

# Random numerical semigroups and sums of subsets of cyclic groups

Santiago Morales Duarte

Universidad de los Andes

*s.morales30@uniandes.edu.co*

December 5, 2023

# Overview

- 1 The probabilistic method
  - Threshold functions
- 2 Numerical semigroups
- 3 Random numerical semigroups
- 4 Experiments
- 5 Results
  - Lower bound
  - Upper bound

# The probabilistic method

## Definition

The Erdős-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  $p(n) \in o(r(n))$ ,  $\lim_{n \rightarrow \infty} \Pr[G(n, p(n)) \models A] = 0$ ,
- 2 When  $r(n) \in o(p(n))$ ,  $\lim_{n \rightarrow \infty} \Pr[G(n, p(n)) \models A] = 1$ ,

or vice versa.

## Example

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

# Second moment method

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

## Theorem

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

**Proof.** Apply Chebyshev'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  $\text{Var}[X] \in o(\mathbb{E}[X]^2)$ ,  $X > 0$  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  $A = \{a_1, \dots, a_n\}$  such that  $S = \langle A \rangle$ , where

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

## 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

44	45	46	47	48	49
38	39	40	41	42	43
32	33	34	35	36	37
26	27	28	29	30	31
20	21	22	23	24	25
14	15	16	17	18	19
8	9	10	11	12	13
2	3	4	5	6	7
-4	-3	-2	-1	0	1

Figure: Visualization of the McNugget semigroup.

## 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

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

## 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

44	45	46	47	48	49
38	39	40	41	42	43
32	33	34	35	36	37
26	27	28	29	30	31
20	21	22	23	24	25
14	15	16	17	18	19
8	9	10	11	12	13
2	3	4	5	6	7
-4	-3	-2	-1	0	1

Figure: Visualization of the McNugget semigroup.

# Numerical semigroup invariants

Note that

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

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

## 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.



## 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  $\lim_{M \rightarrow \infty} p(M) = 0$ , then  $\mathcal{S}$  is co-finite almost always and

$$\lim_{M \rightarrow \infty} E[e(\mathcal{S})] = \lim_{M \rightarrow \infty} E[g(\mathcal{S})] = \lim_{M \rightarrow \infty} E[F(\mathcal{S})] = \infty.$$

- ③ If  $\lim_{M \rightarrow \infty} p(M) > 0$ , then

$$\lim_{M \rightarrow \infty} E[e(\mathcal{S})] < \infty, \quad \lim_{M \rightarrow \infty} E[g(\mathcal{S})] < \infty \quad \text{and} \quad \lim_{M \rightarrow \infty} E[F(\mathcal{S})] < \infty,$$

and each limit is bounded by explicit rational functions in  $p$ .

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,

$$\begin{aligned}\Theta(1) &\leq \lim_{M \rightarrow \infty} \mathbb{E}[e(\mathcal{S})] \leq \Theta\left(\frac{1}{p}\right), \\ \Theta\left(\frac{1}{p}\right) &\leq \lim_{M \rightarrow \infty} \mathbb{E}[g(\mathcal{S})] \leq \Theta\left(\frac{1}{p^2}\right), \\ \Theta\left(\frac{1}{p}\right) &\leq \lim_{M \rightarrow \infty} \mathbb{E}[F(\mathcal{S})] \leq \Theta\left(\frac{1}{p^2}\right).\end{aligned}$$

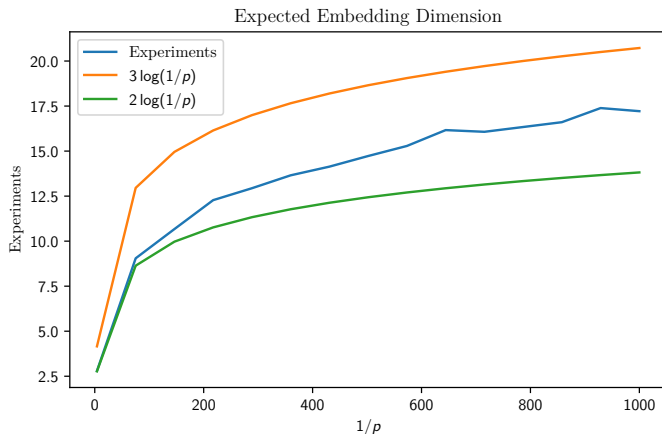
# Experiments

# Average embedding dimension

$1/p$	Lower Bound	$e(S)$	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



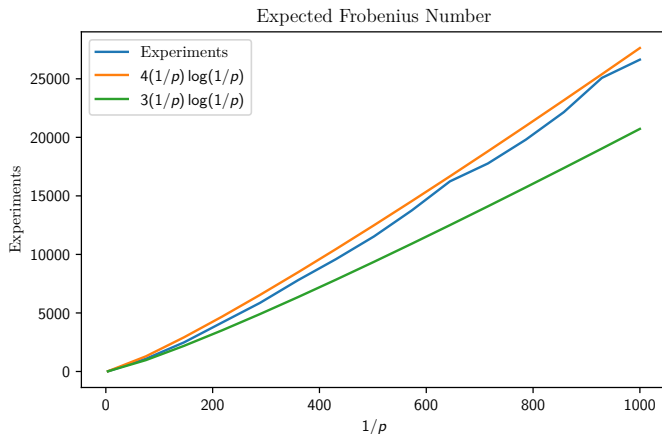
**Figure:** Average embedding dimension of random numerical semigroups generated using the ER-type model vs  $\log(1/p)$ .

# Average Frobenius number

$1/p$	Lower Bound	$F(S)$	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  $(1/p) \log(1/p)$ .

## Conjecture

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



# Results

## Theorem

Let  $S \sim S(M, p)$ , where  $p = p(M)$  is a monotone decreasing function of  $M$  and  $\frac{1}{M} \in o(p(M))$ . Then,

- ① If  $\lim_{M \rightarrow \infty} p(M) = 0$ , then for every  $K \in \mathbb{N}$ ,

$$\lim_{M \rightarrow \infty} \Pr[e(S) > K] = \lim_{M \rightarrow \infty} \Pr[g(S) > K] = \lim_{M \rightarrow \infty} \Pr[F(S) > K] = 1.$$

- ② If  $\lim_{M \rightarrow \infty} p(M) > 0$ , then  $e(S)$ ,  $g(S)$  and  $F(S)$  are bounded in probability, i.e., for every  $\varepsilon > 0$ , there exists  $K_\varepsilon$  such that

$$\Pr[e(S) < K_\varepsilon] > 1 - \varepsilon, \quad \Pr[g(S) < K_\varepsilon] > 1 - \varepsilon \quad \text{and} \quad \Pr[F(S) < K_\varepsilon] > 1 - \varepsilon.$$

Furthermore, for every  $\varphi, \psi$  such that  $(\log x)^2 \in o(\varphi(x))$ ,  $x(\log x)^2 \in o(\psi(x))$ ,

$$\lim_{p \rightarrow 0} \Pr \left[ e(S) < \varphi \left( \frac{1}{p} \right) \right] = \lim_{p \rightarrow 0} \Pr \left[ g(S) < \psi \left( \frac{1}{p} \right) \right] = \lim_{p \rightarrow 0} \Pr \left[ F(S) < \psi \left( \frac{1}{p} \right) \right] = 1.$$

## Lower bound

Let  $\mathcal{S} \sim \mathcal{S}(M, p)$ , where  $p = p(M)$  is a monotone decreasing function of  $M$  and  $\frac{1}{M} \in o(p(M))$ . If  $\lim_{M \rightarrow \infty} p(M) = 0$ , then for every  $K \in \mathbb{N}$ ,

$$\lim_{M \rightarrow \infty} \Pr[e(\mathcal{S}) > K] = \lim_{M \rightarrow \infty} \Pr[g(\mathcal{S}) > K] = \lim_{M \rightarrow \infty} \Pr[F(\mathcal{S}) > K] = 1.$$

To prove this theorem, we show that, for each fixed number of generators  $a$ , there is a high probability that at least  $a$  minimal generators are chosen as  $p \rightarrow 0$ .

# Sums of random subsets of cyclic groups

## Lemma

Let  $q$  be a prime number and  $\mathcal{A}$  be a random subset of  $\mathbb{Z}_q$  of size  $4\lfloor 3\log_2 q \rfloor$ . As  $q$  tends to infinity,  $2\lfloor 3\log_2 q \rfloor \mathcal{A}$  covers  $\mathbb{Z}_q$  almost always.

## Example

Let  $q = 17$ .

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

# Upper bound

## Upper bound

Let  $\psi(x)$  be a function for which  $x(\log x)^2 \in o(\psi(x))$ . Then

$$\lim_{p \rightarrow 0} \Pr \left[ F(\mathcal{S}) \leq \psi \left( \frac{1}{p} \right) \right] = 1.$$

# Upper bound proof strategy

The strategy is to prove that the Ápery set of a subsemigroup of  $S$  is completed before step  $\psi\left(\frac{1}{p}\right)$  with high probability, since  $F(\mathcal{S})$  is less than the maximum element of this Ápery set. The proof has the following structure:

- 1 Using the prime number theorem, we find a step for which a prime  $q$  is chosen with high probability.
- 2 In the spirit of the cyclic group lemma we find a step such that a set  $\mathcal{A}$  of  $\log q$  elements which are different modulo  $q$  are chosen with high probability.
- 3 We apply the cyclic group lemma to  $\text{Ap}(\langle \mathcal{A} \cup \{q\} \rangle)$ .

# Numerical semigroup visualization

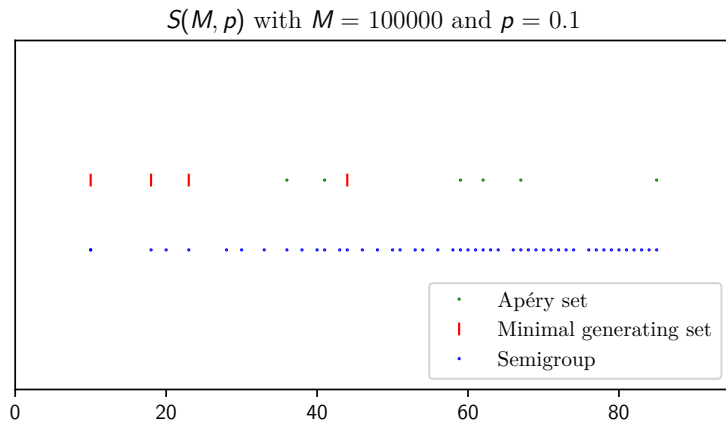


Figure: ER-type random numerical semigroup,  $1/p = 10$ .

# Numerical semigroup visualization

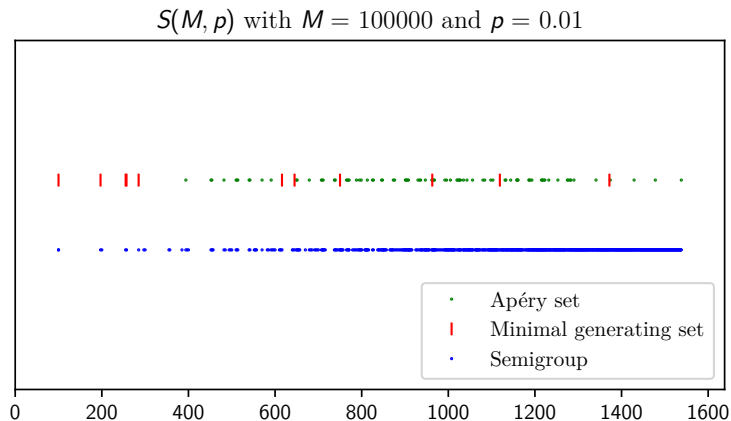


Figure: ER-type random numerical semigroup,  $1/p = 100$ .



# 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