Improved upper bounds on key invariants of Erdös-Rényi numerical semigroups

Santiago Morales

University of California, Davis

moralesduarte@ucdavis.edu

October 1, 2024

Overview

The probabilistic method

Erdös-Rényi model

Definition

The Erdös-Rényi model for random graphs $G(n,p)$ is a probability space over the set of graphs on n labeled vertices determined by

$$\Pr[\{i,j\} \in G] = p$$

with these events mutually independent.

Threshold functions

As n grows, we let p be a function of n, $p = p(n)$.

Definition

$r(n)$ is a threshold function for a graph property A if

1. When $p(n) \in o(r(n))$, $\lim_{n\to\infty} \Pr[G(n, p(n)) \vDash A] = 0$,
2. When $r(n) \in o(p(n))$, $\lim_{n\to\infty} \Pr[G(n, p(n)) \vDash A] = 1$,

or vice versa.

Example

$r(n) = \dfrac{\ln n}{n}$ is a threshold for having no isolated vertices.

Second moment method

Let $X \geq 0$ be a random variable with finite variance.

Theorem

$$\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2}.$$

Proof. Apply Chebyschev's inequality with $\lambda = \frac{\mu}{\sigma}$:

$$\Pr[X = 0] \leq \Pr[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2} = \frac{\sigma^2}{\mu^2}. \quad \square$$

Corollary

If $\mathrm{Var}[X] \in o(\mathrm{E}[X]^2)$, $X > 0$ *asymptotically almost always.*

Numerical semigroups

Numerical semigroups

Definition

A *numerical semigroup* is a subset $S \subseteq \mathbb{N}$ for which

1. $0 \in S$,
2. S is closed under addition, i.e. $a, b \in S$ implies $a + b \in S$, and
3. S has finite complement in $\mathbb{N}$.

Example

$\mathbb{N}$ and $\mathbb{N} \setminus \{1\}$. Subsets of $\mathbb{N}$ which are not numerical semigroups include the set of even numbers, any finite set and $\mathbb{N}_0 \setminus \{2\}$.

Example

The *McNugget Semigroup* is the set of all non-negative integers which can be expressed as a sum of non-negative multiples of 6, 9 and 20.

Generating sets

The McNugget semigroup is an example of a numerical semigroup which is *finitely generated*. This means that there exists a finite set $A = \{a_1, \ldots, a_n\}$ such that $S = \langle A \rangle$, where

$$\langle A \rangle = \{c_1 a_1 + \cdots + c_n a_n : c_1, \ldots, c_n \in \mathbb{N}\}.$$

Theorem

All numerical semigroups are finitely generated.

Theorem

Let $A \subseteq \mathbb{N}$ be a non-empty finite set. Then $\langle A \rangle$ is a numerical semigroup if and only if $\gcd(A) = 1$.

The McNugget semigroup

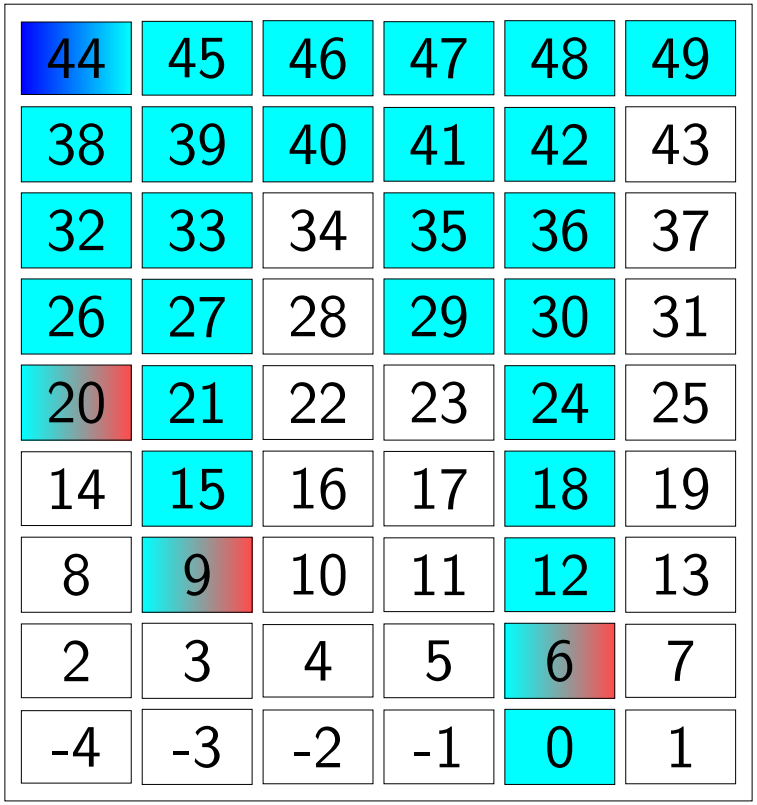

Figure: Visualization of the McNugget semigroup.

Invariants

Definition

The *multiplicity* of S, denoted by $m(S)$, is the smallest non-zero element of S.

Theorem

There exists a unique minimal generating set A with $S = \langle A \rangle$.

Definition

The *embedding dimension* of S, denoted by $e(S)$, is the cardinality of the minimal generating set of S.

Numerical semigroup invariants

Note that

$$e(S) \leq m(S).$$

Definition

The *Apéry set* is defined as

$$\mathrm{Ap}(S) = \{s \in S : s - m(S) \notin S\}.$$

Definition

The *Frobenius number* of S, denoted by $F(S)$, is the largest element of the complement of S in $\mathbb{N}$.

The McNugget semigroup

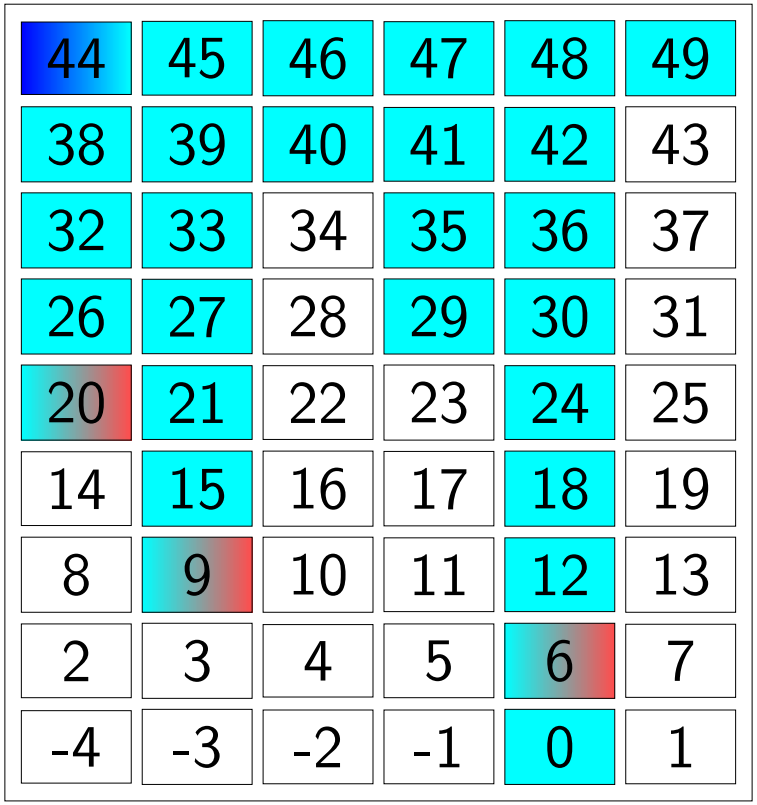

Figure: Visualization of the McNugget semigroup.

Numerical semigroup invariants

Note that

$$F(S) = \max(\mathrm{Ap}(S)) - m(S),$$

for any $n \in S$. Finding the Frobenius number of a numerical semigroup is **NP-hard**.

Definition

The *genus* of S, denoted by $g(S)$, is the cardinality $\mathbb{N} \setminus S$.

Proposition

$$g(S) \leq F(S) \leq 2g(S).$$

Random numerical semigroups

ER-type random numerical semigroups

Definition

For $p \in (0, 1]$ and $M \in \mathbb{N}$, an ER-type random numerical semigroup $\mathcal{S}(M, p)$ is obtained by using the following procedure:

1. Initialize an empty set $\mathcal{A}$.
2. As i goes from 1 to M, add i to $\mathcal{A}$ with probability p, independently of the other steps.
3. Return the semigroup $\mathcal{S} = \langle \mathcal{A} \rangle$.

Note that this definition does not require a numerical semigroup $\mathcal{S}$ to be co-finite.

A theorem of De Loera, O'Neill and Wilburne

Theorem

Let $\mathcal{S} \sim \mathcal{S}(M, p)$, where $p = p(M)$ is a monotone decreasing function of M. Then,

1. *If $p(M) \in o\left(\frac{1}{M}\right)$, then $\mathcal{S} = \{0\}$ almost always.*
2. *If $\frac{1}{M} \in o(p(M))$ and $\lim_{M\to\infty} p(M) = 0$, then $\mathcal{S}$ is co-finite almost always and*
$$\lim_{M\to\infty} \mathrm{E}[e(\mathcal{S})] = \lim_{M\to\infty} \mathrm{E}[g(\mathcal{S})] = \lim_{M\to\infty} \mathrm{E}[F(\mathcal{S})] = \infty.$$
3. *If $\lim_{M\to\infty} p(M) > 0$, then*
$$\lim_{M\to\infty} \mathrm{E}[e(\mathcal{S})] < \infty, \quad \lim_{M\to\infty} \mathrm{E}[g(\mathcal{S})] < \infty \quad \text{and} \quad \lim_{M\to\infty} \mathrm{E}[F(\mathcal{S})] < \infty,$$
and each limit is bounded by explicit rational functions in p.

Upper bounds

They also provide explicit upper bounds when p is constant.

Theorem

Let $\mathcal{S} \sim \mathcal{S}(M, p)$, where p is a constant. Then,

$$\Theta(1) \leq \lim_{M\to\infty} \mathrm{E}[e(\mathcal{S})] \leq \Theta\left(\frac{1}{p}\right),$$

$$\Theta\left(\frac{1}{p}\right) \leq \lim_{M\to\infty} \mathrm{E}[g(\mathcal{S})] \leq \Theta\left(\frac{1}{p^2}\right),$$

$$\Theta\left(\frac{1}{p}\right) \leq \lim_{M\to\infty} \mathrm{E}[F(\mathcal{S})] \leq \Theta\left(\frac{1}{p^2}\right).$$

Experiments

Average embedding dimension

$1/p$	Lower Bound	$\overline{e(\mathcal{S})}$	Upper bound
4.00	2.21	2.79	7.75
75.14	2.95	9.05	150.27
146.29	2.97	10.67	292.56
217.43	2.98	12.28	434.85
288.57	2.99	12.94	577.14
359.71	2.99	13.65	719.43
430.86	2.99	14.14	861.71
502.00	2.99	14.73	1,004.00
573.14	2.99	15.29	1,146.28
644.29	2.99	16.17	1,288.57
715.43	2.99	16.07	1,430.86
786.57	2.99	16.34	1,573.14
857.71	3.00	16.61	1,715.43
928.86	3.00	17.39	1,857.71
1,000.00	3.00	17.22	2,000.00

Table: Average embedding dimension of random numerical semigroups generated using the ER-type model (15 samples of 1000 random numerical semigroups).

Average embedding dimension plot

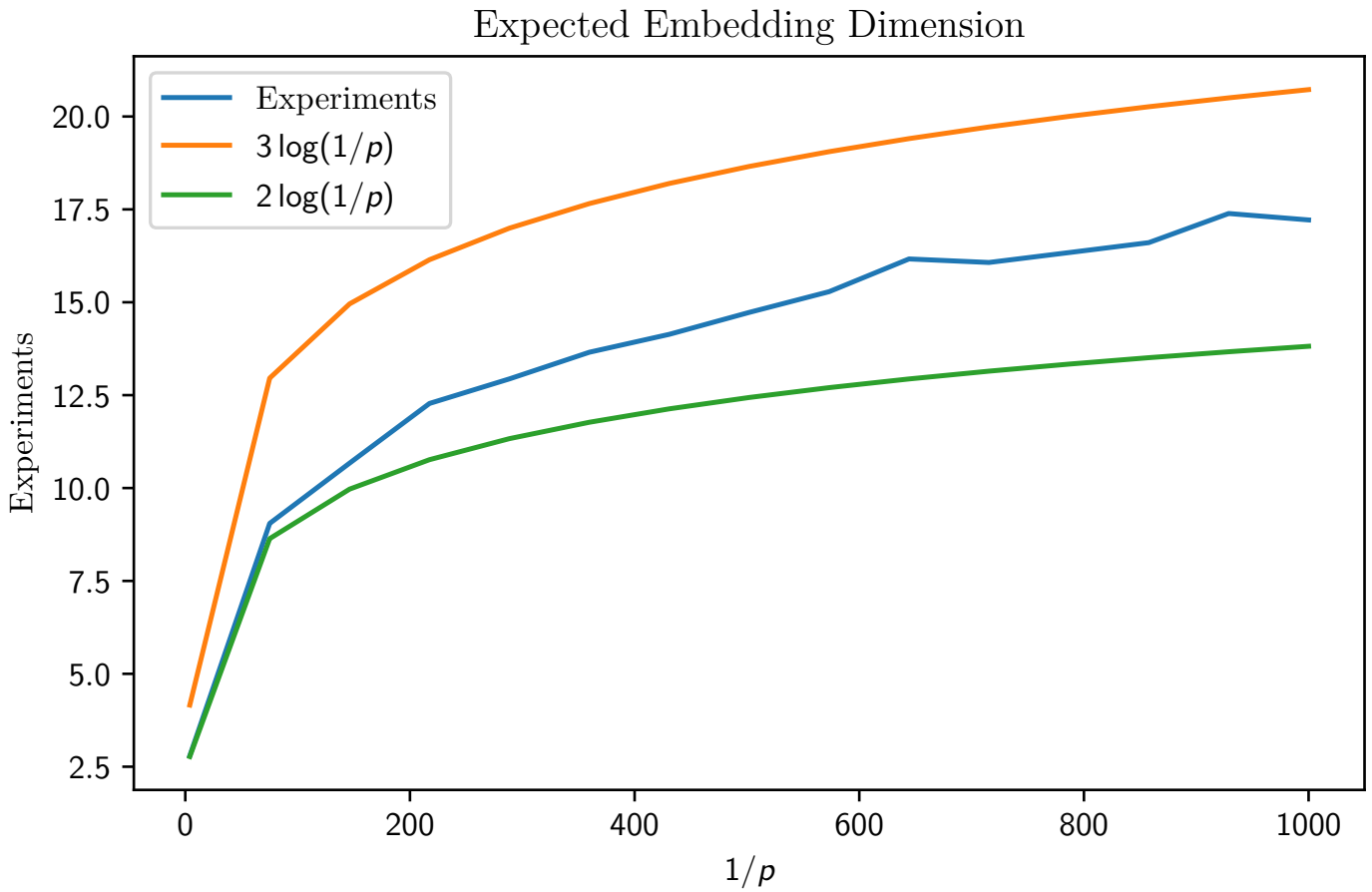

Figure: Average embedding dimension of random numerical semigroups generated using the ER-type model vs $\log(1/p)$.

Average Frobenius number

$1/p$	Lower Bound	$\overline{F(\mathcal{S})}$	Upper bound
4.00	7.26	13.96	46.50
75.14	218.56	1,088.82	22,283.25
146.29	431.93	2,483.26	85,010.91
217.43	645.33	4,174.94	188,229.03
288.57	858.75	5,859.29	331,937.60
359.71	1,072.17	7,794.18	516,136.62
430.86	1,285.59	9,594.56	740,826.09
502.00	1,499.02	11,533.38	1,006,006.00
573.14	1,712.45	13,765.73	1,311,676.37
644.29	1,925.87	16,239.19	1,657,837.19
715.43	2,139.30	17,769.34	2,044,488.45
786.57	2,352.73	19,806.19	2,471,630.17
857.71	2,566.15	22,157.78	2,939,262.33
928.86	2,779.58	25,079.10	3,447,384.94
1,000.00	2,993.01	26,637.46	3,995,998.00

Table: Average Frobenius number of random numerical semigroups generated using the ER-type model (15 samples of 1000 random numerical semigroups).

Average Frobenius number plot

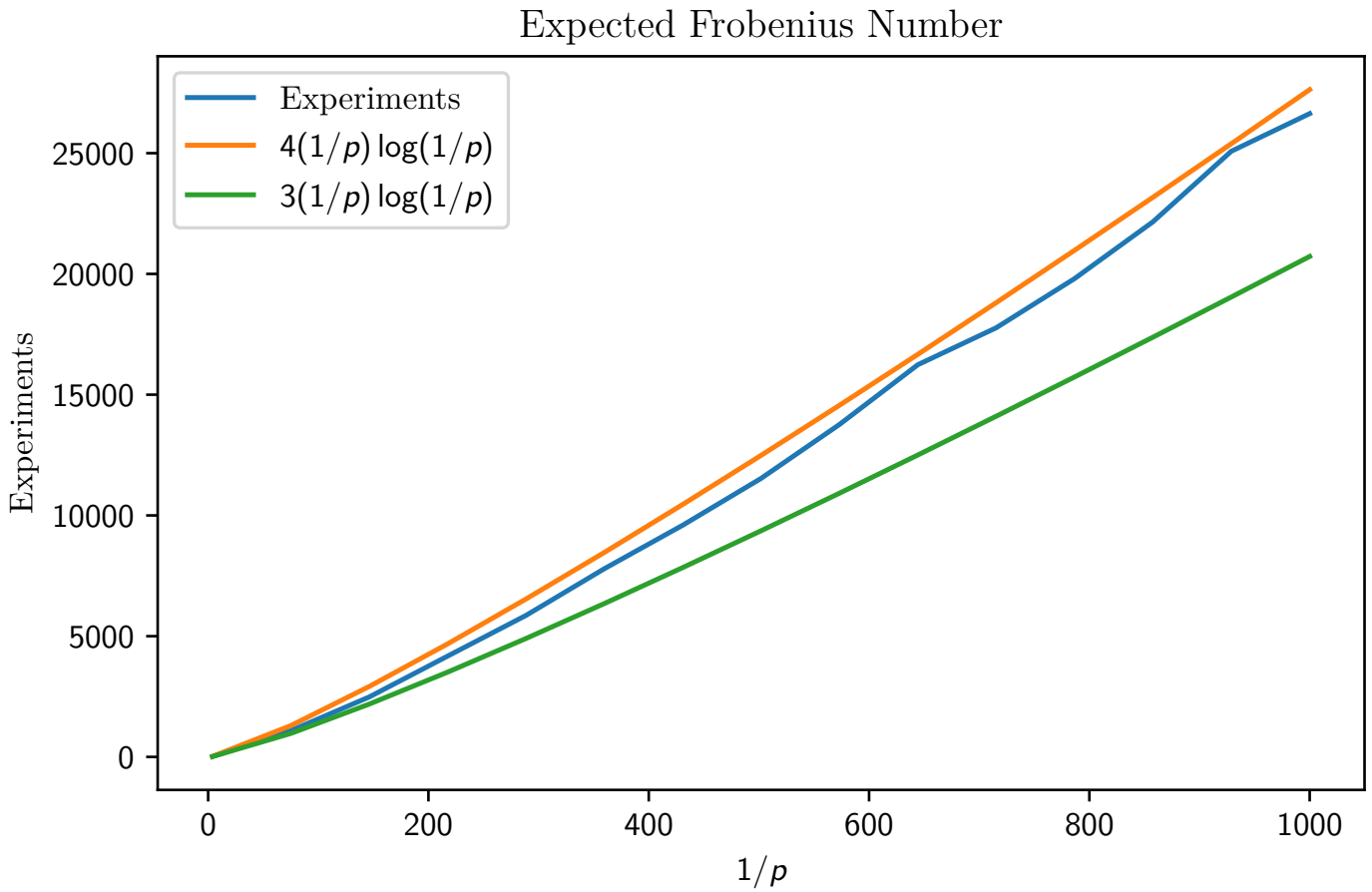

Figure: Average Frobenius number of random numerical semigroups generated using the ER-type model vs $(1/p)\log(1/p)$.

A conjecture

Conjecture

- $\lim_{M\to\infty} \mathbb{E}[e(\mathcal{S}(M,p))] \in \Theta\left(\log \frac{1}{p}\right)$ and
- $\lim_{M\to\infty} \mathbb{E}[F(\mathcal{S}(M,p))] \in \Theta\left(\frac{1}{p}\log \frac{1}{p}\right)$.

Results

A new definition

Definition

For $p \in (0,1]$, an *unconstrained ER-type random numerical semigroup* $S(p)$ is a probability sapce over the set of semigroups $S = \langle \mathcal{A} \rangle$ with $\mathcal{A} \subseteq \mathbb{N}$ determined by $\Pr[n \in \mathcal{A}] = p$ for each $n \in \mathbb{N}$ with these events mutually independent.

Main result

Theorem

Let $S \sim S(p)$ for any $p < 24e^{-8}$. Then,

$$\mathrm{E}[e(S)] \leq 5000 \left(\ln\left(\frac{24}{p}\right)\right)^3 \text{ and}$$

$$\mathrm{E}[g(S)] \leq \mathrm{E}[f(S)] \leq \frac{5000}{p}\left(\ln\left(\frac{24}{p}\right)\right)^3.$$

For both Frobenius number and genus, we reduce the ratio between the lower and upper bounds from order $1/p$ to order $(\ln(1/p))^3$ at the cost of large constant factors and an assumption that p is sufficiently small.

Proof strategy

Our proofs are purely probablilistic. We separate two processes:

1. Random selection of generators.
2. Creation of new elements via addition.

First, we show that with high probability, both a prime q and a set $\mathcal{A}$ of roughly $\log q$ elements are chosen. The set $\mathcal{A}$ can be interpreted as a random subset of $\mathbb{Z}_q$, and we prove a result about $k-$fold sums of such sets. We use that to show that the Frobenius number is bounded by a function of q and $\log q$.

Sumsets

Definition

Let $\mathcal{A}$ be a subset of an abelian group G. The $k-$fold sumset of $\mathcal{A}$ is defined as

$$k\mathcal{A} = \{a_1 + \cdots + a_k : a_1, \ldots, a_k \in \mathcal{A}\}.$$

Sumsets are a central object of study in additive number theory. However, there are fewer results about sumsets of random sets. We prove the following theorem which shows that if a random subset of $\mathcal{A} \subset \mathbb{Z}_q$ whose size as well as k are a multiple of $\log q$, then with high probability, $k\mathcal{A} = \mathbb{Z}_q$.

Random sumset theorem

Theorem

Let q be a prime number and $\mathcal{A}$ be a random subset of $\mathbb{Z}_q$ of size $2b\log_2 q$. Then

$$\Pr\left[(b\log_2 q)\mathcal{A} \neq \mathbb{Z}_q\right] \leq \frac{2b\log_2 q + 2}{q^{b-2}}.$$

Example

Let $q = 17$.

- $\mathcal{A} = \{0, 3, 5, 11\}$,
- $2\mathcal{A} = \{0, 3, 5, 6, 8, 10, 11, 14, 16\}$,
- $4\mathcal{A} = \mathbb{Z}_q$.

Proof

1. We use the second moment method to show that with high probability, the $k-$fold sumset of $\mathcal{A}$ covers $\mathbb{Z}_q$.
2. Our proof uses combinatorial bounds on binomial coefficients in order that the inequalities hold for all q. Assymptoically, as q tends to infinity one could instead use Stirling's formula to obtain the improved bound

$$\Pr\left[(b\log_2 q)\mathcal{A} \neq \mathbb{Z}_q\right] \in O\left(\frac{\log q}{q^{b-2}}\right).$$

3. Our proof actually bounds the probability that not every element of $\mathbb{Z}_q$ is a sum of $b\log_2 q$ *distinct* elements of $\mathcal{A}$, so it yields a stronger statement.

Upper bound proof strategy

The strategy is to prove that the Ápery set of a subsemigroup of S is completed before a certain step with high probability, since $F(\mathcal{S})$ is less than the maximum element of this Ápery set. The proof has the following structure:

Let

$$f(p) = \frac{d}{p}\left(\ln\left(\frac{d}{p}\right)\right)^2.$$

Upper bound proof strategy

1. Define D_1 to be the event that at least one prime number between $f(p)+1$ and $cf(p)$ is selected as a generator of $\mathcal{S}$
2. Assuming D_1, let q be the largest prime selected in that range. Define D_2 to be the event that at least $2b\log_2 q$ are selected between 1 and $q-1$.
3. Assuming D_1 and D_2, let $\mathcal{S}'$ be the subsemigroup of $\mathcal{S}$ generated by a random subset of $2b\log_2 q$ elements between 1 and q along with q itself. Let D_3 be the event that the largest element of the Ápery set of $\mathcal{S}'$ with respect to q is at most $bq\log_2 q$.

If all of these events occur, then we can bound the frobenius number by

$$F[\mathcal{S}] \leq F[\mathcal{S}'] \leq bq\log_2 q \leq bcf(p)\log_2(cf(p)).$$

Some visualizations

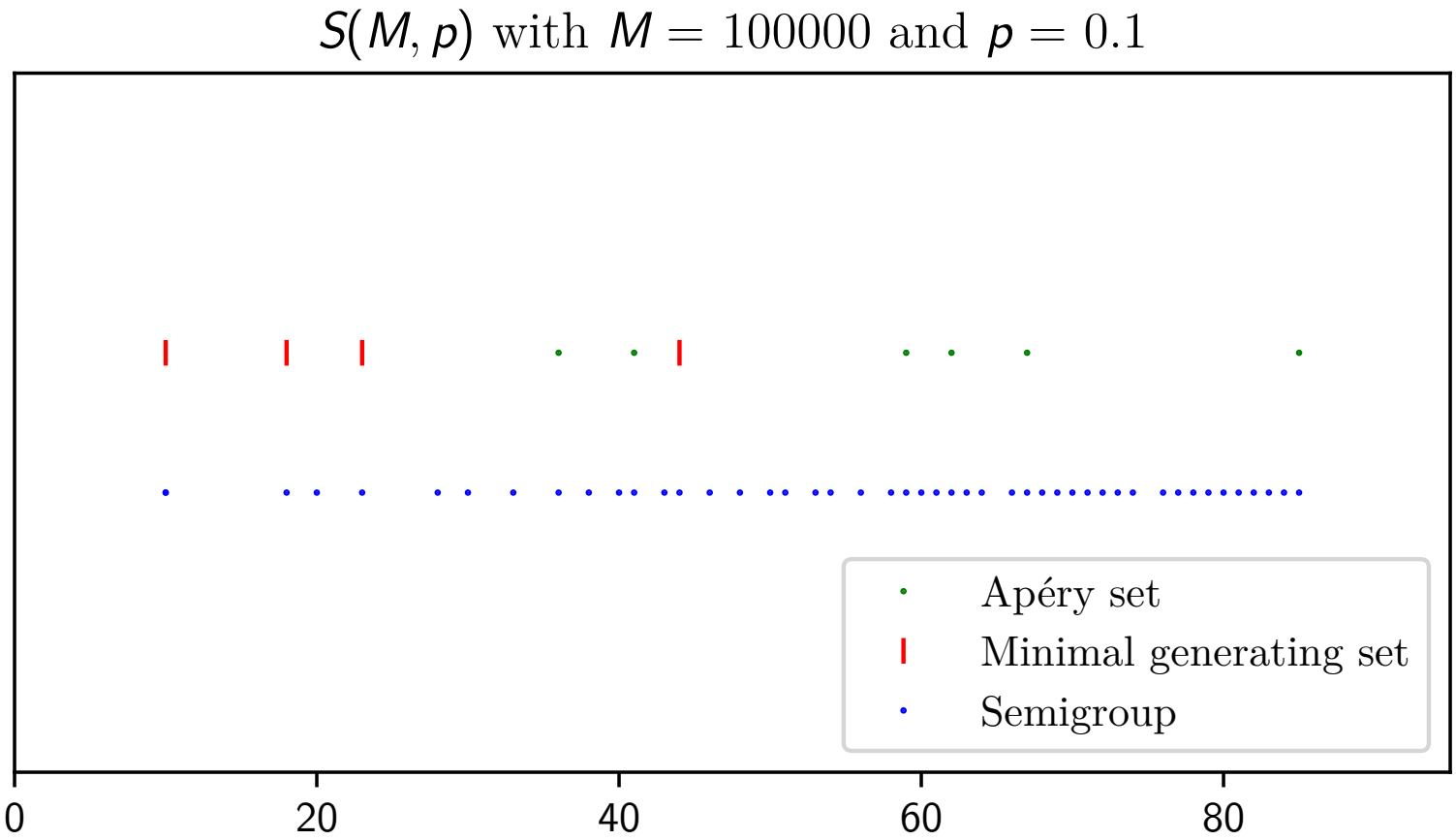

Figure: ER-type random numerical semigroup, $1/p = 10$.

Some visualizations

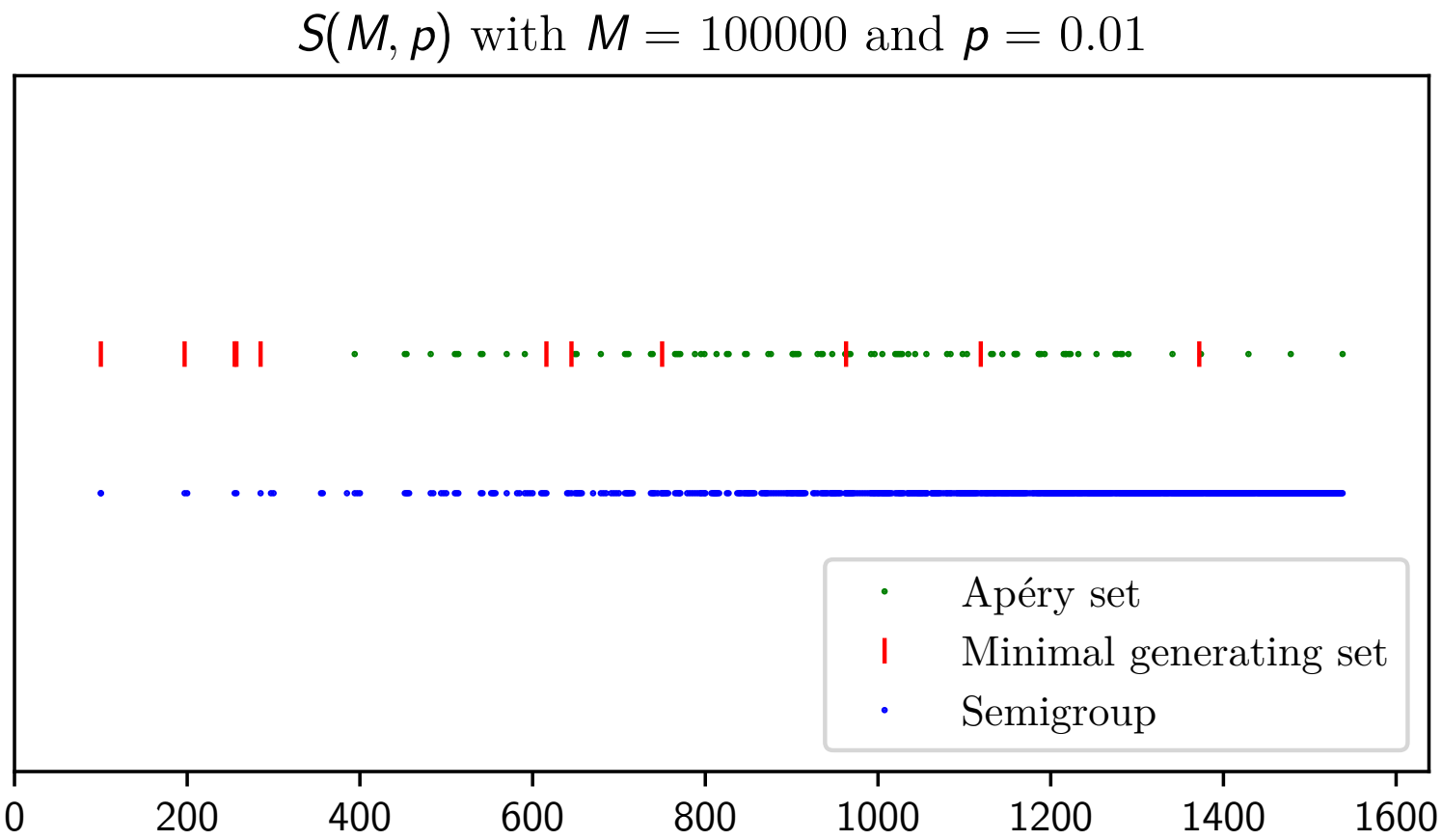

Figure: ER-type random numerical semigroup, $1/p = 100$.

Summary

We use a range of probabilistic methods to improve the upper bounds to within a polylogarithmic factor of the lower bounds. As one of the tools to do this, we prove that for any prime q, if $\mathcal{A}$ is a random subset of $\mathbb{Z}_q$, whose size is of the order of $\log q$ and k is also of order $\log q$, then with high probability, $k\mathcal{A} = \mathbb{Z}_q$.
One strategy to prove the conjecture would be to try to bound the Ápery set with respect to the smallest generator of $\mathcal{S}$. This might be feasible via some generalization of the random sumset theorem in which q is not required to be prime. The proof would need to be modified since it relies on the symmetry of prime cyclic groups.

Selected References

- J. De Loera, C. O'Neill, and D. Wilburne, "Random numerical semigroups and a simplicial complex of irreducible semigroups," *The Electronic Journal of Combinatorics*, P4–37, 2018
- J. L. Ramírez-Alfonsín, "Complexity of the Frobenius problem," *Combinatorica*, vol. 16, pp. 143–147, 1996
- V. I. Arnold, "Weak asymptotics for the numbers of solutions of Diophantine problems," *Functional Analysis and Its Applications*, vol. 33, no. 4, pp. 292–293, 1999
- N. Alon and J. H. Spencer, *The Probabilistic Method*. John Wiley & Sons, 2016
- A. Assi, M. D'Anna, and P. A. García-Sánchez, *Numerical semigroups and applications*. Springer Nature, 2020, vol. 3