A Short Proof of Kahn-Kalai Conjecture

Phuc Tran[a] Van Vu[a]

Submitted: Jul 30, 2023; Accepted: Jun 25, 2024; Published: Jul 12, 2024

Abstract

In a recent paper, Park and Pham famously proved Kahn-Kalai conjecture. In this note, we simplify their proof, using an induction to replace the original analysis. This reduces the proof to one page and from the argument it is also easy to read that one can set the constant K in the conjecture to ≈ 3.998, which could be the best value under the current method. Our argument also applies to the ϵ-version of the Park-Pham result, studied by Bell.

Mathematics Subject Classifications: 05C80, 05D40, 60C05.

1 Introduction

Let X be a set of N elements and $0 \leqslant p \leqslant 1$. The measure μ_p is defined on the subsets of X by $\mu_p(S) = p^{|S|}(1-p)^{N-|S|}$. This is the measure generated by choosing each element of X independently with probability p. For a family F of subsets, $\mu_p(F) := \sum_{S \in F} \mu_p(S)$. Furthermore, let $\mathbf{E}_p(|F|) = \sum_{S \in F} p^{|S|}$ be the expectation for the number of elements of F in the chosen set. We call F *increasing family* if F satisfies the increasing property, meaning that if $B \supset A \in F$, then $B \in F$. Given a family F, let $\langle F \rangle$ be the collection of subsets of X which contain some elements of F, namely $\langle F \rangle := \{T : T \supset S, S \in F\}$. We say that G covers F if $F \subset \langle G \rangle$.

Define $p_c(F)$ to be point where $\mu_{p_c}(\langle F \rangle) = 1/2$, and $p_E(G)$ be the point where $\mathbf{E}_{p_E}(|G|) = 1/2$. Let $q(F) = \max\{p_E(G) \mid G \text{ covers } F\}$. It is clear that

$$p_E(F) \leqslant q(F) \leqslant p_c(F). \tag{1}$$

Finally, we say that F is l-bounded if its elements have size at most l.

Example. Let X be the set of all edges of the complete graph K_n on a set V of n vertices. Thus, $|X| = N = \binom{n}{2}$ and each subset $S \subset X$ corresponds to a graph on V. For each $0 \leqslant p \leqslant 1$, μ_p is the measure of the random graph $G(n,p)$. Let F be the collection of Hamiltonian cycles on V. It is clear that F is n-bounded and $\mathbf{E}_p(F) = \frac{(n-1)!p^n}{2}$

[a]Department of Mathematics, Yale University, 10 Hillhouse Ave, New Haven, Connecticut, U.S.A. (phuc.tran@yale.edu, van.vu@yale.edu).

(since $|F| = \frac{(n-1)!}{2}$) is the expectation of the number of Hamiltonian cycles in $G(n,p)$. Furthermore, $p_c(F)$ is the critical point where $G(n,p)$ contains a Hamiltonian cycle with probability 1/2.

By Stirling's formula, $q(F) \geqslant p_E(F) = (1+o(1))\frac{e}{n}$. The computation of $p_c(F)$ is harder, and classical theorems in random graph theory show that

$$p_c(F) = (1+o(1))\frac{\ln n}{n} = (\ln 2 + o(1))\frac{\log n}{n} \approx .693\frac{\log n}{n},$$

where (following tradition) we assume that $\log x$ has base 2.

Kahn and Kalai [2] conjectured that there is a constant K such that for any increasing family F, $p_c(F) \leqslant Kq(F)\log n$. This was the central question in random graph theory for many years. In 2021, a weaker version of this conjecture was proved by Frankston, Kahn, Narayanan and Park [3], inspired by an exciting development from [4]. This version already contained the most interesting applications in the literature. A year later, Park and Pham [5], also using ideas from [4], settled the conjecture in an even stronger form.

Theorem 1 (Park-Pham [5])**.** *There is a constant $K > 0$ such that for any l-bounded increasing family F, $p_c(F) \leqslant Kq(F)\log l$.*

This note grew out of our attempt to teach Theorem 1 in class. We found a shortcut using induction, avoiding the relatively technical analysis in [5]. This simplifies the proof of the main result of [5] (Theorem 3 below) and reduces its length to about one page. The other ingredients remain the same.

This short proof also reveals that when $l \to \infty$ (which is the interesting case in applications) we can set $K \approx 3.998$, which we believed to be the best constant with respect to the current approach. The previous record is $K = 8$, see [1].

2 The covering theorem

Fix a positive number $p \leqslant 1$. In the power set 2^X, the m-level, denoted by L_m, is the family of all subsets of size m. Clearly $|L_m| = \binom{N}{m}$. For a family H of subsets of X, the cost to cover H, denoted by $f_p(H)$, is the minimum expectation with respect to p of a family G such that $H \subset \langle G \rangle$. This cost function is sub-additive, if H is partitioned into H_1 and H_2, then $f_p(H) \leqslant f_p(H_1) + f_p(H_2)$. If H is empty, then $f_p(H) = 0$. On the other hand, if H contains the empty set, then $f_p(H) = 1$.

For an increasing family F, let $c_t(F) = \frac{|F \cap L_t|}{|L_t|}$ be the fraction of level t in F. By double counting, it is easy to show that

Fact 2. *$c_t(F)$ is increasing with t.*

Set $m_l := \lfloor LpN\log(l+1) \rfloor$ where L is a sufficiently large constant. The main result of [5] is the following

Theorem 3 (Park-Pham [5]). *There is a constant L such that the following holds. Assume that H is l-bounded and $f_p(H) \geqslant 1/2$. Then $\langle H \rangle$ contains at least $\frac{2}{3}$ fraction of the m_l-level (L_{m_l}). In other words*

$$|\langle H \rangle \cap L_{m_l}| \geqslant \frac{2}{3}|L_{m_l}|.$$

As shown in [5], Theorem 1 follows quickly from Theorem 3; see Remark (6) for details. We will prove an (artificially) stronger variant, whose parameters are set for induction.

Theorem 4 (Covering theorem). *Let $0 \leqslant l \leqslant N$ be integers and X be a set of size N. Assume that H is an l-bounded family of subsets of X and $f_p(H) \geqslant \frac{1}{2} - \frac{1}{2^{l+2}}$. Then $\langle H \rangle$ contains at least $\frac{2}{3} + \frac{1}{2^{l+2}}$ fraction of the m_l-level.*

We prove Theorem 4 by induction on l and N. For $l = 0$, $f_p(H) \geqslant 1/4 > 0$. This means that H consists of the empty set, and the conclusion is trivial (for any N) as $\langle \emptyset \rangle$ contains L_m for all $0 \leqslant m \leqslant N$. Now we prove for a pair $1 \leqslant l \leqslant N$, assuming that the hypothesis holds for all values $l' < l$ and $N' \leqslant N$.

For a set $W \subset X$, define $H_W := \{S \backslash W : S \in H\}$. Let H'_W be the family of minimal sets of H_W and let $G_W := \{T : T \in H'_W, |T| > .9l\}$ and $\tilde{H}_W := H'_W \backslash G_W$. The following lemma is the key estimation.

Lemma 5 (Double counting lemma). *For $L > 1000$ and $w = \lfloor .1LpN \rfloor$, we have*

$$\sum_{W:|W|=w} \mu_p(\langle G_W \rangle) \leqslant \binom{N}{w} \frac{1}{8 \times 16^l}.$$

Proof. Notice that $\mu_p(\langle G_W \rangle) \leqslant \sum_{S' \in G_W} p^{|S'|}$. Therefore,

$$\sum_{W:|W|=w} \mu_p(\langle G_W \rangle) \leqslant \sum_{W:|W|=w} \sum_{k} \sum_{S':S' \in G_W, |S'|=k} p^k. \tag{2}$$

To bound the RHS, we bound the number of pairs (W, S'), in which W is a set of size w and $S' \in G_W$ has exactly k elements. To determine (W, S'), we first fix the union $W' = W \cup S'$ and then choose $S' \in G_{W' \backslash S'} : |S'| = k$. There are $\binom{N}{w+k}$ ways to choose W'. Once this union is fixed, pick a set $S \in H$ inside the union (there must be at least one, namely, the one that defines S', but we can choose any). For each possible $S' \subset W'$ (i.e. $W = W' \setminus S', S' \in G_W, |S'| = k$), if $S' \not\subset S$, then $S' \cap S = S \setminus W$ is minimal. It contradicts the fact that $S' \in G_W$ is a minimal set. Thus, S' must be a subset of S (regardless of whether S' is defined by S or not). This gives at most $\binom{l}{k}$ choices of S'. Finally, by the definition of the system, we only need to consider $k > .9l$. Therefore, with $w = \lfloor .1LpN \rfloor$, the RHS of (2) is at most

$$\sum_{.9l<k\leqslant l} \binom{N}{w+k}\binom{l}{k} p^k \leqslant \binom{N}{w} \sum_{.9l<k\leqslant l} (.1L)^{-k} \binom{l}{k} \leqslant \binom{N}{w} (.1L)^{-.9l} 2^{l+1} \leqslant \frac{\binom{N}{w}}{8 \times 16^l}, \tag{3}$$

given that $L \geqslant 1000$. This proves the lemma.

□

Back to Theorem 4, we say that W is *good* if $\mu_p(\langle G_W\rangle) \leqslant \sum_{S'\in G_W} p^{|S'|} \leqslant \frac{1}{2^{l+2}}$. By averaging, at most an $\frac{1}{2\times 8^l}$ fraction of all W are *bad*. For a good W, by subadditivity, we have

$$f_p(\tilde{H}_W) \geqslant f_p(H'_W) - f_p(G_W) \geqslant f_p(H) - \frac{1}{2^{l+2}} \geqslant \frac{1}{2} - \frac{1}{2^{l+1}} \geqslant \frac{1}{2} - \frac{1}{2^{l_1+2}},$$

with $l_1 := \lfloor .9l \rfloor < l$.

By the induction hypothesis, $\langle \tilde{H}_W\rangle$ contains $\frac{2}{3} + \frac{1}{2^{l_1+2}}$ fraction of the m_{l_1}-level of the ground set $X\backslash W$, $|X\backslash W| = N - w$. By taking the union with W (for good W), it follows that $\langle H\rangle$ contains at least

$$\frac{2}{3} + \frac{1}{2^{l_1+2}} - \frac{1}{2\times 8^l} \geqslant \frac{2}{3} + \frac{1}{2^{l+2}}$$

fraction of the $(m_{l_1} + w)$-level, in which $m_{l_1} + w = \lfloor Lp(N-w)\log(l_1+1)\rfloor + \lfloor .1LpN\rfloor \leqslant \lfloor LpN \log(l+1)\rfloor = m_l$, given that $L \geqslant 1000$. By Fact 2, our proof is complete.

Remark 6. One can easily deduce Kahn-Kalai conjecture from the main theorem, with $K = L(1+\epsilon)$, for any fixed $\epsilon > 0$. Let us consider $G(n,p)$, the argument for random hypergraphs is similar. Set $N = \binom{n}{2}$. Notice that if we choose each edge with probability $p = \rho(1+\epsilon)$ where $pN \to \infty$, then with probability $1 - o(1)$, the resulting graph has at least $m = \rho N$ edges. Thus we can generate $G(n,p)$ (barring an event of probability $o(1)$) by first generating a random number $\bar{m}$ of value at least m (according to the binomial distribution, but this does not really matter), and then hit a uniform random point on the $\bar{m}$ level and take the corresponding graph. So with probability at least $2/3 - o(1)$ we hit a point in $\langle H\rangle$.

3 Reducing the constant K

In this section, we show that when $l \to \infty$, we can reduce L to approximately 3.998, and then K (by Remark 6) also approximates 3.998. This seems to be the best value with respect to the current method.

Let $0 < \delta < 1$ be a constant and set L (with foresight) such that $3 > \epsilon = \frac{(L\log(1/\delta))^\delta}{2} - 1 > 0$. The smallest value for L so that this holds for some $1 > \delta > 0$ is $L \approx 3.998\ldots$ which is slightly larger than the minimal value of $\frac{2^{1/\delta}}{\log(1/\delta)}$ over the interval $(0,1)$.

Let l_0 be a natural number such that $2 \leqslant (1+\epsilon/3)^{l_0 - \lfloor \delta l_0\rfloor}$. Set $m_l = \lfloor LpN\log(l+1) + 1000pN\log(l_0+1)\rfloor$. With a small modification, we prove the following ϵ-version of Theorem 4.

Theorem 7. *Let $l_0 \leqslant l \leqslant N$ be integers and X be a set of size N. Assume that H is l-bounded and $f_p(H) \geqslant \frac{1}{2} - (1+\epsilon/3)^{-l}$. Then $\langle H\rangle$ contains at least a $\frac{2}{3} + (1+\epsilon/3)^{-l}$ fraction of the m_l-level.*

Proof. We start the induction at $l = l_0$. This base case is covered by Theorem 4 and results in the term $1000pN\log(l_0+1)$. Next, we replace .9 by δ and consider $w := \lfloor LcpN\rfloor$ with

$c := \log(1/\delta)$. Thus, in (3), instead of $\sum_{.9l<k\leqslant l}(.1L)^{-k}\binom{l}{k}$, we end up with

$$\begin{aligned}\sum_{W:|W|=w}\mu_p(\langle G_W\rangle) &\leqslant \binom{N}{w}\sum_{\delta l<k\leqslant l}(Lc)^{-k}\binom{l}{k}\\ &\leqslant \binom{N}{w}(Lc)^{-l\delta}\sum_{\delta l<k\leqslant l}\binom{l}{k}\\ &\leqslant \binom{N}{w}(Lc)^{-l\delta}2^l\\ &= \binom{N}{w}\left[\frac{(Lc)^\delta}{2}\right]^{-l}\\ &= \binom{N}{w}(1+\epsilon)^{-l}\ \text{(by Definition of } L \text{ and } \epsilon\text{)}\\ &\leqslant \binom{N}{w}(1+\epsilon/3)^{-l}(1+\epsilon/3)^{-l}\ \text{(since }\ \epsilon<3\text{).}\end{aligned} \tag{4}$$

We say that W is *good* if $\mu_p(\langle G_W\rangle) \leqslant \sum_{S'\in G_W} p^{|S'|} \leqslant (1+\epsilon/3)^{-l}$. By averaging, at most an $(1+\epsilon/3)^{-l}$ fraction of all W are *bad*. For a good W, by subadditivity, we have

$$f_p(\tilde{H}_W) \geqslant f_p(H'_W) - f_p(G_W) \geqslant f_p(H) - (1+\epsilon/3)^{-l} \geqslant \frac{1}{2} - 2(1+\epsilon/3)^{-l} \geqslant \frac{1}{2} - (1+\epsilon/3)^{-l_1},$$

with $l_1 = \lfloor \delta l \rfloor$, thanks to the assumption $2 \leqslant (1+\epsilon/3)^{l_0-\lfloor \delta l_0\rfloor}$. By applying the induction hypothesis for l_1 and taking union with W (for the good W), it follows that $\langle H\rangle$ contains at least

$$\frac{2}{3} + (1+\epsilon/3)^{-l_1} - (1+\epsilon/3)^{-l} \geqslant \frac{2}{3} + (1+\epsilon/3)^{-l},$$

fraction of the $(m_{l_1}+w)$-level for $m_{l_1}+w = \lfloor Lp(N-w)\log(l_1+1) + 1000pN\log(l_0+1)\rfloor + w$. By the settings of l_1 and c, it is easy to check that $m_{l_1}+w$ is at most $m_l = \lfloor LpN\log(l+1) + 1000pN\log(l_0+1)\rfloor$. We complete the induction and the proof. □

4 Covering theorem for arbitrary small ϵ_1

In the previous sections, we proved that if $f_p(H) > 1/2$, then $\langle H\rangle$ contains at least $2/3 = 1-\frac{1}{3}$ fraction of the m_l-level, for sufficiently large m_l. In [1], Bell considered the question of how large should m_l be if we replace $\frac{1}{3}$ by an arbitrary $\epsilon_1 > 0$. He proved in [1, Theorem 3] that the covering theorem still holds for $m_l = \lfloor 48pN\log l + 48pN\log\left(\frac{1}{\epsilon_1}\right)\rfloor$. By combining our induction with Bell's result, we can prove the following bound for sufficiently small ϵ_1.

Define $L \approx 3.998$ and its corresponding l_0 as in the beginning of Section 3. Set $m_l = \lfloor LpN\log(l+1) + 96pN\log\left(\frac{1}{\epsilon_1}\right)\rfloor$.

Theorem 8. *Let* $0 < \epsilon_1 < 1/l_0$ *be a positive number which may depend on* N*. Assume that* H *is* l*-bounded and* $f_p(H) > \frac{1}{2}$*. Then* $\langle H \rangle$ *contains at least* $1 - \epsilon_1$ *fraction of the* m_l*-level.*

Proof of Theorem 8 . We start the induction at $l = \lfloor \frac{1}{\epsilon_1} \rfloor - 1$. This base case is covered by Bell's theorem and results in the term $96pN \log\left(\frac{1}{\epsilon_1}\right)$. When $l \geqslant \lfloor \frac{1}{\epsilon_1} \rfloor - 1$, we follow the proof of Theorem 7 with $1 - \epsilon_1$ replacing $\frac{2}{3}$ and obtain the remaining term $LpN \log(l + 1)$. □

The first author could further improve the constant from 96 to 16; details will appear in a sequential work.

Acknowledgements

The research is partially supported by Simon Foundation award SFI-MPS-SFM-00006506 and NSF grant AWD 0010308. We thank J. Park for comments and pointing out reference [1].

References

[1] T. Bell, The Park-Pham theorem with optimal convergence rate. *The Electronic Journal of Combinatorics*, (2) **30** (2023), #P2.25.

[2] J. Kahn and G. Kalai, Thresholds and expectation thresholds, *Combin. Probab. Comput.* **16** (2007), 495-502.

[3] K. Frankston, J. Kahn, B. Narayanan, J. Park, Thresholds versus fractional expectation-thresholds. *Ann. Mathematics* **194** (2021), 475-495.

[4] R. Alweiss, L. Schachar, K. Wu and J. Zhang, Improved bounds for the sunflower lemma, *Ann. of Math.* (2) **194** (2021), no 3, 795-815.

[5] J. Park, H. Pham, A Proof of the Kahn-Kalai conjecture, *Journal of the American Mathematical Society* (1) **37** (2024), 235-243.