Graduation Project

Santiago Morales - 201913369

May 3, 2023

1 Introduction

2 Kahn-Kalai Conjecture

2.1 Thresholds

Let $n \in \mathbb{N}$ and $0 \leq p \leq 1$. The random graph $G(n, p)$ 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 [1]. Given a graph theoretic property A, there is a probability that $G(n, p)$ satisfies A, which we write as $\Pr[G(n, p) \vDash A]$.

Definition 2.1. $r(n)$ is a threshold function for a graph theoretic 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. [1]

We give an example of a threshold function which illustrates a common method for proving that a function is a threshold.

2.1.1 Threshold function for having isolated vertices

Let G be a graph on n labeled vertices. An isolated vertex of G is a vertex which does not belong to any of the edges of G. Let A be the property that G contains an isolated vertex. We will prove that $r(n) = \dfrac{\ln n}{n}$ is a threshold for A.

For each vertex i in G define the variable

$$X_i = \begin{cases} 1 & \text{if } i \text{ is an isolated vertex,} \\ 0 & \text{if } i \text{ is not an isolated vertex.} \end{cases}$$

Now, the probability that a vertex i is isolated is $(1-p)^{n-1}$ since it is the probability that none of the other $n-1$ vertices is connected to i. Let $X = \sum_{i=1}^{n} X_i$, then the expected number of isolated vertices is

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \Pr[X_i] = n(1-p)^{n-1}.$$

Let $p = k\dfrac{\ln n}{n}$ for $k \in \mathbb{R}_{>0}$. Then

$$\begin{aligned} \lim_{n\to\infty} E[X] &= \lim_{n\to\infty} n\left(1 - k\frac{\ln n}{n}\right)^{n-1} \\ &= ne^{-k\ln n} = n^{1-k}. \end{aligned}$$

Therefore, $\lim_{n\to\infty} E[X] = 0$ if $k > 1$. Since $E[X] \geq \Pr[X > 0]$, we conclude that

$$\lim_{n\to\infty} \Pr[G(n,p) \vDash A] = \lim_{n\to\infty} \Pr[X > 0] = 0.$$

Now, for $k < 1$, the fact that $\lim_{n\to\infty} E[X] = \infty$ is not enough to conclude that $\lim_{n\to\infty} \Pr[G(n,p) \vDash A] = 1$. We have to use the second moment method.

Theorem. *If $E[X] \to \infty$ and $Var[X] = o(E[X]^2)$, then $\lim_{n\to\infty} Pr[X > 0] = 1$. [1]*

Proof. We will prove that, in this case, $Var[X] = o(E[X]^2)$. First,

$$\begin{aligned} \sum_{i\neq j} E[X_i X_j] &= \sum_{i \neq j} \Pr[X_i = X_j = 1] \\ &= n(n-1)(1-p)^{n-1}(1-p)^{n-2} \\ &= n(n-1)(1-p)^{2n-3}, \end{aligned}$$

for if i is an isolated vertex, then there is no edge between i and j so we only have to account for the remaining $n-2$ edges that contain j.

Thus, since $\sum_{i=1}^n E[X_i^2] = \sum_{i=1}^n E[X_i] = E[X]$ and $\lim_{n\to\infty} p = 0$,

$$\lim_{n\to\infty} \frac{\mathrm{Var}[X]}{E[X]^2} = \lim_{n\to\infty} \frac{E(X^2) - E[X]^2}{E[X]^2} = \lim_{n\to\infty} \frac{\sum_{i=1}^n E[X_i^2] + \sum_{i\neq j} E[X_i X_j]}{E[X]^2} - 1$$
$$= 0 + \lim_{n\to\infty} \frac{n(n-1)(1-p)^{2n-3}}{n^2(1-p)^{2n-2}} - 1 = \lim_{n\to\infty} \frac{1}{1-p} - 1 = 0.$$

We conclude that $\mathrm{Var}[X] \in o(E[X]^2)$ and so, if $k < 1$,

$$\lim_{n\to\infty} \Pr[G(n,p) \vDash A] = \lim_{n\to\infty} \Pr[X > 0] = 1.$$

Therefore, $r(n) = \frac{\ln n}{n}$ is a threshold function for property A.

2.2 The expectation threshold

Let X be a set and $p \in [0,1]$. For $A \subset X$, let $P(A) = p^{|A|}(1-p)^{|X\setminus A|}$. A class $\mathcal{F}$ of subsets of X is increasing if for $A \in \mathcal{F}$, $A \subset B$ implies $B \in \mathcal{F}$. If $\mathcal{F} \neq \emptyset$, $P(\mathcal{F}) = \sum_{A\in\mathcal{F}} P(A)$ is a strictly increasing function of p.

TODO Define increasing class and general definition of threshold [2]

TODO Define expectation threshold and show inequalities. [3]

Example 2.1. H_1 example

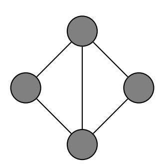

Figure 1: H_1

Example 2.2. H_2 example

Theorem 2.1 (Park Theorem [2], originally Kahn-Kalai Conjecture)**.** ***TODO***

TODO How to prove something is p-small

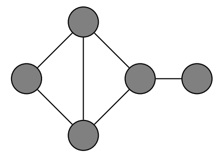

Figure 2: H_2

2.2.1 An application of the Park Theorem

TODO Prove the threshold for perfect matchings for $G(n, p)$

3 Numerical Semigroups

A *numerical semigroup* is a subset $S \subset \mathbb{N}_0$ which is closed under addition, i.e. $a, b \in S$ implies $a + b \in S$. For instance, $\mathbb{N}_0$, $\mathbb{N}_0 \setminus \{0\}$, $2\mathbb{N}_0$ are all numerical semigroups, but $\mathbb{N}_0 \setminus \{2\}$ is not. Some literature requires that a semigroup has a finite complement in $\mathbb{Z}_{\geq 0}$ [4], but we prefer the more general definition.

Example 3.1. TODO McNugget Semigroup

As seen in the previous example, we can describe a numerical semigroup with a set of *generators*:

$$S = \langle a_1, ..., a_n \rangle := \left\{ \sum_{i=1}^{n} k_i a_i : k_1, ..., k_n \in \mathbb{N}_0 \right\}.$$

TODO give some examples.

The *Frobenius number* $F(S)$ of a numerical semigroup $S = \langle a_1, ..., a_n \rangle$ is defined as the largest integer divisible by $d(S) := \gcd(a_1, ..., a_n)$ which does not belong to S. In other terms,

$$F(S) = \max(d\mathbb{Z} \setminus S).$$

4 Probability Spaces over Numerical Semigroups

We generate a random numerical semigroup with a model similar to the Ërdos-Renyi model for random graphs.

Definition 4.1. For $p \in [0,1]$ and $M \in \mathbb{N}$, a random numerical semigroup $S(M,p)$ is a probability space over the set of semigroups $S = \langle \mathcal{A} \rangle$ with $\mathcal{A} \subset \{1, ..., M\}$, determined by

$$\Pr[n \in \mathcal{A}] = p,$$

with these events mutually independent.

5 Expected Frobenius Number

We prove a Theorem found in [5] without the use of the simplicial complex.

Theorem 5.1. *Let $S \sim S(M,p)$, where $p = p(M)$ is a monotone decreasing function of M. If $\frac{1}{M} \ll p \ll 1$, then S is cofinite, i.e., the set of gaps is finite, a.a.s and*

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

Proof. Let $X := \min(S \setminus \{0\})$ be a random variable. Then, for $0 < n \leq M$,

$$\Pr[X = n] = p(1-p)^{n-1},$$

and so

$$\begin{aligned} E[X] &= \sum_{n=0}^{\infty} n\Pr[X = n] = \sum_{n=0}^{M} np(1-p)^{n-1} = p\frac{d}{dp}\left[-\sum_{n=0}^{M}(1-p)^n\right] \\ &= p\frac{d}{dp}\frac{(1-p)^{M+1}-1}{p} = p\frac{1-(1-p)^{M+1}-(M+1)(1-p)^M p}{p^2} \\ &= \frac{1-(1-p)^M - M(1-p)^M p}{p} \geq \frac{1-e^{-Mp}-Mpe^{-Mp}}{p}. \end{aligned}$$

Thus, since $\lim_{M\to\infty} Mp = \infty$, then $\lim_{M\to\infty} Mpe^{-Mp} = \lim_{M\to\infty} e^{-Mp} = 0$, which implies that

$$\lim_{M\to\infty} E[X] = \lim_{M\to\infty} \frac{1-e^{-Mp}-Mpe^{-Mp}}{p} = \infty.$$

Also, note that if $p = \frac{c}{M}$, $c \in \mathbb{R}_+$ $(0 < e^{-c} + ce^{-c} < 1)$,

$$\lim_{M\to\infty} E[X] = \lim_{M\to\infty} \frac{1-e^{-c}-ce^{-c}}{p} = \infty.$$

Proof. Fix $a \in \mathbb{N}$ such that $a > 11$ and let $A = \{1, \ldots, \lfloor \frac{a}{p} \rfloor\}$. Since $\frac{1}{M} \ll p$, we have that $\lfloor \frac{a}{p} \rfloor \leq M$ for large enough M. Consider the following events:

- E_1: No generator selected is less than $\frac{1}{ap}$.

  Let X_1 be the number of generators selected from $\{1, \ldots, \lfloor \frac{1}{ap} \rfloor\}$. Then

  $$\Pr[\overline{E_1}] = \Pr[X_1 > 0] \leq E[X_1] \leq p \cdot \frac{1}{ap} = \frac{1}{a}.$$

- E_2 : At most $\frac{3a}{2}$ generators are selected from A.

  Let X_2 be the number of generators selected in A, then X_2 is a binomial random variable with $n = \frac{a}{p}$ and we can use the bound (Feller I can add this to the appendix)

  $$\Pr[\overline{E_2}] = \Pr\left[X_2 > \frac{3a}{2}\right] \leq \frac{\frac{3a}{2}(1-p)}{(\frac{3a}{2} - a)^2} \leq \frac{6}{a}.$$

  Also, note that by the union bound

  $$\Pr[E_1 \wedge E_2] \leq 1 - \frac{1}{a} - \frac{6}{a} = 1 - \frac{7}{a}.$$

- E_3 : At least $\frac{a}{2}$ generators are selected from A.

  Similarly, we can use the bound for the other tail of the distribution so that

  $$\Pr[\overline{E_3}] = \Pr\left[X_2 < \frac{a}{2}\right] \leq \frac{(n - \frac{a}{2})p}{(np - \frac{a}{2})^2} = \frac{a - (\frac{a}{2})p}{(\frac{a}{2})^2} \leq \frac{4}{a}.$$

- E_4 : The generators selected from A are minimal.

  Let $Y_{(1)}, Y_{(2)}, \ldots, Y_{(k)}$ denote the first k generators selected in A. Assume E_1 and E_2. We have that E_1 implies $Y_{(1)} \geq \frac{1}{ap}$ and E_2 implies $k \leq \frac{3a}{2}$.

  First we bound for the probability that, given E_1 and E_2, $b \in A$ is selected as a generator. By conditional probability

  $$\begin{aligned}\Pr[b \text{ is selected}] = {} & \Pr[b \text{ is selected} | E_1 \wedge E_2] \Pr[E_1 \wedge E_2] \\ & + \Pr[b \text{ is selected} | \overline{E_1 \wedge E_2}] \Pr[\overline{E_1 \wedge E_2}],\end{aligned}$$

and so

$$\Pr[b \text{ is selected}|E_1 \wedge E_2] \leq \frac{\Pr[b \text{ is selected}]}{\Pr[E_1 \wedge E_2]} \leq \frac{p}{1-\frac{7}{a}}.$$

Then, using the same reasoning, for any Now, note that $Y_{(2)}$ is not minimal if a multiple of $Y_{(1)}$ is selected in A. Thus, if we fix $Y_{(1)} = y_1 \geq \frac{1}{ap}$, $Y_{(1)}$ is not minimal if $b \in \{2y_1, 3y_1, \ldots, c_1y_1\}$ is selected, where c_1y_1 is the largest multiple of y_1 which does not exceed $\frac{a}{p}$. Since $y_1 \geq \frac{1}{ap}$, we have that $c_1 \leq a^2$. Then, using the union bound,

$$\Pr[Y_{(2)} \text{ is not minimal}|E_1 \wedge E_2 \wedge Y_{(1)} = y_1] \leq \frac{pa^2}{1-\frac{7}{a}}.$$

If we sum over all possible y_1, we get that

$$\Pr[Y_{(2)} \text{ is not minimal}|E_1 \wedge E_2] \leq \frac{pa^2}{1-\frac{7}{a}}.$$

Similarly, for $2 \leq t \leq k$ and fixed $Y_{(1)} = y_1, \ldots, Y_{(t-1)} = y_{t-1}$, $Y_{(t)}$ is not minimal if the first $t-1$ numbers selected from A can generate $Y_{(t)}$. For the possible numbers generated by the first t numbers selected, there are at most a^2 choices for each coefficient, so there are at most a^{2t} such linear combinations. Then

$$\Pr[Y_{(t)} \text{ is not minimal}|E_1 \wedge E_2] \leq \frac{pa^{2t}}{1-\frac{7}{a}}.$$

Therefore, since $Y_{(1)}$ is always minimal, we can use the union bound and $k \leq \frac{3a}{2}$ to conclude that

$$\Pr[E_4|E_1 \wedge E_2] \geq 1 - \frac{p}{1-\frac{7}{a}} \sum_{t=1}^{\frac{3a}{2}-1} a^{2t} = 1 - o(1).$$

Thus,

$$\Pr[E_4] = \Pr[E_4|E_1 \wedge E_2]\Pr[E_1 \wedge E_2] \geq 1 - \frac{7}{a} - o(1).$$

Finally, note that by union bound,

$$\Pr[E_4 \wedge E_3] \geq 1 - \frac{11}{a} - o(1).$$

Therefore, for every $N \in \mathbb{N}$ and $\varepsilon > 0$, there exists K such that $M \geq K$ implies

$$\Pr[f(S) > N],\ \Pr[g(S) > N],\ \Pr[e(S) > N] > 1 - \varepsilon.$$

.

References

[1] N. Alon and J. H. Spencer, *The probabilistic method.* John Wiley & Sons, 2016.

[2] J. Park and H. T. Pham, "A proof of the kahn-kalai conjecture," *arXiv e-prints*, pp. arXiv–2203, 2022.

[3] K. Frankston, J. Kahn, B. Narayanan, and J. Park, "Thresholds versus fractional expectation-thresholds," *Annals of Mathematics*, vol. 194, no. 2, pp. 475–495, 2021.

[4] S. Chapman, R. Garcia, and C. O'Neill, "Beyond coins, stamps, and chicken mcnuggets: An invitation to numerical semigroups," *A Project-Based Guide to Undergraduate Research in Mathematics: Starting and Sustaining Accessible Undergraduate Research*, pp. 177–202, 2020.

[5] J. De Loera, C. O'Neill, and D. Wilburne, "Random numerical semigroups and a simplicial complex of irreducible semigroups," *arXiv preprint arXiv:1710.00979*, 2017.