produced ternary functional graphs were separated out. Statistical tests have not yet been done on this data, but relative error calculations look very promising.



# Permutation Measurements: Distributions

(Standard Result from Literature) 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  $\infty$ .



# Cycle Distribution Data

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

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

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.



# Cycle Distribution Example

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



# Cycle Distribution Results

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



# Cycle Distribution Theorems

This is perhaps not that surprising, given the extra structure associated with 1-cycles (fixed points) and 2-cycles. For example, consider the following theorem:

**Theorem (Holden and Moree, 2004)**

*There are  $\gg x / \ln x$  primes  $p \leq x$  such that the average number of 1-cycles (fixed points) in the map  $x \mapsto g^x \pmod p$  for 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.



# Extensions

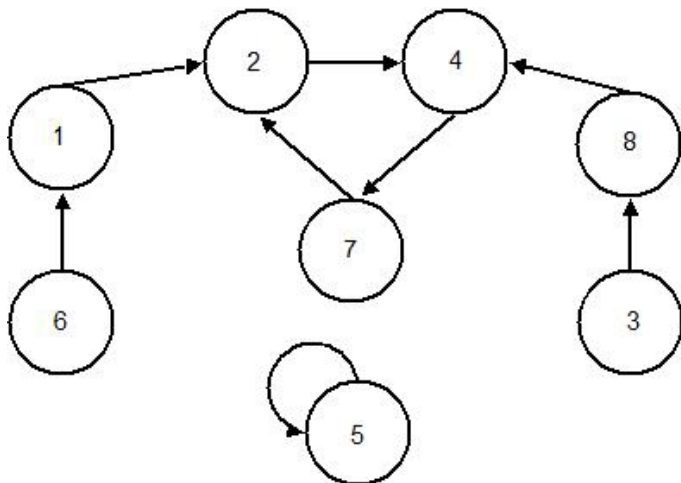
The original problem can be also be asked about other variations of the discrete logarithm problem, e.g.:

- Composite moduli
- Finite fields
- Elliptic curves?

We have started investigation of the prime power case. (Marc Mace, 2009)



# Example: $x \mapsto 2^x \pmod{9}$



# Prime Power Moduli

The map

$$\{1, \dots, p^r - 1\} \rightarrow \{1, \dots, p^r - 1\}$$

given by

$$x \mapsto g^x \pmod{p^r}$$

does not in general produce an  $m$ -ary graph (for any  $m$ ).

However, it does produce such a graph if taken on the sets

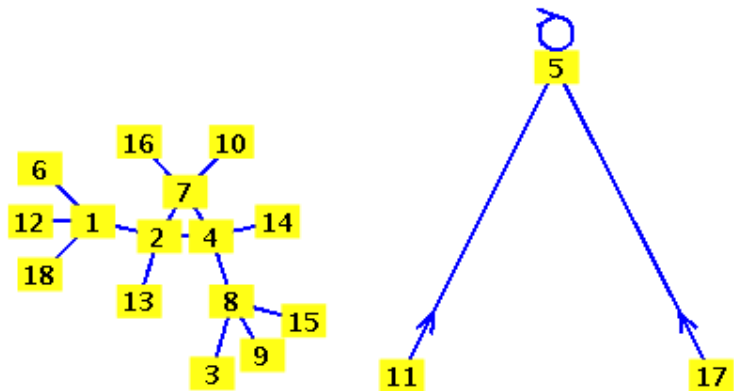
$$\{1, \dots, p^r(p-1)\} \rightarrow \{1, \dots, p^r(p-1)\}.$$

## 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  $n$ .*

# Example: $x \mapsto 2^x \pmod{9}$ (II)



# Prime Power Functional Graph Data

Data was collected on number of components, average number of cyclic nodes, average cycle length, and average tail length for  $3^r$ ,  $r = 6, 7, 8, 9, 10$  by Mace (2009).

The values of  $g$  which produced tfg's (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 Prime Power Graph

Another possibility is to 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  $x$ .

This results in a “multivalued function”

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

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.





## Another 1-cycle Theorem

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).*



# Conclusions

- Random  $m$ -ary graphs provide the best model for the prime case so far.
- 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).
- 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.
- The distribution of cycles of various (small) sizes does not always agree with predictions, perhaps due to the structure of the factorization of  $p - 1$ .



# Future Work

- Missing measurements, especially variances
- Distributions in the binary and ternary cases
- Statistical tests for the ternary case
- The quaternary case (and beyond)
- Data collection for new measurements, e.g. average component size as seen from a node, maximum component size
- Composite moduli, e.g. prime powers, RSA numbers
- Graphs of “multivalued functions”
- Finite fields, elliptic curves, other groups



# And the Big Questions

- Why are the variances (within each prime) smaller than expected?
- Can we exploit this to attack the Discrete Log Problem?



# Thanks!

