

©Copyright 2018
Christopher Fowler

Random Permutations and Simplicial Complexes

Christopher Fowler

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Christopher Hoffman, Chair

Isabella Novik

Ioana Dumitriu

Program Authorized to Offer Degree:
Mathematics

University of Washington

Abstract

Random Permutations and
Simplicial Complexes

Christopher Fowler

Chair of the Supervisory Committee:
Professor Christopher Hoffman
Mathematics

We study the asymptotic behavior of distributions on two different combinatorial objects, permutations and simplicial complexes.

First we study strong α -logarithmic measures on the symmetric group, including the well-studied Ewens sampling formula (see [21] and [20] for reference). We show that, for almost every α , precisely $\lceil(1 - \alpha \log 2)^{-1}\rceil$ are need to invariably generate S_n asymptotically with positive probability. The corollary is that any fewer permutations will with high probability not invariably generate S_n . In particular, for several measures on S_n no finite number is sufficient.

Then we direct our attention to a multi-parameter measure on simplicial complexes. This measure is an interpolation between random clique complexes and Linial-Meshulam random k -dimensional complexes, both subjects of considerable attention over the last two decades. We extend results for each into this new model, establishing upper and lower thresholds for the appearance of nontrivial cohomology in each dimension and characterize the behavior at one window of criticality. We also manage to establish the size of homology within these regimes. Notably, unlike these other distributions, multi-parameter complexes can exhibit nontrivial homology in numerous dimensions simultaneously.

TABLE OF CONTENTS

	Page
Chapter 1: Introduction	1
1.1 Invariable generation and the Ewens sampling formula	1
1.2 Homology of multiparameter random simplicial complexes	4
Bibliography	9
Chapter 2: Ewens Sampling and Invariable Generation	14
Chapter 3: Homology of Multi-Parameter Random Simplicial Complexes	62

ACKNOWLEDGMENTS

I wish to express my sincerest appreciation to Christopher Hoffman. For being a patient, incredibly smart, and helpful advisor. For tolerating the countless times I stormed into his office claiming to have solved it all, only to return the following day, tail between my legs.

I also owe a debt to Matthew Kahle, who introduced me to my first research problem and urged me to finish it before it was swiped from under me. I'm grateful for our time working together on the bouquet of spheres project that, while ultimately unsuccessful, has had a profound impact on me. Plus, it was simply a fun problem to work on. His enthusiasm for the project, and collaboration in general, was truly infectious.

I'd be remiss if I didn't thank my family and friends. My father has been fostering my development as a mathematician since I was three, and my mom has always kept me grounded and cognizant that there is more than one type of intelligence. They have been unerringly supportive throughout my academic life, from my obsession with Thoreau, to majoring in political science, to graduate school. I can never thank them enough. Finally, in many ways Matt Junge served as a role model for the type of graduate student I wanted to be. Most importantly, he taught me to derive pleasure and value from my teaching responsibilities, when it's so easy to view it as an onerous chore. That's something that will stick with me forever.

DEDICATION

to my mother and father

Chapter 1

INTRODUCTION

This thesis concerns two disparate problems lying in the intersection of probability and combinatorics. Chapter 2 is the work of [15], completed in collaboration with Brito, Levy and Junge. We establish the number of random permutations required to invariably generate the symmetric group with positive probability when the distribution of the cycle counts has the strong α -logarithmic property. In doing so, we show that for many probability measures on the symmetric group, no fixed number of permutations is sufficient to invariably generate S_n with high probability. Chapter 3 contains the work in [31], exploring a new probability measure on random simplicial complexes, one that serves as an interpolation between two models of recent interest. It extends many of the results for these distributions to this generalized regime, as well as exhibiting the potential for unique homological behavior.

1.1 Invariable generation and the Ewens sampling formula

Elements g_1, \dots, g_m of a group G are said to *generate* the group if the smallest subgroup containing them is G itself. The *symmetric group* S_n is the group of permutations of the set $[n] := \{1, \dots, n\}$. We say a set of permutations π_1, \dots, π_m *generate* S_n *invariably* if π'_1, \dots, π'_m generate S_n whenever π'_i is conjugate to π_i for all i [23].

The investigation of random permutations has a storied history, as does the accompanying question of invariable generation. In 1934, van der Waerden raised the question of the minimal number of uniformly random permutations required to invariably generate S_n with positive probability [58]. Recent breakthroughs by Pemantle, Peres, Rivin [57] and Eberhard, Ford, Green [25] resolved van der Waerden's question, showing that precisely four permutations are required. We consider the same question in the context of more general

measures related to the cycle structure of a permutation. The primary example is the Ewens sampling formula introduced in [Ewe72].

The Ewens sampling formula aligns a permutation's cycle structure with conditioned independent Poisson random variables. Let $\text{Poi}(\lambda)$ denote the law of a Poisson random variable with mean λ . Fix $\alpha > 0$ and let X_1, \dots, X_n be independent random variables in which X_k has law $\text{Poi}(\alpha/k)$. For a permutation $\pi \in S_n$ let C_k denote the number of k -cycles in π . More precisely, C_k is the number of disjoint k -element subsets of $[n]$ on which π restricts to a cyclic permutation. We say that the vector $\mathbf{C} = (C_1, \dots, C_n)$ has distribution $\text{ESF}(\alpha, n)$ if it satisfies the joint distribution:

$$\mathbf{P}[C_1 = x_1, \dots, C_n = x_n] = \mathbf{P}\left[X_1 = x_1, \dots, X_n = x_n \mid \sum_1^n kX_k = n\right]$$

The work of [25] and [57] rely on a nice property of the small-cycle structure of uniformly random permutations. We state the result as presented in [25, Lemma 1.3]. Let (X_1, X_2, \dots) be a sequence of independent Poisson random variables, where X_i has parameter $1/i$. Then for a uniformly random permutation $\pi \in S_n$ and a fixed constant k , the distribution of (C_1, \dots, C_k) converges to that of (X_1, \dots, X_k) . So the cycle structure of a uniformly random permutation is highly reminiscent of the definition of Ewens sampling formula, a crucial fact in both proofs. This is because the uniformly random distribution on S_n is precisely $\text{ESF}(\alpha, n)$ with $\alpha = 1$.

The Ewens sampling formula appears throughout mathematics, statistics and the sciences. In the words of Harry Crane in his survey [21], it “exemplifies the harmony of mathematical theory, statistical application, and scientific discovery.” That survey and its followup [20] give a nice tour of the formula's universal character, describing applications to evolutionary molecular genetics, the neutral theory of biodiversity, Bayesian nonparametrics, combinatorial stochastic processes, and inductive inference, to name a few. The sampling formula also underpins foundational mathematics in number theory, stochastic processes, and algebra.

We present our primary result as it relates to Ewens sampling formula. We define m_α to

be

$$m_\alpha = \inf_{m \geq 2} \left\{ m : \inf_{n \geq 2} \mathbf{P} [\{\pi_1, \dots, \pi_m\} \subset S_n, \pi_i \sim \text{ESF}(\alpha, n), \text{invariably generate } S_n] > 0 \right\}.$$

Let $h(\alpha)$ be a function on α given by

$$h(\alpha) = \begin{cases} \lceil (1 - \alpha \log 2)^{-1} \rceil, & 0 < \alpha < 1/\log 2 \\ \infty, & \alpha \geq 1/\log 2 \end{cases}.$$

Theorem 1 *For the continuity points of h it holds that $m_\alpha = h(\alpha)$. At points of discontinuity we have $h(\alpha) \leq m_\alpha \leq h(\alpha) + 1$.*

In other words, suppose $\pi_1, \dots, \pi_m \sim \text{ESF}(\alpha, n)$. If $m < h(\alpha)$, then with high probability the set of permutations do not invariably generate S_n . But if $m \geq h(\alpha)$, then they invariably generate with probability bounded away from 0.

Invariable generation is defined to be conjugacy invariant, and the conjugacy classes of the symmetric group are precisely the permutations of the same cycle type. A permutation fixes a set of size k if some combination of its cycle lengths sums up to k . If a collection of permutations π_1, \dots, π_m all fix sets of size k , then there exist a set of conjugates π'_1, \dots, π'_m that all fix the same set of size k , and thus do not generate all of S_n . For this reason, invariable generation concerns only the cycle structure of permutations.

Much of our work in establishing an upper bound for m_α involved understanding the fixed sets of permutations. Creative arguments were employed to extend several classical results on uniformly random permutations to the realm of Ewens sampling formula. One crucial generalization was of [49, Theorem 1].

Theorem 2 *Let $\pi \sim \text{ESF}(\alpha, n)$. Then with probability $1 - o(1)$ the only transitive subgroups containing π are A_n and S_n .*

Among other pieces required in proving this was an adaptation of a result by Erdős and Turán in [27, Theorem V].

Theorem 3 *Let $\pi \sim \text{ESF}(\alpha, n)$ and $\omega(n)$ be an arbitrary function with $\omega(n) \rightarrow \infty$. The probability that the largest prime which divides $\prod_{\ell=1}^n \ell^{C_\ell}$ is larger than $n \exp(-\omega(n)\sqrt{\log n})$ is $1 - o(1)$.*

As stated above, the sizes of a permutation's fixed sets correspond to the sums of its cycle lengths. Moreover, the cycle counts of $\text{ESF}(\alpha, n)$ -permutations can be cleanly related by the Feller coupling, established in [6, Theorem 3.1, Theorem 3.2], to independent $\text{Poi}(\alpha/k)$ random variables. Therefore, invariable generation is related to sumsets formed from random multisets. For the lower bound of Theorem 1, we direct our attention to sequences of independent Poisson random variables $\mathbf{X}(\alpha) = (X_1, X_2, \dots)$ where X_j has mean α/j . We then model the sizes of fixed sets of our permutations by the random sumset

$$\mathcal{L}(\mathbf{X}(\alpha)) = \left\{ \sum_{j \geq 1} jx_j : 0 \leq x_j \leq X_j \right\}.$$

Now consider independent realizations $\mathbf{X}^{(1)}(\alpha), \dots, \mathbf{X}^{(m)}(\alpha)$. Our bounds correspond to the extremal values of m such that

$$\bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)}(\alpha)) = \{0\} \text{ with positive probability,}$$

and

$$\left| \bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)}(\alpha)) \right| = \infty \text{ almost surely.}$$

As mentioned before, we extend numerous results for uniform permutations to the Ewens sampling formula and beyond. We adapt large chunks of [27, 49, 14] to this new regime, replacing counting arguments with a tool known as the Feller coupling, introduced in [30, p. 815].

1.2 Homology of multiparameter random simplicial complexes

The study of random topological spaces began with random graphs, the seminal example of which is $G(n, p)$, the Erdős–Rényi model. Given a probability parameter $p \in (0, 1)$, typically a function of n , we consider a graph on n vertices where every edge between two vertices of

G is added independently with probability p . This defines a probability measure on the set of all simple graphs on n vertices, and we say $G \stackrel{dist}{=} G(n, p)$ to indicate G is a random graph with law $G(n, p)$.

The purpose of the second work is to understand the homological behavior of a generalized model for random simplicial complexes, mentioned in [43] and first explored concurrently in [19] and [31]. We define $X(n, p_1, p_2, \dots)$ to be the probability distribution over simplicial complexes on vertex set $[n] = \{1, \dots, n\}$, with the distribution on 1-skeletons agreeing with $G(n, p_1)$. The distribution on higher dimensional skeletons is iteratively constructed: for an integer $k > 1$, any k -simplex with boundary contained in our complex is present with probability p_k . This provides a measure on all simplicial complexes on n vertices.

This measure also provides a bridge between two well studied distributions, random k -complexes $Y_k(n, p)$ and random clique complexes $X(n, p)$. The former, called Linial–Meshulam complexes for $k = 2$ and Meshulam–Wallach complexes for higher dimensions, starts with a simplicial complex on n vertices with full $(k - 1)$ -skeleton, and contains every possible k -simplex independently with probability p . This can be realized as $X(n, 1, \dots, 1, p_k = 0, 0, \dots)$ under our new framework. The random clique complex model $X(n, p)$ is defined as the clique complex of the graph $G(n, p)$, where every complete subgraph on $k + 1$ vertices defines a k -simplex of the simplicial complex. This distribution can be interpreted as $X(n, p, 1, \dots)$. Similarly the random graph $G(n, p)$ is $X(n, p, 0, \dots)$.

A formative result of random graph theory, proven by Erdős and Rényi in [27], was the sharp threshold of $p = \log n/n$ for connectivity in $G(n, p)$: if $p \geq (\log n + \omega(1))/n$ then $G(n, p)$ is with high probability connected, and if $p \leq (\log n - \omega(1))/n$ then $G(n, p)$ is with high probability disconnected. The result pertains to the asymptotic behavior of this model based on the parameter p . Indeed, previous work on all three aforementioned random simplicial complexes often pertains to the asymptotic behavior of the model, ie. what happens as the number of vertices n tends to infinity. Given some property \mathcal{A} of simplicial complexes, we

say that $X \in \mathcal{A}$ with high probability, or *w.h.p.*, if

$$\lim_{n \rightarrow \infty} \mathbf{P} [X \in \mathcal{A}] = 1.$$

Significant work has been done on the behavior of random graphs since [27], and random simplicial complexes emerged as a natural high dimension analog. Some the most natural questions to ask relate to the homological or cohomological behavior of these complexes. Indeed, this is effectively an extension of the graph connectivity question for $G(n, p)$. The 0-homology of a graph G is defined by $H_0(G, \mathbf{Z}) = \mathbf{Z}^m$ where m is the number of connected components of G .

In this context there are two different types of phase transitions that occur in these models. For any given dimension, there can be a threshold at which homology or cohomology changes from trivial to nontrivial. Conversely, there can be a threshold at which it goes from nontrivial back to trivial, or *vanishes*. Extensive work on has been done to establish lower bounds on the thresholds at which homology appears and upper bounds on the thresholds at which homology vanishes for various models.

Notation: We write $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ to indicate that X is chosen from the distribution $X(n, p_1, p_2, \dots)$.

Our theorems deal with the $(k - 1)$ -th homology or cohomology of $X(n, p_1, p_2, \dots)$. As mentioned above there are two types of phase transitions, and we work to develop bounds on the thresholds for both. Since the $(k - 1)$ -th (co)homology of a simplicial complex depends only on its k -skeleton, these theorems only depend on probabilities p_1 through p_k .

Theorem 4 *Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with $p_i = n^{-\alpha_i}$ and $\alpha_i \geq 0$ for all i . If*

$$\sum_{i=1}^k \alpha_i \binom{k}{i} < 1, \tag{1.1}$$

then w.h.p. $H^{k-1}(X, \mathbf{Q}) = 0$.

We prove this threshold is essentially the best possible by establishing nontrivial cohomology on the other side of (1.1). Moreover, the second regime for which cohomology exists is establishes the potential for $H^k(X, \mathbf{Q}) \neq 0$ simultaneously for several k .

Theorem 5 Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with $p_i = n^{-\alpha_i}$, $\alpha_i \geq 0$ for all i , and

$$1 \leq \sum_{i=1}^k \alpha_i \binom{k}{i}. \quad (1.2)$$

If

$$\sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i} < 1, \quad (1.3)$$

then w.h.p. $H^{k-1}(X, \mathbf{Q}) \neq 0$. Furthermore, when $\alpha_k > 0$ we can relax the condition in (1.3) to

$$\sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1} < k+1. \quad (1.4)$$

A common question to ask concerning phase transitions is what happens at the boundary between phases. Given a complex X , we let β_k denote the k -th Betti number of X , given by $\beta_k := \dim(H^k(X, \mathbf{Q}))$. Allowing the p_i to be more varied functions of n , we identify this critical region and establish a limit theorem for the Betti number.

Theorem 6 Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with

$$p_i = (\rho_1 \log n + \rho_2 \log \log n + c)^{\nu_i} n^{-\alpha_i},$$

where

$$\rho_1 = k - \sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}, \quad \rho_2 = \sum_{i=1}^{k-1} \nu_i \binom{k}{i+1} \quad \text{and} \quad \sum_{i=1}^k \alpha_i \binom{k}{i} = \sum_{i=1}^k \nu_i \binom{k}{i} = 1.$$

Then β_{k-1} approaches a Poisson distribution

$$\beta_{k-1} \rightarrow \text{Poi}(\mu)$$

with mean

$$\mu = \frac{\rho_1^{\rho_2} e^{-c}}{k!}.$$

This distribution differs from previously explored ones in one important way. Namely, there are clear windows for these parameters in which a complex will w.h.p. exhibit nontrivial homology in several dimensions simultaneously.

Theorem 7 *Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with $p_i = n^{-\alpha_i}$, for any integer l there exists an integer k and an open set of α_i for which X w.h.p. has non-trivial cohomology in dimensions k through $k + l$.*

BIBLIOGRAPHY

- [1] Tom M. Apostol. *Introduction to analytic number theory*. Undergraduate texts in mathematics. Springer, New York, Berlin, Heidelberg, 1976. Autres tirages : 1984, 1986, 1995, 1998, 2010.
- [2] Lior Aronshtam and Nathan Linial. When does the top homology of a random simplicial complex vanish? *Random Structures & Algorithms*, 46(1):26–35, 2015.
- [3] Lior Aronshtam, Nathan Linial, Tomasz Łuczak, and Roy Meshulam. Collapsibility and vanishing of top homology in random simplicial complexes. *Discrete & Computational Geometry*, 49(2):317–334, 2013.
- [4] R. A. Arratia. Independent process approximations for combinatorial structures. *Advances in Applied Probability*, 24(4):764–765, 007 2016.
- [5] Richard Arratia, Andrew Barbour, and Simon Tavaré. Poisson process approximations for the ewens sampling formula. *Ann. Appl. Probab.*, 2(3):519–535, 08 1992.
- [6] Richard Arratia, Andrew Barbour, and Simon Tavaré. Limits of logarithmic combinatorial structures. *Ann. Probab.*, 28(4):1620–1644, 10 2000.
- [7] Richard Arratia, Andrew Barbour, and Simon Tavaré. Exploiting the feller coupling for the ewens sampling formula. *Statist. Sci.*, 31(1):27–29, 02 2016.
- [8] Richard Arratia and Simon Tavaré. The cycle structure of random permutations. *Ann. Probab.*, 20(3):1567–1591, 07 1992.
- [9] Richard Arratia and Simon Tavaré. Independent process approximations for random combinatorial structures. *Adv. Math.*, 104:90–154, 1994.
- [10] László Babai. On the order of uniprimitive permutation groups. *Annals of Mathematics*, 113(3):553–568, 1981.
- [11] Eric Babson, Christopher Hoffman, and Matthew Kahle. The fundamental group of random 2-complexes. *Journal of the American Mathematical Society*, 24(1):1–28, 2011.

- [12] Werner Ballmann and J Świątkowski. On l_2 -cohomology and property (t) for automorphism groups of polyhedral cell complexes. *Geometric & Functional Analysis GAFA*, 7(4):615–645, 1997.
- [13] Volker Betz, Daniel Ueltschi, and Yvan Velenik. Random permutations with cycle weights. *Ann. Appl. Probab.*, 21(1):312–331, 02 2011.
- [14] JD Bovey. The probability that some power of a permutation has small degree. *Bulletin of the London Mathematical Society*, 12(1):47–51, 1980.
- [15] Gerandy Brito, Christopher Fowler, Matthew Junge, and Avi Levy. Ewens sampling and invariable generation. *Combinatorics, Probability and Computing*, pages 1–39, 2018.
- [16] A Costa and M Farber. Large random simplicial complexes, i. *Journal of Topology and Analysis*, 8(03):399–429, 2016.
- [17] A Costa and M Farber. Large random simplicial complexes, ii; the fundamental group. *Journal of Topology and Analysis*, 9(03):441–483, 2017.
- [18] A Costa and M Farber. Large random simplicial complexes, iii the critical dimension. *Journal of Knot Theory and Its Ramifications*, 26(02):1740010, 2017.
- [19] Armindo Costa and Michael Farber. Random simplicial complexes. pages 129–153, 2016.
- [20] Harry Crane. Rejoinder: The ubiquitous ewens sampling formula. *Statist. Sci.*, 31(1):37–39, 02 2016.
- [21] Harry Crane. The ubiquitous ewens sampling formula. *Statist. Sci.*, 31(1):1–19, 02 2016.
- [22] J.H. Davenport and G.C. Smith. Fast recognition of alternating and symmetric galois groups. *Journal of Pure and Applied Algebra*, 153(1):17 – 25, 2000.
- [23] John D. Dixon. Random sets which invariably generate the symmetric group. *Discrete Mathematics*, 105(1):25 – 39, 1992.
- [24] Sean Eberhard, Kevin Ford, and Ben Green. Permutations fixing a k -set. *International Mathematics Research Notices*, 2015.
- [25] Sean Eberhard, Kevin Ford, and Ben Green. Invariable generation of the symmetric group. *Duke Math. J.*, 166(8):1573–1590, 06 2017.

- [26] Nicholas M. Ercolani and Daniel Ueltschi. Cycle structure of random permutations with cycle weights. *Random Structures & Algorithms*, 44(1):109–133, 2014.
- [27] P Erdős and P Turán. On some problems of a statistical group-theory. ii. *Acta Mathematica Hungarica*, 18(1-2):151–163, 1967.
- [28] Leonhard Euler. Observationes circa series, infinitas. *Commentarii Academiae Scientiarum Petropolitanae*, 9:160–188, 1737.
- [29] W.J. Ewens. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3(1):87 – 112, 1972.
- [30] William Feller. The fundamental limit theorems in probability. *Bulletin of the American Mathematical Society*, 51(11):800–832, 1945.
- [31] Christopher F Fowler. Generalized random simplicial complexes. *arXiv preprint arXiv:1503.01831*, 2015.
- [32] Howard Garland. p-adic curvature and the cohomology of discrete subgroups of p-adic groups. *Annals of Mathematics*, pages 375–423, 1973.
- [33] A. Gladkich and R. Peled. On the Cycle Structure of Mallows Permutations. *ArXiv e-prints*, January 2016.
- [34] A. Gladkich and R. Peled. On the Cycle Structure of Mallows Permutations. *ArXiv e-prints*, January 2016.
- [35] Andrew Granville. The anatomy of integers and permutations. *preprint.*, 2008.
- [36] Robert Guralnick and Kay Magaard. On the minimal degree of a primitive permutation group. *Journal of Algebra*, 207(1):127–145, 1998.
- [37] Allen Hatcher. *Algebraic topology*. Cambridge University Press, 2002.
- [38] Joos Heintz. On polynomials with symmetric galois group which are easy to compute. *Theoretical Computer Science*, 47:99 – 105, 1986.
- [39] Christopher Hoffman, Matthew Kahle, and Elliot Paquette. Spectral gaps of random graphs and applications to random topology. *arXiv preprint arXiv:1201.0425*, 2012.
- [40] Christopher Hoffman, Matthew Kahle, and Elliot Paquette. The threshold for integer homology in random d-complexes. *arXiv preprint arXiv:1308.6232*, 2013.

- [41] Matthew Kahle. Topology of random clique complexes. *Discrete Mathematics*, 309(6):1658–1671, 2009.
- [42] Matthew Kahle. Sharp vanishing thresholds for cohomology of random flag complexes. *Ann. of Math.*, 179:1085–1107, 2014.
- [43] Matthew Kahle. Topology of random simplicial complexes: a survey. *AMS Contemp. Math*, 620:201–222, 2014.
- [44] Matthew Kahle, Elizabeth Meckes, et al. Limit theorems for betti numbers of random simplicial complexes. *Homology, Homotopy and Applications*, 15(1):343–374, 2013.
- [45] R. Kenyon, D. Kral, C. Radin, and P. Winkler. Permutations with fixed pattern densities. *ArXiv e-prints*, June 2015.
- [46] Dmitry Kozlov. The threshold function for vanishing of the top homology group of random d-complexes. *Proceedings of the American Mathematical Society*, 138(12):4517–4527, 2010.
- [47] Martin W Liebeck and Jan Saxl. Minimal degrees of primitive permutation groups, with an application to monodromy groups of covers of riemann surfaces. *Proceedings of the London Mathematical Society*, 3(2):266–314, 1991.
- [48] Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. *Combinatorica*, 26(4):475–487, 2006.
- [49] Tomasz Luczak and Lászlò Pyber. On random generation of the symmetric group. *Combinatorics, Probability and Computing*, 2:505–512, 12 1993.
- [50] H. Maier and G. Tenenbaum. On the set of divisors of an integer. *Inventiones mathematicae*, 76(1):121–128, 1984.
- [51] C. L. Mallows. Non-null ranking models. i. *Biometrika*, 44(1/2):114–130, 1957.
- [52] Roy Meshulam and Nathan Wallach. Homological connectivity of random k-dimensional complexes. *Random Structures & Algorithms*, 34(3):408–417, 2009.
- [53] S. Mukherjee. Fixed points and cycle structure of random permutations. *ArXiv e-prints*, September 2015.

- [54] Sumit Mukherjee. Fixed points and cycle structure of random permutations. *Electron. J. Probab.*, 21:18 pp., 2016.
- [55] David R. Musser. On the efficiency of a polynomial irreducibility test. *J. ACM*, 25(2):271–282, April 1978.
- [56] G. Olshanski. Random permutations and related topics. *ArXiv e-prints*, April 2011.
- [57] Robin Pemantle, Yuval Peres, and Igor Rivin. Four random permutations conjugated by an adversary generate sn with high probability. *Random Structures & Algorithms*, pages n/a–n/a, 2015.
- [58] B.L. van der Waerden. Die seltenheit der gleichungen mit affekt. *Mathematische Annalen*, 109:13–16, 1934.

Chapter 2

EWENS SAMPLING AND INVARIABLE GENERATION

EWENS SAMPLING AND INVARIABLE GENERATION

GERANDY BRITO, CHRISTOPHER FOWLER, MATTHEW JUNGE, AND AVI LEVY

ABSTRACT. We study the number of random permutations needed to invariably generate the symmetric group S_n when the distribution of cycle counts has the strong α -logarithmic property. The canonical example is the Ewens sampling formula, for which the special case $\alpha = 1$ corresponds to uniformly random permutations.

For strong α -logarithmic measures, and almost every α , we show that precisely $\lceil (1 - \alpha \log 2)^{-1} \rceil$ permutations are needed to invariably generate S_n with asymptotically positive probability. A corollary is that for many other probability measures on S_n no fixed number of permutations will invariably generate S_n with positive probability. Along the way we generalize classic theorems of Erdős, Tehran, Pyber, Luczak and Bovey to permutations obtained from the Ewens sampling formula.

CONTENTS

1. Introduction	2
1.1. Ewens sampling and the logarithmic property	3
1.2. Statement of Theorem 1	5
1.3. Generalizations for the Ewens sampling formula	7
1.4. Overview of proof and result for sumsets	8
1.5. Further questions	10
1.6. Notation	12
1.7. Outline of paper	12
1.8. Acknowledgements	13
2. Proof of Theorem 1	13
3. The upper bound for m_α	15
3.1. The Feller Coupling	16
3.2. Proofs of Theorem 3, Theorem 5, and Theorem 6	24
3.3. Proving Proposition 8	34

4. The upper bound for s_α	35
5. The lower bound for s_α	39
References	46

1. INTRODUCTION

In 1934, van der Waerden raised the question of the minimal number of uniformly random permutations required to invariably generate S_n with positive probability [Wae34]. Recent breakthroughs by Pemantle, Peres, Rivin [PPR15] and Eberhard, Ford, Green [EFG17] resolved van der Waerden's question, showing that precisely four permutations are required. We consider the same question in the context of more general permutation measures related to the cycle structure. The primary example is the Ewens sampling formula (1) introduced in [Ewe72]. It aligns a permutation's cycle structure with conditioned independent Poisson random variables and is defined at (1).

Elements g_1, \dots, g_m of a group G generate the group if the smallest subgroup containing them is G itself. The symmetric group S_n is the group of permutations of the set $[n] := \{1, \dots, n\}$. We say a set of permutations π_1, \dots, π_m generate S_n invariably if π'_1, \dots, π'_m generate S_n whenever π'_i is conjugate to π_i for all i [Dix92]. Invariable generation is conjugacy invariant, and conjugacy classes of the symmetric group are precisely permutations of the same *cycle type*, defined to be the multiset of cycle lengths in the cycle decomposition of a permutation. A permutation fixes a set of size k if some combination of the permutations cycle lengths sums up to k . If a collection of permutations π_1, \dots, π_m all fix sets of size k , then we can choose a set of conjugates π'_1, \dots, π'_m that all fix the same set of size k , and thus do not generate all of S_n . For this reason, invariable generation concerns only the cycle structure of permutations, rather than the action on $[n]$.

Recent developments show that cycle counts converge to independent Poisson random variables for many measures on S_n as n tends to infinity. Arratia, Barbour and Tavaré have shown that strong α -logarithmic measures have this property and numerous further results [AT92, ABT92, ABT00, ABT16]. Measures with general cycle weights are considered in [BUV11, EU14] and related permutations are introduced in [KKRW15]. Following this work, Mukherjee used Stein's method to deduce that a variety of permutation measures have Poisson limiting cycle counts [Muk16]. Furthermore, the structure of the short cycles in the Mallows measure was characterized in [GP16].

In addition to its intrinsic interest, the minimal number of permutations required for invariable generation has applications to computational Galois theory. One may use this quantity to estimate the runtime of a Monte Carlo polynomial factorization algorithm [Mus78, Hei86, DS00, PPR15]. This establishes a connection between the irreducible factors of polynomials and the cycle structure of permutations. The cycle structure of a permutation is also compared with prime factorization of integers in the survey [Gra08]. An analogy between divisors of large integers and invariable generation of permutations is then mentioned in the introduction of [EFG17], postulating a link between the threshold for invariable generation and small divisors of a set of random integers.

Despite the age of van der Waerden’s question, it was only recently shown that the number of uniformly random permutations required to invariably generate S_n is bounded. A bound of $O(\sqrt{\log n})$ was first established in [Dix92]. This was improved to $O(1)$ in [LP93], but the constant was large ($\approx 2^{100}$). Pemantle, Peres and Rivin reduced the bound to four in [PPR15], which was then shown to be sharp by Eberhard, Ford and Green in [EFG17].

We extend the results for uniform permutations to the Ewens’ sampling formula and beyond. Along the way, we extend results in [ET67, LP93, Bov80], replacing counting arguments with a tool known as the Feller coupling (10) introduced in [Fel45, p. 815].

1.1. Ewens sampling and the logarithmic property. The Ewens sampling formula appears throughout mathematics, statistics and the sciences. In the words of Harry Crane it, “exemplifies the harmony of mathematical theory, statistical application, and scientific discovery.” This is stated in his survey [Cra16b]. It and its followup [Cra16a] give a nice tour of the formula’s universal character, describing applications to evolutionary molecular genetics, the neutral theory of biodiversity, Bayesian nonparametrics, combinatorial stochastic processes, and inductive inference, to name a few. The sampling formula also underpins foundational mathematics in number theory, stochastic processes, and algebra.

We start by formally defining the *Ewens sampling formula*, as well as the strong α -logarithmic property. Let $\text{Poi}(\lambda)$ denote the law of a Poisson random variable with mean λ . Fix $\alpha > 0$ and let X_1, \dots, X_n be independent random variables in which X_k has law $\text{Poi}(\alpha/k)$. For a permutation $\pi \in S_n$ let C_k denote the number of k -cycles in π . More precisely, C_k is the number of disjoint k -element subsets of $[n]$ on which π restricts to a cyclic permutation. We say that the vector

$\mathbf{C} = (C_1, \dots, C_n)$ has distribution $\text{ESF}(\alpha, n)$ if it satisfies the joint distribution:

$$(1) \quad \mathbb{P}[C_1 = x_1, \dots, C_n = x_n] = \mathbb{P}\left[X_1 = x_1, \dots, X_n = x_n \mid \sum_1^n kX_k = n\right].$$

As observed in [Ewe72] this can be written explicitly as

$$\mathbb{P}[C_1 = x_1, \dots, C_n = x_n] = \frac{n!}{\alpha_{(n)}} \prod_{j=1}^n \binom{\alpha}{j} \frac{1}{j} \frac{1}{x_j!} \mathbf{1}_{\{\sum_{\ell=1}^n \ell \cdot x_\ell = n\}}$$

where $\alpha_{(n)} = \alpha(\alpha + 1) \cdots (\alpha + n - 1)$. The formulation in (1) will be most useful for our purposes.

The tuple (C_1, \dots, C_n) uniquely specifies the cycle type of a permutation. Furthermore, it is easy to see that as π ranges over S_n , the tuple (C_1, \dots, C_n) associated to π ranges over the set

$$\left\{ (C_1, \dots, C_n) \in \{0, \dots, n\}^n : \sum_{k=1}^n kC_k = n \right\}.$$

Thus, (1) specifies a measure on the conjugacy classes of S_n . We extend this to a measure on S_n (also denoted by $\text{ESF}(\alpha, n)$) by weighting all elements of the same class uniformly. Note that $\text{ESF}(1, n)$ is the uniform measure, and therefore the results we establish regarding $\text{ESF}(\alpha, n)$ will generalize the corresponding results about uniform permutations appearing in [PPR15, EFG17].

A more general family that includes $\text{ESF}(\alpha, n)$ is obtained by replacing the X_k in (1) with an arbitrary sequence of independent nonnegative integer valued random variables Z_1, Z_2, \dots . To avoid intractable distortions when conditioning on $\sum kZ_k = n$, logarithmic growth of $\sum Z_k$ is traditionally assumed (see [ABT00]). For our purposes, we require the Z_k satisfy the *strong α -logarithmic condition*:

$$(2) \quad \begin{aligned} |i\mathbb{P}[Z_i = 1] - \alpha| &< e(i)c_1, \\ i\mathbb{P}[Z_i = \ell] &\leq e(i)c_\ell, \quad \ell \geq 2, \end{aligned}$$

where $\sum i^{-1}e(i)$ and $\sum \ell c_\ell$ are both finite.

Besides Poisson, the most commonly used distributions for the Z_i are the negative-binomial and binomial distributions. These are called assemblies, multisets, and selections, respectively (see [ABT00] for more discussion). The proof that all three of these different distributional choices for the Z_i satisfy the strong α -logarithmic property is in [ABT00, Proposition 1.1].

It is no trouble to work at this level of generality, because the limiting cycle structure of any family satisfying the strong α -logarithmic property is the same as that for an $\text{ESF}(\alpha, n)$ permutation.

We will describe this in more detail in Section 1.4, but in short, the number of ℓ -cycles converges to an independent Poisson with mean α/ℓ .

1.2. Statement of Theorem 1. We state our result in terms of random variables Z_1, Z_2, \dots satisfying the strong α -logarithmic property at (2). The example to keep in mind though is the measure $\text{ESF}(\alpha, n)$, since all of these have the same limiting cycle counts.

Given such a collection let $\mu_n(\alpha, Z_1, Z_2, \dots, Z_n)$ be the measure induced on S_n via the relation at (1) with X_k replaced by Z_k . We define $m_\alpha = m_\alpha(Z_1, Z_2, \dots)$ to be the minimum number of permutations sampled according to μ_n to invariably generate S_n with positive probability as $n \rightarrow \infty$:

$$(3) \quad m_\alpha = \inf_{m \geq 2} \left\{ m : \inf_{n \geq 2} \mathbb{P}[\{\pi_1, \dots, \pi_m\} \subset S_n, \pi_i \sim \mu_n, \text{ invariably generate } S_n] > 0 \right\}.$$

Our theorem gives a closed formula for $m_\alpha = h(\alpha)$, save for a countable exceptional set. Because we will reference it several times, we write the expression here:

$$(4) \quad h(\alpha) = \begin{cases} \lceil (1 - \alpha \log 2)^{-1} \rceil, & 0 < \alpha < 1/\log 2 \\ \infty, & \alpha \geq 1/\log 2 \end{cases}.$$

See Figure 1.

Theorem 1. *Let Z_1, Z_2, \dots be a collection of random variables satisfying the strong α -logarithmic condition at (2). Let $m_\alpha = m_\alpha(Z_1, Z_2, \dots)$, as in (3), be the minimum number of permutations to invariably generate S_n with positive probability. Let $h(\alpha)$ be as in (4). For points of continuity of h it holds that $m_\alpha = h(\alpha)$. At points of discontinuity we have $h(\alpha) \leq m_\alpha \leq h(\alpha) + 1$.*

This has broader implications beyond measures that have the strong α -logarithmic property. We use it to deduce that any finite collection of random permutations with cycle counts asymptotically dominating a $\text{Poi}(1/j \log 2)$ random variable will fail to generate a transitive subgroup of S_n .

Corollary 2. *Suppose that $c_j \geq 1/\log 2$ and $X_k \sim \text{Poi}(c_j/j)$. Let κ_n be a sequence of probability measures on S_n such that for random permutations $\pi \sim \kappa_n$ the cycle structure, \mathbf{C} satisfies*

$$d((C_1, \dots, C_{k_0}), (X_1, \dots, X_{k_0}))_{TV} \rightarrow 0$$

for all fixed k_0 . Then, the probability that any fixed number of permutations sampled according to κ_n invariably generate a transitive subgroup of S_n goes to 0 as $n \rightarrow \infty$.

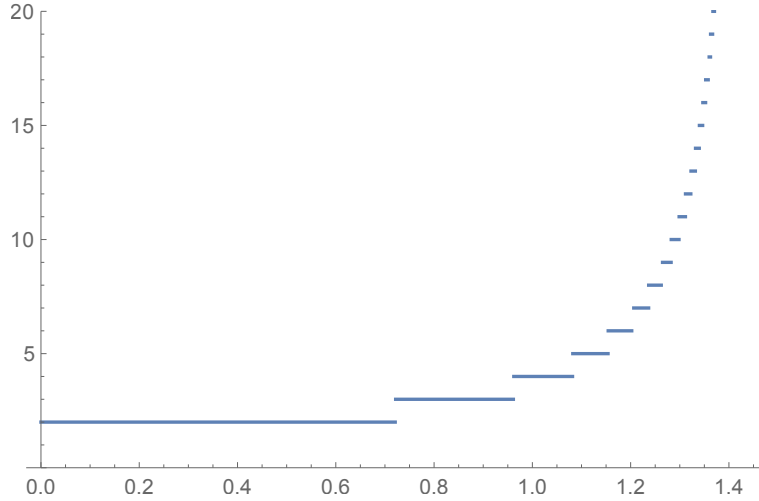


FIGURE 1. The graph $h(\alpha)$ defined at (4). We show that $m_\alpha = h(\alpha) = s_\alpha$, but are not sure about the value at the discontinuities of h . See Question 1 for more discussion.

This corollary is particularly relevant to the results in [BUV11] and [EU14]. They describe measures on S_n that are formed by weighting cycle lengths by parameters c_j . The limiting cycle structure has $C_j \sim \text{Poi}(c_j/j)$. Thus, we can create any limiting Poisson cycle structure we like. In particular, those satisfying Corollary 2.

Another source of alternate measures with limiting Poisson cycle counts comes from [Muk16]. This fits into the larger body of work on permutons initiated in [KKRW15]. Of particular interest is Mallow's measure (introduced in [Mal57]). This measure specifies a parameter $q_n > 0$ that biases towards more or less inversions in a permutation. A permutation π has probability proportional to $q_n^{\text{inv}(\pi)}$, where $\text{inv}(\pi) := |\{(s, t) | s < t \text{ and } \pi s > \pi t\}|$.

The recent article [GP16] begins to characterize the cycle structure of Mallow's measure. They find that for $(1 - q_n)^{-2} \ll n$ all cycles are on the order of $o(n)$. The high density of small cycles, along with [Muk16, Theorem 1.4], which proves a limiting Poisson cycle profile, suggests that no finite collection of permutations sampled according to Mallow's measure in this regime will invariably generate S_n . Note that in the regime $(1 - q_n)^{-2} \gg n$ the cycle counts converge to those of a uniformly random permutation. This is a new and exciting area. We are hopeful our result will find more applications as these objects become better understood.

1.3. Generalizations for the Ewens sampling formula. In proving Theorem 1 we connect an approximation sumset model (described in Section 1.4) back to fixed-set sizes in random permutations. This also takes place in [PPR15] and [EFG17]. These two articles use a small-cycle limit theorem for uniformly random permutations ([AT92, Theorem 1]). Namely, that the number of ℓ -cycles in a uniformly random permutation converges to a Poisson random variable with mean $1/\ell$ that is independent of the other small cycles.

Permutations sampled from μ_n have an analogous limit theorem ([ABT00, Theorem 3.2]). Except now the ℓ -cycles behave like Poisson random variables with mean α/ℓ . This lets us use similar ideas to analyze the corresponding sumset model. However, our approach diverges significantly when we connect back to random permutations. Especially in obtaining an upper bound on m_α . The difference is that both [PPR15] and [EFG17] have access to a large canon of results for uniformly random permutations. We do not. This requires a deep excavation where we extend classical results to $\pi \sim \text{ESF}(\alpha, n)$. Along the way we prove many new estimates (see Lemma 11, Lemma 12, Lemma 13, Proposition 14, Lemma 19, and Lemma 21) for larger cycle lengths in $\text{ESF}(\alpha, n)$ permutations.

We begin by showing that with high probability the only transitive subgroups containing a permutation π sampled with distribution $\text{ESF}(\alpha, n)$ are A_n and S_n . This is the analogue of what [LP93, Theorem 1] proves for uniformly random permutations. This was recently studied in more detail by Eberhard, Ford, Green in [EFG15]. We plan to explore this vein for $\text{ESF}(\alpha, n)$ permutations in future work. For our current purposes, the following result suffices.

Theorem 3. *Let $\pi \sim \text{ESF}(\alpha, n)$. Then with probability $1 - o(1)$ the only transitive subgroups containing π are A_n and S_n .*

Pyber and Luczak rely on an important theorem of Bovey regarding primitive subgroups. We need the analogue of it for $\text{ESF}(\alpha, n)$ permutations. The *minimal degree* of $\langle \pi \rangle = \{\pi^k : k \in \mathbb{Z}\}$ is the minimum number of elements of $\{1, \dots, n\}$ displaced by some power of π that is not the identity.

Example 4. *Consider $\pi = (1234)(567)(89)$ and $\tau = (1234567)(89)$ in S_9 . The order of π displacing the fewest elements of $\{1, \dots, 9\}$ is $\pi^4 = (567)$ so the minimal degree of $\langle \pi \rangle$ is 3. Meanwhile, the order of τ displacing the fewest elements is $\tau^7 = (89)$, so the minimal degree of $\langle \tau \rangle$ is 2.*

[Bov80, Theorem 1] says that if π is a uniformly random permutation, then for each $\varepsilon > 0$ and $0 < \beta < 1$ we have

$$\mathbb{P}[\text{minimal degree of } \langle \pi \rangle > n^\beta] < C_{\varepsilon, \beta} n^{\varepsilon - \beta}.$$

We establish a weaker analogue of this result.

Theorem 5. *Let $\pi \sim \text{ESF}(\alpha, n)$. For each $0 < \beta < 1$ and any α it holds that*

$$\mathbb{P}[\text{minimal degree of } \langle \pi \rangle > n^\beta] = o(1).$$

Establishing Theorem 3 also requires a generalization of a classical theorem of Erdős and Turán (see [ET67, Theorem V]).

Theorem 6. *Let $\pi \sim \text{ESF}(\alpha, n)$ and $\omega(n)$ be an arbitrary function with $\omega(n) \rightarrow \infty$. The probability that the largest prime which divides $\prod_{\ell=1}^n \ell^{C_\ell}$ is larger than $n \exp(-\omega(n)\sqrt{\log n})$ is $1 - o(1)$.*

The proof of Theorem 3 uses both Theorem 5 and Theorem 6. For all three generalizations we follow a somewhat similar blueprint to their predecessors. However, the previous work often uses special features of uniformly random permutations. Counting arguments are heavily employed. In general, these techniques do not translate to $\text{ESF}(\alpha, n)$ permutations. Our approach is to use the Feller coupling (see Section 3.1) to directly relate the cycle structure to independent Poisson random variables. In some places this yields more elegant proofs, in others it becomes a bit technical. Keeping in mind that $\text{ESF}(1, n)$ is the uniform measure, this is a nice high-level approach to extending these results.

1.4. Overview of proof and result for sumsets. Because a fixed set's size is a sum of cycle lengths of a permutation, invariable generation is related to sumsets formed from random multisets. The link is developed over a collaborative arc of Arratia, Barbour and Tavaré, [ABT92, ABT00, AT94, ABT16]. They study the relationship between α -logarithmic structures and Poisson random variables. Most relevant for our purposes are the descriptions of the cycle counts $\mathbf{C} = (C_1, \dots, C_n)$ for permutations sampled according to μ_n .

They prove in [ABT00, Theorem 3.1, Theorem 3.2] that, for permutations induced by the strong α -logarithmic property, the small cycle counts evolve to be independent Poisson random variables, and the large cycle counts behave according to Ewens sampling formula. Furthermore, [ABT92, Lemma 1] (also discussed in [ABT16]) shows that the Ewens sampling formula can be cleanly related via the Feller coupling to independent $\text{Poi}(\alpha/k)$ random variables. The payoff is that we can model the sizes of fixed sets in a permutation with independent Poisson random variables.

Let $\mathbf{X}(\alpha) = (X_1, X_2, \dots)$ be a sequence of Poisson random variables where X_j has mean α/j . The coupling discussed in the previous paragraph allows us to model the sizes of fixed sets in a

permutation with the random sumset

$$\mathcal{L}(\mathbf{X}(\alpha)) = \left\{ \sum_{j \geq 1} j x_j : 0 \leq x_j \leq X_j \right\}.$$

Returning to our opening question, we can bound the required number of permutations from above if no fixed set size is common to all permutations. And we can bound the number from below if there is guaranteed to be a common fixed set size. Consider independent realizations $\mathbf{X}^{(1)}(\alpha), \dots, \mathbf{X}^{(m)}(\alpha)$. The upper and lower bounds correspond to finding extremal values of m such that

- (i) $\bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)}(\alpha)) = \{0\}$ with positive probability, and
- (ii) $|\bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)}(\alpha))| = \infty$ almost surely.

Now the cycle counts of these permutations converge to independent Poisson random variables [AT92, ABT92, ABT00, ABT16]. Thus common elements among independent sumsets correspond to common fixed-set sizes of independent permutations. So invariable generation is analogous to a trivial sumset intersection. On the other hand, an infinite intersection means that common fixed-sets persist among the permutations. For the case $\alpha = 1$, the results of [PPR15, Theorem 1.6] and [EFG17, Corollary 2.5] each imply that four sets suffice in (i). That (ii) holds with three sets is due to [EFG17, Corollary 3.9]. Here is the analogue for general $\alpha > 0$.

Theorem 7. *Let $s_\alpha := \inf\{m: \mathbb{P}[\bigcap_1^m \mathcal{L}(\mathbf{X}^{(i)}(\alpha)) = \{0\}] > 0\}$ be the smallest number of i.i.d. sumset intersections so that the resulting set is trivial with positive probability. Let $h(\alpha)$ be as in (4), and m_α as in Theorem 1. At points where h is continuous we have $s_\alpha = h(\alpha)$, and at the discontinuities $h(\alpha) \leq s_\alpha \leq h(\alpha) + 1$.*

To establish this result for sumsets, we build upon the ideas in [PPR15] and [EFG17] to prove matching lower and upper bounds for s_α . These bounds can be found in Proposition 25 and Proposition 26. Parts of the argument are as simple as swapping 1's for α 's, but in others, particularly the lower bound, some care is required.

A relevant quantity for establishing the upper bound is $p_k = p_k(\alpha)$, the probability the element k appears in $\mathcal{L}(\mathbf{X})$. Estimating $p_k(1)$ is the major probabilistic hurdle in [PPR15]. There are two difficulties. The first is called a lottery effect; certain unlikely events greatly skew p_k . This is circumvented by using *quenched* probabilities \tilde{p}_k , which ignore an $o(1)$ portion of the probability

space (that $\sum X_j$ and $\sum jX_j$ are uncharacteristically large). With this restriction, it is shown in [PPR15, Lemma 2.3] that $\tilde{p}_k(1) \leq k^{\log(2)-1}$. Notice that $4(\log 2 - 1) < -1$ and a Borel–Cantelli type argument implies finiteness of the intersection of the m -independent $\mathcal{L}(\mathbf{X}^{(i)}(\alpha))$. Establishing the quenched formula is the second hurdle. It requires a clever counting and partitioning of $\mathcal{L}(\mathbf{X})$. We generalize these ideas in Lemma 24 to obtain the upper bound

$$\tilde{p}_k(\alpha) \leq C_\varepsilon k^{-1+\alpha \log 2+\varepsilon}$$

for any $\varepsilon > 0$. Taking into account the lottery effect, and ignoring a vanishing portion of the probability space, we can write $\mathbb{P}[k \in \cap_1^m \mathcal{L}(\mathbf{X}^{(i)})] = \prod_1^m \tilde{p}_k$. When $m \geq \lceil (1 - \alpha \log 2 + \varepsilon)^{-1} \rceil$, the product is smaller than n^{-1} , and the probabilities are summable. With this, the Borel–Cantelli lemma implies the intersection is almost surely finite, and thus it is trivial with positive probability.

The lower bound in Proposition 26 requires more care. We generalize the framework in [EFG17, Section 3] to higher dimensions to incorporate more than three sumsets as well as different values of α . Given $I \subseteq \mathbb{N}$ and m sumsets we consider the set of differences

$$S(I; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}) := \{(n_1 - n_m, n_2 - n_m, \dots, n_{m-1} - n_m) : n_i \in \mathcal{L}(\mathbf{X}^{(i)}) \cap I\}.$$

We use Fourier analysis (as in [EFG17] and [MT84]) to study a smoothed version of the indicator function of this set. In particular, we will work with its Fourier transform $F: \mathbb{T}^{m-1} \rightarrow \mathbb{C}$. After obtaining a pointwise bound on F , we integrate around that torus to obtain, for sufficiently large k and $I = (k^{1-\beta}, k]$, that $|S(I, M_1, \dots, M_m)| \gg k^{m-1}$ and lies in the cube $[-ck, ck]^{m-1}$ for some $c > 0$. The density of these sets is sufficiently high that if we look in a larger interval, with positive probability we can find $j_1, \dots, j_m \in (k, Dk]$ such that $X_{j_1}^{(1)}, \dots, X_{j_m}^{(m)} > 0$ and

$$(j_m - j_1, \dots, j_m - j_{m-1}) \in S(I; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}).$$

Combining this with the corresponding point in the set of differences then produces a number in $\mathcal{L}(\mathbf{X}^{(1)}) \cap \dots \cap \mathcal{L}(\mathbf{X}^{(m)})$ made solely from values in $(k^{1-\beta}, Dk]$ for some constants $\beta, D > 0$. By partitioning \mathbb{N} into infinitely many disjoint sets of this form and repeating this argument for each of them, we conclude that the set of intersections is almost surely infinite.

1.5. Further questions. There is still much to be done for both components of this paper—random permutations and random sumsets. Recall that Theorem 1 and Theorem 7 do not pin down the exact behavior at points of discontinuity of h . A glaring question is to characterize m_α and s_α at

these points. We are doubtful our approach generalizes. Characterizing the behavior at the critical values is likely difficult, and will require a new idea.

Question 1. *What are m_α and s_α at discontinuity points of h ?*

For random permutations we would like a more complete characterization than in Corollary 2. In particular an upper bound.

Question 2. *Show that if a permutations has cycle counts that are asymptotically dominated by Poisson random variables with mean $h^{-1}(2)/k$, then two will invariably generate S_n with positive probability.*

Corollary 2 shows that for many permutation measures no finite number will invariably generate S_n . Possibly if we let the number of permutations grow with n we will see different behavior.

Question 3. *Let $\alpha > 1/\log 2$. Prove that there are constants $\beta < \beta'$ such that the probability $\beta \log n$ Ewens- α permutations generate S_n is bounded away from 0, while $\beta' \log n$ fail to generate S_n with high probability?*

We believe that $\log n$ is the correct order by the following approximation model. When $\alpha > 1/\log 2$ the fixed set sizes in a random permutation are dense in $[n]$. So, we can model the occurrence of each fixed size by independent Bernoulli random variables with some parameter $p_\alpha > 0$. If one performs N independent thinnings of $[n]$ by including each integer k with probability p_α , then the probability k belongs to all N subsets is p_α^N . Hence, the probability there are no common elements in the N -fold intersection is $(1 - p_\alpha^N)^n$. If we choose $N = c \log n$ this will converge to either 0 or a positive number as we increase c . The difficulty in answering the question for permutations is we do not have independence for fixing different fixed set sizes.

We also have a question in the simpler setting of generating S_n . Dixon's theorem [Dix92] implies that two uniformly random permutations will generate either A_n or S_n with high probability. Does the analogue hold for $\text{ESF}(\alpha, n)$ permutations? Because generation relies on more than just cycle structure, this may be a hard question.

Question 4. *How many $\text{ESF}(\alpha, n)$ permutations are needed to generate one of A_n or S_n with high probability?*

1.6. **Notation.** Note that $\alpha > 0$ is fixed throughout. As already discussed, we denote cycle counts of a permutation by $\mathbf{C} = (C_1, \dots, C_n)$ with C_j the number of j -cycles. Unless otherwise noted, our permutations and cycle vectors come from an $\text{ESF}(\alpha, n)$ distribution. We will let $\mathbf{Y}(\alpha)$ denote a vector of independent $\text{Poi}(\alpha/k)$ random variables that are obtained from the Feller coupling in (10). We will also take $\mathbf{X}(\alpha)$ to be a vector of independent $\text{Poi}(\alpha/k)$ random variables. This is done to distinguish the permutation and random sumset settings. Henceforth we will suppress the α dependence unless there is reason to call attention to it. For an infinite vector, such as \mathbf{Y} , we will use the notation $\mathbf{Y}[i, j]$ to denote the sub-vector $(Y_i, Y_{i+1}, \dots, Y_j)$.

For I an interval in \mathbb{N} define the sumset

$$\mathcal{L}(I, \mathbf{X}) := \left\{ \sum_{j \in I} jx_j : 0 \leq x_j \leq X_j \right\},$$

and set $\mathcal{L}(\mathbf{X}) := \mathcal{L}(\mathbb{N}, \mathbf{X})$. Given \mathbf{X} define

$$(5) \quad f_{\ell, k}(\mathbf{X}) = \sum_{\ell < j \leq k} X_j, \quad g_{\ell, k}(\mathbf{X}) = \sum_{\ell < j \leq k} jX_j.$$

It will be convenient to abbreviate $f_k := f_{1, k}(\mathbf{X})$ and $g_k := g_{1, k}(\mathbf{X})$. Thus, f_k is the number of available summands smaller than k , and g_k is the largest sum attainable with them.

We will use the notation $f \ll g$ and $f(n) = O(g(n))$ interchangeably for the existence of c such that for all large enough n it holds that $f(n) \leq cg(n)$. When the c depends on our choice of ε we will use $f \ll_{\varepsilon} g$, similarly $f \ll_{m, \varepsilon} g$ denotes dependence of the constant on both ε and the number of intersections m . We use the little- o notation $f(n) = o(g(n))$ to mean that $f(n)/g(n) \rightarrow 0$. Also, $f(n) = \tilde{O}(g(n))$ means there is $k > 0$ such that $f(n) = O(\log^k(n)g(n))$. The symbol $f(n) \approx g(n)$ means that $f(n)/g(n) \rightarrow c$ for some $c > 0$. A sequence of events E_n occurs *with high probability* if $\liminf \mathbb{P}[E_n] = 1$. Also $i \mid j$ means that i divides j (i.e. there is an integer k with $ik = j$.)

1.7. **Outline of paper.** We will lead off with a proof of Theorem 1. This comes in two parts: an upper bound from Theorem 7, and a matching lower bound in Proposition 10. The upper bound in Theorem 7 has two distinct components. We connect random sumsets to $\text{ESF}(\alpha, n)$ permutations in Section 3.1. This section includes the proofs of Theorem 3, Theorem 5, and Theorem 6. The second component is analyzing the random sumset model. We do this in Section 4 and provide an upper bound on s_{α} at Proposition 25. The lower bound is Proposition 10 and proven in Section 5.

1.8. **Acknowledgements.** We would like to thank Robin Pemantle for several helpful conversations. We are very grateful for Richard Arratia's correspondence. He made us aware of the indispensable Feller coupling. Also, thanks to Clayton Barnes, Emily Dinan, and Jacob Richey for their assistance in the formative stages of this project.

2. PROOF OF THEOREM 1

First we show that $h(\alpha) + 1$ is an upper bound for the number of $\text{ESF}(\alpha, n)$ permutations to invariably generate S_n . We will use a few lemmas and propositions that come later in the paper, but prove this now to give the reader a sense of the important bounds needed.

Proposition 8. *Suppose α is a continuity point of h . A collection of $h(\alpha) < \infty$ independent permutations sampled according to an $\text{ESF}(\alpha, n)$ invariably generates S_n with positive probability as $n \rightarrow \infty$.*

Proof. Let $m = h(\alpha)$. Call the permutations $\pi^{(1)}, \dots, \pi^{(m)}$. Condition $\pi^{(1)}$ to be odd. Lemma 21 guarantees this happens with probability bounded away from zero. Define the events

$$E_{n,k} = \{\pi^{(1)}, \dots, \pi^{(m)} \text{ fix an } \ell\text{-set for some } k \leq \ell \leq n/2\},$$

$$F_{n,k} = \{\pi^{(1)}, \dots, \pi^{(m)} \text{ fix an } \ell\text{-set for some } 1 \leq \ell \leq k\}.$$

The probability of a common fixed set size amongst all of the permutations is bounded above by

$$(6) \quad \mathbb{P}[E_{n,k} \cup F_{n,k}] \leq \mathbb{P}[E_{n,k}] + \mathbb{P}[F_{n,k}].$$

Therefore, the $\pi^{(i)}$ will invariably generate a transitive subgroup of S_n if this quantity is less than 1. By Theorem 3 we know this transitive subgroup is with high probability either A_n or S_n . Because $\pi^{(1)}$ is odd, it must be S_n .

Now α is a continuity point of h , so

$$m = h(\alpha) = \lceil (1 - \alpha \log 2)^{-1} \rceil > (1 - \alpha \log 2)^{-1}.$$

It follows that there exists some $\varepsilon > 0$ such that

$$m = (1 - (\alpha + \varepsilon) \log 2)^{-1}.$$

Fix ε' such that $0 < \varepsilon' < \varepsilon$. By Lemma 20 there exists a c such that for any k and n and $i = 2, \dots, m$,

$$\mathbb{P}[\pi^{(i)} \text{ fixes a } k\text{-set}] \leq ck^{-\delta},$$

where

$$\delta = 1 - (\alpha + \varepsilon') \log 2.$$

Similarly, by Corollary 23 for any k ,

$$\mathbb{P}[\pi^{(1)} \text{ fixes a } k\text{-set}] \leq ck^{-\delta} \cdot \max \left\{ \frac{\alpha + 1}{\alpha}, \frac{\alpha + 1}{1} \right\}.$$

By independence of the $\pi^{(i)}$, we conclude

$$\mathbb{P}[\pi^{(1)}, \dots, \pi^{(m)} \text{ all fix a } k\text{-set}] \leq (ck^{-\delta})^m \cdot \max \left\{ \frac{\alpha + 1}{\alpha}, \frac{\alpha + 1}{1} \right\}.$$

By construction we have that $\delta m > 1$. A union bound for $k_0 \leq k \leq n/2$ ensures that,

$$(7) \quad \mathbb{P}[E_{n,k_0}] \leq Ck_0^{1-\delta m}.$$

As $\delta m > 1$ it follows from (7) that we can make $\mathbb{P}[E_{n,k}]$ as small as we like by fixing a large k .

Now we turn our attention to bounding the probability of $F_{n,k}$. We will show that there exists $\beta > 0$ such that

$$(8) \quad \inf_{k \geq 1} \liminf_{n \rightarrow \infty} \mathbb{P}[F_{n,k}] < 1 - \beta.$$

Once we have (8) we apply (7) with k_0 large enough so that $\mathbb{P}[E_{n,k_0}] < \beta/2$. This then gives (6) is strictly less than 1 as $n \rightarrow \infty$, as desired.

It remains to establish (8). Fix k and let π be an $\text{ESF}(n, \alpha)$. The Feller coupling at (10) ensures that each cycle count $C_\ell \leq Y_\ell + \mathbf{1}\{J_n = 0\}$ for all $1 \leq \ell \leq n$ for a certain random variable J_n (defined just above (10)). Thus, with high probability we have

$$C_\ell \leq Y_\ell \text{ for all } 1 \leq \ell \leq k.$$

Since this dominance relation is with high probability, it also applies to $\pi^{(1)}$ despite being conditioned to be odd. This is because $\{\pi^{(1)} \text{ is odd}\}$ occurs with positive probability by Lemma 21.

The event $F_{n,k}^c$ can be coupled so that it occurs when m independent Poisson sumsets have empty intersection. Theorem 7 ensures that this occurs with some positive probability which is independent of k (since we are considering infinite sets). This probability provides the claimed β in (8) and completes the argument. \square

Remark 9. The argument above implies that $m_\alpha \leq m_{\alpha+\eta}$ for all $\eta > 0$, since any appropriate choice of ε' in the latter case will hold for the former. It follows that $h(\alpha) + 1$ permutations will invariably generate S_n with positive probability at discontinuity points of h .

Now we prove a matching lower bound.

Proposition 10. *A collection of $m < h(\alpha)$ permutations sampled independently according to an $ESF(\alpha, n)$ with high probability do not invariably generate S_n , or any transitive subgroup of S_n .*

Proof. Let $m < h(\alpha)$, and fix some $\varepsilon > 0$. By Proposition 26 the intersection of the i.i.d. sumsets with parameter α , $\bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)})$, is almost surely infinite. It follows that there is some $k_0 = k_0(\varepsilon)$ such that if $k \geq k_0$ then $\bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)}) \cap [1, k] \neq \emptyset$ with probability at least $1 - \varepsilon/2$.

Fix some $k \geq k_0$ and call the random permutations $\pi^{(1)}, \dots, \pi^{(m)}$. By [ABT00, Theorem 3.1] the cycle counts $\mathbf{C}^{(i)}[1, k]$ of $\pi^{(i)}$ converge in total variation to $\mathbf{X}^{(i)} \cap [1, k]$. Then there exists some $n_0 = n_0(\varepsilon, k)$ such that if $n \geq n_0$, with probability greater than $1 - \varepsilon/2$ we have $\mathbf{C}^{(i)}[1, k] = \mathbf{X}^{(i)} \cap [1, k]$ for all $i = 1, \dots, m$. Since

$$\bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)}) \cap [1, k] \subset \bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)} \cap [1, k]),$$

with probability at least $1 - \varepsilon$ we have that $\pi^{(1)}, \dots, \pi^{(m)}$ each fix a set of size ℓ for some $\ell \leq k$. As ε was arbitrary, this completes our proof. \square

Theorem 1 is a straightforward consequence of the last two propositions.

Proof of Theorem 1. Together [ABT00, Theorem 3.1, Theorem 3.2] guarantee that the cycle counts of any measure satisfying the strong α -logarithmic property converge in total variation to the same distribution as an $ESF(\alpha, n)$ permutation. Hence it suffices to consider $ESF(\alpha, n)$ permutations.

By Proposition 8 we have $m_\alpha \leq h(\alpha)$ at point of continuity of h . At points of discontinuity, we use Remark 9 to deduce that $m_\alpha \leq m_{\alpha+\eta} \leq h(\alpha) + 1$. Then, by Proposition 10 we have that $h(\alpha) \leq m_\alpha$ for all α , completing our proof. \square

3. THE UPPER BOUND FOR m_α

The main goal of this section is to prove that $m_\alpha \leq s_\alpha$. Our primary tool for relating Poisson sumsets to permutations is the Feller coupling. This is described at the onset of Section 3.1. We then use it in that section to prove several estimates on the cycle structure of $ESF(\alpha, n)$ permutations. With these we prove Theorem 3, Theorem 5, and Theorem 6 in Section 3.2. We put all of this together and establish Proposition 8 in Section 3.3.

3.1. The Feller Coupling. The Ewens sampling formula can be obtained via an elegant coupling attributed to Feller [Fel45, p. 815] for uniform permutations. The articles [ABT92, ABT16] have nice descriptions. We start with a sequence ξ_1, ξ_2, \dots of mixed Bernoulli random variables with $\mathbb{P}[\xi_i = 1] = \frac{\alpha}{\alpha+i-1}$ and $\mathbb{P}[\xi_i = 0] = \frac{i-1}{\alpha+i-1}$. Define an ℓ -spacing to be any occurrence of $\ell - 1$ zeros enclosed by a 1 on the left and right (e.g. 1, 0, 0, 0, 1 is a 4 spacing). Remarkably, the total number of ℓ -spacings, Y_ℓ , is distributed as $\text{Poi}(\alpha/\ell)$. Moreover, the collection Y_1, Y_2, \dots are independent (see [ABT16, ABT92]).

The counts of spacings in $\xi_1, \dots, \xi_n, 1$ generate a partition of n , which can be filled in uniformly randomly to form a permutation with C_ℓ cycles of size ℓ , where C_ℓ are sampled according to an $ESF(\alpha, n)$ distribution. Explicitly,

$$\begin{aligned} C_\ell &= \text{the number of } \ell\text{-spacings in } \xi_1, \dots, \xi_n, 1 \\ (9) \quad &= \xi_{n-\ell+1}(1 - \xi_{n-\ell+2}) \cdots (1 - \xi_n) + \sum_{k=1}^{n-\ell} \xi_k(1 - \xi_{k+1}) \cdots (1 - \xi_{k+\ell-1})\xi_{k+\ell}. \end{aligned}$$

This gives an intuition for why $ESF(1, n)$ corresponds to a uniformly random permutation. Indeed, we can inductively construct such a permutation by letting ξ_i indicate the decision to complete a cycle, when there is an i -way choice for the next element.

Letting R_n be the index of the rightmost 1 in ξ_1, \dots, ξ_n and $J_n = n + 1 - R_n$ it follows from the previous discussion that

$$(10) \quad C_\ell \leq Y_\ell + \mathbf{1}\{J_n = \ell\}.$$

This says that the cycle counts can be obtained from independent Poisson random variables through a random number of deletions, and possibly one insertion. We require a few estimates on the behavior of these perturbations.

Lemma 11. *Let J_n be as defined above, and $D_n = \sum_{\ell=1}^n Y_\ell - \mathbf{1}\{J_n = \ell\} - C_\ell$ be the number of deletions. The following four inequalities hold:*

$$(11) \quad \mathbb{P}[J_n = \ell] \leq \frac{\alpha}{n - \ell}, \quad \ell < n.$$

$$(12) \quad \mathbb{P}[J_n = \ell] \leq cn^{-\alpha(1-\gamma)}, \quad n - n^\gamma < \ell < n, \quad 0 < \gamma < 1, \quad \text{for some } c = c(\gamma) > 0.$$

$$(13) \quad \mathbb{P}[J_n = \ell] = O(n^{-\alpha/(1+\alpha)}), \quad \ell < n.$$

$$(14) \quad \mathbb{E}D_n = O(1).$$

Proof. We can write the density of J_n explicitly for all $\ell < n$ as

$$(15) \quad \mathbb{P}[J_n = \ell] = \frac{\alpha}{\alpha + n - \ell} \prod_{n-\ell+2}^n \frac{i-1}{\alpha+i-1} \leq \prod_{n-\ell+1}^n \frac{i-1}{\alpha+i-1}.$$

As $\alpha > 0$ we can use the leading term $\frac{\alpha}{\alpha+n-\ell} < \frac{\alpha}{n-\ell}$ to obtain (11).

Fix some $\gamma \in (0, 1)$. For the case $\ell > n - n^\gamma$ in (12) we use that when $J_n = \ell$ we must have $\xi_{n-\ell+1} = 1$ and $\xi_k = 0$ for all larger index terms up to ξ_n . The probability of this can be computed explicitly as

$$\mathbb{P}[J_n = \ell] = \frac{\alpha}{\alpha + n - \ell} \prod_{n-\ell+2}^n \frac{i-1}{\alpha+i-1} \leq \prod_{n-\ell+1}^n \frac{i-1}{\alpha+i-1}.$$

We estimate the product by converting it to a sum

$$\prod_{n-\ell+1}^n \frac{i-1}{\alpha+i-1} = \exp\left(-\sum_{n-\ell+1}^n \log \frac{\alpha+i-1}{i-1}\right) \leq \exp\left(-\int_{n-\ell+1}^n \log \frac{\alpha+x-1}{x-1} dx\right).$$

The last inequality follows from the fact that $\log \frac{\alpha+x-1}{x-1}$ is a decreasing function, and the sum can be viewed as a right-sided Riemann (over-)approximation. This integral has a closed form

$$-\int_{n-\ell+1}^n \log \left(\frac{\alpha+x-1}{x-1}\right) dx = -\log \left(\frac{(\alpha+n-1)^{\alpha+n-1}}{(n-1)^{n-1}}\right) + \log \left(\frac{(\alpha+n-\ell)^{\alpha+n-\ell}}{(n-\ell)^{n-\ell}}\right).$$

Hence, exponentiating the above line gives

$$(16) \quad \prod_{n-\ell+1}^n \frac{i-1}{\alpha+i-1} \leq \frac{(n-1)^{n-1}}{(\alpha+n-1)^{\alpha+n-1}} \times \frac{(\alpha+n-\ell)^{\alpha+n-\ell}}{(n-\ell-1)^{n-\ell}}.$$

The leftmost product is less than $n^{-\alpha}$. The rightmost product can be written as

$$(17) \quad (\alpha + n - \ell)^\alpha \left(\frac{n - \ell + \alpha}{n - \ell - 1} \right)^{n - \ell} = (\alpha + n - \ell)^\alpha \left(\frac{1 + \frac{\alpha}{n - \ell}}{1 - \frac{1}{n - \ell}} \right)^{n - \ell}.$$

The rightmost product in (17) is asymptotic to e^α/e^{-1} as $n - \ell \rightarrow \infty$, and thus may be bounded by some constant for all values of $n - \ell$. Using the hypothesis that $n - \ell \leq n^\gamma$, it follows that (17) is universally bounded by $cn^{\alpha\gamma}$ for some constant c . We conclude that (16) is less than $cn^{-\alpha+\alpha\gamma} = cn^{-\alpha(1-\gamma)}$. This establishes (12).

To obtain the universal bound at (13), observe when $\ell \leq n - n^{\alpha/(1+\alpha)}$ we can input this into (11) to obtain

$$\mathbb{P}[J_n = \ell] \leq \frac{\alpha}{n - (n - n^{\alpha/(1+\alpha)})} = O(n^{-\alpha/(1+\alpha)}).$$

And, for $\ell > n - n^\alpha$ we substitute $\gamma = \alpha/(1 + \alpha)$ into (12) to obtain the same asymptotic inequality: $O(n^{-\alpha/(1+\alpha)})$.

Lastly, (14) is proven in [ABT92, Theorem 2]. \square

We will also require a bound on the joint probability that two cycles occur simultaneously. There will be some poly-log terms that, besides being a minor nuisance, do not effect the tack of our proof. Recall the \tilde{O} notation that ignores logarithmic contribution. It is described carefully in Section 1.6.

Lemma 12. *For any $i < j \leq n$ it holds that*

$$\mathbb{P}[C_i C_j > 0] = \tilde{O}(i^{-1} j^{-\alpha/(1+\alpha)}).$$

Proof. We start by using the formula at (9) to bound C_i with the ξ_k ,

$$C_i \leq \xi_{n-i+1}(1 - \xi_{n-i+2}) \cdots (1 - \xi_n) + \sum_{k=1}^{n-i} \xi_k \xi_{k+i}.$$

Thus,

$$\begin{aligned} C_i C_j &\leq \left(\xi_{n-i+1}(1 - \xi_{n-i+2}) \cdots (1 - \xi_n) + \sum_{k=1}^{n-i} \xi_k \xi_{k+i} \right) \\ &\quad \times \left(\xi_{n-j+1}(1 - \xi_{n-j+2}) \cdots (1 - \xi_n) + \sum_{\ell=1}^{n-j} \xi_\ell \xi_{\ell+j} \right) \\ &= (x + y)(z + w). \end{aligned}$$

The last line is just labeling the terms in the product immediately above in the natural way (i.e. $y = \sum_{k=1}^{n-i} \xi_k \xi_{k+i}$). This is so we can bound $\mathbb{E}[xz + xw + yw + yz]$ in an organized fashion, one term at a time.

Term xz : First, consider $\xi_{n-i+1}(1 - \xi_{n-i+2}) \cdots (1 - \xi_n) \xi_{n-j+1}(1 - \xi_{n-j+2}) \cdots (1 - \xi_n)$. When $i < j$ this product contains the term $\xi_{n-i+1}(1 - \xi_{n-i+1}) \equiv 0$, thus it is always zero. We conclude that

$$(18) \quad \mathbb{E}[xz] \equiv 0.$$

Term xw : Next, we consider the cross-term $\xi_{n-i+1}(1 - \xi_{n-i+2}) \cdots (1 - \xi_n) \sum_{\ell=1}^{n-j} \xi_\ell \xi_{\ell+j}$. Many of the summands are zero. In fact, the overlapping terms with $n - j - i + 1 < \ell \leq n - j$ are identically zero. So, for the sake of obtaining disjoint terms, we consider the smaller sum

$$\xi_{n-i+1}(1 - \xi_{n-i+2}) \cdots (1 - \xi_n) \sum_{\ell=1}^{n-j-i+1} \xi_\ell \xi_{\ell+j}.$$

Aside from the $\ell = n - j - i + 1$ term, all the indices are all disjoint. Ignoring this term for now, the expected value is

$$(19) \quad \frac{\alpha}{\alpha + n - i} \frac{n - i + 1}{\alpha + n - i + 1} \cdots \frac{n - 1}{\alpha + n - 1} \sum_{\ell=1}^{n-j-i} \frac{\alpha^2}{(\alpha + \ell - 1)(\alpha + \ell - 1 + j)}.$$

Now i and j correspond to cycle lengths, so we must have $i + j \leq n$ to have a nonzero expectation at all. Since $i < j$ we conclude that $i < n/2$. A simple bound on the product outside the sum is to plug this value into the first term and ignore the rest. This gives the bound

$$\frac{\alpha}{\alpha + n - i} \frac{n - i + 1}{\alpha + n - i + 1} \cdots \frac{n - 1}{\alpha + n - 1} < \frac{\alpha}{n - (n/2) + 1} < \frac{2\alpha}{n}.$$

Ignoring the product outside of it, the sum can be compared to the integral $\int_1^n \frac{1}{x(x+j)} dx$ to prove it is $O(\log(j)/j)$. Thus the expression in (19) is $\tilde{O}(1/nj)$.

Rather than using the rough bound from above, the full $\ell = n - j - i + 1$ term is

$$\xi_{n-j-i+1}(1 - \xi_{n-j-i+2}) \cdots (1 - \xi_{n-i}) \xi_{n-i+1}^2 (1 - \xi_{n-i+2}) \cdots (1 - \xi_n).$$

The expectation of this term is

$$\frac{\alpha}{\alpha + n - j - i} \left(\prod_{k=n-j-i+2}^{n-i} \frac{k-1}{\alpha + k - 1} \right) \frac{\alpha}{\alpha + n - i} \left(\prod_{k=n-i+2}^n \frac{k-1}{\alpha + k - 1} \right).$$

But this is just

$$\begin{aligned} \frac{\alpha}{\alpha+n-i} \frac{\alpha+n-i}{n-i} \mathbb{P}[J_n = j+i] &= \frac{\alpha}{n-i} \mathbb{P}[J_n = j+i] \\ &\leq \frac{2\alpha}{n} \mathbb{P}[J_n = j+i]. \end{aligned}$$

Then by Lemma 11 (13) we have the bound

$$\mathbb{P}[J_n = j+i] = O(n^{-\alpha/(1+\alpha)}).$$

Thus this term is $O(i^{-1}j^{-\alpha/(1+\alpha)})$, and

$$(20) \quad \mathbb{E}[xw] = \tilde{O}(i^{-1}j^{-\alpha/(1+\alpha)}).$$

Term yw : Because $i < j$ the term $\xi_{n-j+1}(1 - \xi_{n-j+2}) \cdots (1 - \xi_n) \sum_{\ell=1}^{n-i} \xi_\ell \xi_{\ell+i}$ behaves differently than the case just prior. In fact, it is the largest order term amongst the four. The first step is the same as before though. A few of the terms are identically zero. So, we can write yw as

$$\xi_{n-j+1}(1 - \xi_{n-j+2}) \cdots (1 - \xi_n) \sum_{\ell=1}^{n-i-j+1} \xi_\ell \xi_{\ell+i}.$$

Except for when $\ell = n - i - j + 1$, all of the indices now disjoint. When we take expectation of the terms with disjoint indices we obtain

$$\frac{\alpha}{\alpha+n-j} \frac{n-j+1}{\alpha+n-j+1} \cdots \frac{n-1}{\alpha+n-1} \sum_{\ell=1}^{n-j-i} \frac{\alpha^2}{(\alpha+\ell-1)(\alpha+\ell+i-1)}.$$

Notice that j could be quite large. The best bound we have for the product outside the sum comes from (13). This implies it is $O(n^{-\alpha/(1+\alpha)})$. As with (19), the sum term is $O(\log i/i)$. This yields that the whole expression is $\tilde{O}(i^{-1}j^{-\alpha/(1+\alpha)})$.

As before, the full $\ell = n - i - j + 1$ term is

$$\xi_{n-i-j+1}(1 - \xi_{n-j-i+2}) \cdots (1 - \xi_{n-j}) \xi_{n-j+1}^2 (1 - \xi_{n-i+2}) \cdots (1 - \xi_n).$$

The expectation of this term is

$$\frac{\alpha}{\alpha+n-j-i} \left(\prod_{k=n-j-i+2}^{n-j} \frac{k-1}{\alpha+k-1} \right) \frac{\alpha}{\alpha+n-i} \left(\prod_{k=n-j+2}^n \frac{k-1}{\alpha+k-1} \right).$$

This simplifies to

$$\frac{\alpha}{\alpha+n-j} \frac{\alpha+n-j}{n-j} \mathbb{P}[J_n = j+i] = \frac{\alpha}{n-j} \mathbb{P}[J_n = j+i].$$

By construction $i \leq n - j$, and again by Lemma 11 (13), we have that $\mathbb{P}[J_n = j + i] = O(n^{-\alpha/(1+\alpha)})$. Thus this term has expectation that is $O(i^{-1}j^{-\alpha/(1+\alpha)})$. It follows that

$$(21) \quad \mathbb{E}[yw] = \tilde{O}(i^{-1}n^{-\alpha/(1+\alpha)}).$$

Term yz : The last step is to bound the product

$$(22) \quad \left(\sum_{k=1}^{n-i} \xi_k \xi_{k+i} \right) \left(\sum_{\ell=1}^{n-j} \xi_\ell \xi_{\ell+j} \right) = \sum_{k,\ell=1}^{n-i,n-j} \xi_k \xi_{k+i} \xi_\ell \xi_{\ell+j}.$$

Call the index set $\Lambda = \{(k, \ell) : k \in [1, \dots, n - i], \ell \in [1, \dots, n - j]\}$. As $i < j$, we can divide Λ into disjoint sets

$$\Lambda_0 = \{k, k + i, \ell, \text{ and } \ell + j \text{ are all distinct}\},$$

$$\Lambda_1 = \{k = \ell\},$$

$$\Lambda_2 = \{k = \ell + j\},$$

$$\Lambda_3 = \{k + i = \ell\},$$

$$\Lambda_4 = \{k + i = \ell + j\}.$$

This yields the following upper bound on (22):

$$\sum_{\Lambda_0} \xi_k \xi_{k+i} \xi_\ell \xi_{\ell+j} + \sum_{\Lambda_1} \xi_k \xi_{k+i} \xi_{\ell+j} + \sum_{\Lambda_2} \xi_k \xi_{k+i} \xi_\ell + \sum_{\Lambda_3} \xi_k \xi_{k+i} \xi_{\ell+j} + \sum_{\Lambda_4} \xi_k \xi_{k+i} \xi_\ell.$$

The payoff of this partition is that each product of ξ above is of distinct, independent terms. This lets us tidily compute the expectation of each. For $(k, \ell) \in \Lambda_0$ we have the $k, k + i, \ell, \ell + j$ are all distinct. This gives

$$\mathbb{E}[\xi_k \xi_{k+i} \xi_\ell \xi_{\ell+j}] \leq \frac{A}{k(k+i)\ell(\ell+j)}$$

for some constant A depending only on α . We can take the expectation of the entire sum

$$\mathbb{E} \sum_{\Lambda_0} \xi_k \xi_{k+i} \xi_\ell \xi_{\ell+j} \leq \sum_{k=1, \ell=1}^n \mathbb{E}[\xi_k \xi_{k+i} \xi_{k+j}] \leq \sum_{k=1, \ell=1}^n \frac{A}{k(k+i)\ell(\ell+j)} = \tilde{O}(1/ij).$$

We can parametrize Λ_1 by k to write it as $\Lambda_1 = \{(k, k) : k = 1, \dots, n - j\}$, this gives the bound

$$\mathbb{E} \sum_{\Lambda_1} \xi_k \xi_{k+i} \xi_{\ell+j} \leq \sum_{k=1}^n \mathbb{E}[\xi_k \xi_{k+i} \xi_{k+j}] \leq \sum_{k=1}^n \frac{A}{k(k+i)(k+j)} = \tilde{O}(1/ij).$$

We can similarly parametrize the sums over Λ_2, Λ_3 and Λ_4 to obtain

$$(23) \quad \mathbb{E}[yz] = \tilde{O}(1/ij)$$

All of the expressions at (18), (20), (21), and (23) are $\tilde{O}(i^{-1}j^{-\alpha/(1+\alpha)})$. Markov's inequality yields

$$(24) \quad \mathbb{P}[C_i C_j > 0] = \tilde{O}(i^{-1}j^{-\alpha/(1+\alpha)}).$$

This proves the claimed inequality. \square

We also need a bound on the probability that cycles of the same size occur.

Lemma 13. *For any $i \leq n$ it holds that*

$$\mathbb{P}[C_i \geq 2] = \tilde{O}(i^{-2} + n^{-1-\alpha/(1+\alpha)}).$$

Proof. We start by bounding $\mathbb{P}[C_i > 2]$. By (10) we have

$$(25) \quad \mathbb{P}[Y_i + \mathbf{1}\{J_n = i\} > 2] \leq \mathbb{P}[Y_i > 1] = O(1/i^2).$$

It remains to bound $\mathbb{P}[C_i = 2]$. Conditioning on the value of J_n , we obtain the decomposition

$$(26) \quad \mathbb{P}[C_i = 2] = \mathbb{P}[C_i = 2 \mid J_n = i]\mathbb{P}[J_n = i] + \mathbb{P}[C_i = 2 \mid J_n \neq i]\mathbb{P}[J_n \neq i].$$

First we estimate the first summand on the right hand side of (26). Because two cycles larger than $n/2$ cannot occur, we must have $i \leq n/2$. Then by Lemma 11 (11), $\mathbb{P}[J_n = i] \leq 2\alpha/n$. When $J_n = i$, it ensures that the sequence ξ_1, \dots, ξ_n has $\xi_{n-i+1} = 1$ and the larger index ξ_k are zero. The smaller index ξ_i are independent of this event. The value of C_i conditioned on $J_n = i$ is thus one more than the number of i -spacings in the sequence

$$\xi_1, \dots, \xi_{n-i}, 1,$$

with ξ_1, \dots, ξ_{n-i} independent and distributed according the Feller coupling of an $ESF(\alpha, n-i)$ distribution. Call the number of i -spacings in the above sequence C_i^- . So, conditional on $J_n = i$ we have $C_i = C_i^- + 1$. By writing explicitly what it means for an i -spacing to arise, we can bound C_i^- in terms of the ξ_k :

$$C_i^- \leq \xi_{n-2i+1}(1 - \xi_{n-2i+2}) \cdots (1 - \xi_{n-i}) + \sum_{k=1}^{n-2i} \xi_k \xi_{k+i}.$$

Looking to apply Markov's inequality, we take the expectation of the above line and obtain

$$\mathbb{E}C_i^- \leq \frac{\alpha}{\alpha + n - 2i} \frac{n - 2i + 1}{\alpha + n - 2i + 1} \cdots \frac{n - i - 1}{\alpha + n - i - 1} + \sum_{k=1}^{n-2i} \frac{\alpha^2}{(\alpha + k - 1)(\alpha + k - 1 + i)}.$$

The product above is equivalent to $\mathbb{P}[J_{n-i} = i]$, then we use Lemma 11 (13) and the fact that $i < n/2$ to bound it by $\tilde{O}((n-i)^{-\alpha/(1+\alpha)}) = \tilde{O}(n^{-\alpha/(1+\alpha)})$. The sum, as we have seen from (19), has order $\tilde{O}(1/i)$. Recalling that (11) implies $\mathbb{P}[J_n = i] = O(1/n)$, it follows that

$$(27) \quad \mathbb{P}[C_i = 2 \mid J_n = i] \mathbb{P}[J_n = i] = \tilde{O}(n^{-1-\alpha/(1+\alpha)} + 1/in) = \tilde{O}(i^{-2} + n^{-1-\alpha/(1+\alpha)}).$$

Now we estimate the second summand on the right hand side of (26). Under the event $\{J_n \neq i\}$, for C_i to equal 2, it is necessary that two i -spacings occur in the infinite sequence ξ_1, ξ_2, \dots . The count of i -spacings is distributed as Y_i from (10). Therefore,

$$(28) \quad \mathbb{P}[C_i = 2 \text{ and } J_n \neq i] \leq \mathbb{P}[Y_i \geq 2] = O(1/i^2).$$

Then (25), (27), and (28) imply that

$$(29) \quad \mathbb{P}[C_i \geq 2] = \tilde{O}(i^{-2} + n^{-1-\alpha/(1+\alpha)})$$

as desired. \square

Using the previous two lemmas we may prove the following.

Proposition 14. *Let $\pi \sim ESF(\alpha, n)$ and $a = 1 - \frac{\alpha}{4\alpha+3}$. The probability that π has two cycles whose lengths have a common divisor larger than n^a is $o(1)$.*

Proof. Call E_a the event that π has two cycles with common divisor larger than n^a . We utilize a union bound, then apply Lemma 12 and Lemma 13 to write

$$\begin{aligned} \mathbb{P}[E_a] &\leq \sum_{d=\lceil n^a \rceil}^n \left(\sum_{1 \leq i < j}^{\lfloor n^{1-a} \rfloor} \mathbb{P}[C_{id} C_{jd} > 0] + \sum_{i=1}^{\lfloor n^{1-a} \rfloor} \mathbb{P}[C_{id} \geq 2] \right) \\ &= \tilde{O} \left(\sum_{d=\lceil n^a \rceil}^n \sum_{1 \leq i < j}^{\lfloor n^{1-a} \rfloor} (id)^{-1} (jd)^{-\alpha/(1+\alpha)} + \sum_{d=\lceil n^a \rceil}^n \sum_{i=1}^{\lfloor n^{1-a} \rfloor} \left(\frac{1}{(id)^2} + n^{-1-\alpha/(1+\alpha)} \right) \right). \end{aligned}$$

Since $d \geq n^a$ the above line is

$$\tilde{O} \left(n^1 n^{2(1-a)} n^{-a-a\alpha/(1+\alpha)} + n^1 n^{1-a} n^{-2a} + n^{1-a-\alpha/(1+\alpha)} \right).$$

We can simplify the exponents in each term to

$$3 - 3a - a\alpha/(1 + \alpha), \quad 2 - 3a, \quad \text{and} \quad 1 - a - \alpha/(1 + \alpha),$$

respectively. Some algebra shows that all are negative as long as $a > 1 - \frac{\alpha}{4\alpha+3}$. For such a we have $\mathbb{P}[E_a] = o(1)$. \square

3.2. Proofs of Theorem 3, Theorem 5, and Theorem 6. We now have enough to establish our bound on the minimal degree of an $\text{ESF}(\alpha, n)$ permutation.

Lemma 15. *Let $\pi \sim \text{ESF}(\alpha, n)$. For each $0 < \beta < 1$, there exists $\lambda \in (0, 1)$ such that if we define the set*

$$Q_n = \{d < n^\beta : \text{there exists a prime } p > n^{\lambda\beta} \text{ with } p \mid d\},$$

then

$$(30) \quad \mathbb{P}[\exists C_d = 1 \text{ with } d \in Q_n] = 1 - o(1).$$

Proof. We start with [ABT00, Theorem 3.1] which implies

$$(31) \quad \sum_{d \in Q_n} C_d \xrightarrow{TV} \text{Poi}(\kappa_n),$$

where

$$\kappa_n = \sum_{d \in Q_n} \frac{\alpha}{d}.$$

Borrowing an argument from [Bov80, p. 50] we have

$$\begin{aligned} \sum_{d \in Q_n} \frac{1}{d} &= \sum_{d < n^\beta} \frac{1}{d} - \sum_{\substack{d < n^\beta \\ p \mid d \Rightarrow p \leq n^{\lambda\beta}}} \frac{1}{d} \\ &\geq \sum_{d < n^\beta} \frac{1}{d} - \prod_{p \leq n^{\lambda\beta}} (1 - 1/p)^{-1} \\ &= \beta \log n - e^{\gamma^*} \lambda \beta \log n + O(1) \end{aligned}$$

where γ^* is Euler's constant. When λ is such that $e^{\gamma^*} \lambda \beta < \beta/2$ we have $\kappa_n > \frac{\beta}{2} \log n$ for all n . It follows from (31) and standard estimates on the concentration of a Poisson random variable that

$$\mathbb{P}\left[\sum_{d \in Q_n} C_d > (\beta/4) \log n\right] \rightarrow \mathbb{P}[\text{Poi}(\kappa_n) > (\beta/4) \log n] = 1 - o(1).$$

To obtain (30), observe that the convergence statement in [ABT00, Theorem 3.1] implies that $\mathbb{P}[C_d > 1] \rightarrow \mathbb{P}[Y_d > 1] = O(1/d^2)$. Here Y_d is a Poisson random variable as in (10). The Borel-Cantelli lemma implies that only finitely many Y_d are larger than 1. Hence $\sum_{Y_i > 1} Y_i$ is almost surely bounded. We have seen that, for n large, the sum of C_d with $d \in Q_n$ is at least $\text{Poi}(\frac{\beta}{4} \log n)$ with probability $1 - o(1)$. This diverges, and so one or more of the positive C_d must be equal to one. Our claim follows. \square

For a random permutation $\pi \sim \text{ESF}(\alpha, n)$, we define $\Phi(\pi) = \prod_{\ell=1}^n \ell^{C_\ell}$. To prove Theorem 5, it suffices to show that with high probability π has a cycle of length d such that

$$(32) \quad d \leq n^\beta,$$

$$(33) \quad d^2 \nmid \Phi(\pi).$$

First we will need an estimate on d^2 dividing $\Pi(\pi_n)$ conditional on $C_d = 1$.

Lemma 16. *Let π , β , λ , and Q_n be as in Lemma 15, then for any $d \in Q_n$*

$$\mathbb{P}[d^2 \text{ divides } \Phi(\pi_n) \mid C_d = 1] = o(n^{-\lambda\beta + \varepsilon})$$

for all $\varepsilon > 0$.

Proof. It follows from [AT92, Theorem 3] that the cycle counts C_i with $i \neq d$ are with high probability distributed like $\text{ESF}(\alpha, n - d)$. Accordingly let $n' = n - d$ and $\pi_{n'}$ be a permutation sampled from this distribution. Note that from our construction of Q_n , there exists a prime $p > n^{\lambda\beta}$ such that

$$\{d \mid \Phi(\pi_{n'})\} \subset \{p \mid \Phi(\pi_{n'})\}.$$

For such prime number p , the event on the right hand side above satisfies

$$\{p \mid \Phi(\pi_{n'})\} \subset \left\{ \sum_{i=1}^{n'/p} C_{ip} > 0 \right\}$$

Note we are omitting $i = d/p$ from the above summation, and every subsequent one.

Call the last event $G_{n'}$. The Feller coupling at (10) allows us to bound the probability by

$$(34) \quad \mathbb{P}[G_{n'}] \leq \mathbb{P} \left[\sum_{i=1}^{\lceil n'/p \rceil} Y_{ip} + \mathbf{1}\{J_{n'} = ip\} > 0 \right].$$

This, in turn, is bounded by

$$\mathbb{P}\left[\text{Poi}\left(\sum_{i=1}^{\lceil n'/p \rceil} \frac{\alpha}{ip}\right) > 0\right] + \mathbb{P}\left[\sum_{i=1}^{\lceil n'/p \rceil} \mathbf{1}\{J_{n'} = ip\} > 0\right].$$

The mean of the above Poisson random variable is asymptotically bounded by $\alpha \log(n/p)/p$. Since $p \geq n^{\lambda\beta}$, this ensures that the left summand is $o(n^{-\lambda\beta+\varepsilon})$. To bound the right summand we write

$$\mathbb{P}\left[\sum_{i=1}^{\lceil n/p \rceil} \mathbf{1}\{J_n = ip\} > 0\right] \leq \sum_{ip \leq n-n^\gamma} \mathbb{P}[J_n = ip] + \sum_{ip > n-n^\gamma} \mathbb{P}[J_n = ip].$$

Here γ is a constant we will determine in a moment. We begin by bounding the first summand of the right hand side. By Lemma 11 (11),

$$\begin{aligned} \sum_{ip \leq n-n^\gamma} \mathbb{P}[J_{n'} = ip] &\leq \sum_{i=1}^{\lfloor (n'-n^\gamma)/p \rfloor} \frac{\alpha}{n' - ip} \\ &\leq \int_1^{\frac{n-n^\gamma}{p}+1} \frac{\alpha}{n-d-px} dx \\ &= O\left(\frac{\log n}{p}\right) \\ &= o(n^{-\lambda\beta+\varepsilon}). \end{aligned}$$

The last line follows since $p \geq n^{\lambda\beta}$. For the second summand, from Lemma 11 (12) there exists some constant $c > 0$ such that

$$\begin{aligned} \sum_{ip > n-n^\gamma} \mathbb{P}[J_n = ip] &\leq \sum_{ip > n-n^\gamma} cn^{-\alpha(1-\gamma)} \\ &\leq cn^{\gamma-\lambda\beta} n^{-\alpha(1-\gamma)}. \end{aligned}$$

The last line follows as there are at most $n^{\gamma-\lambda\beta}$ different values of i such that $n' - n^\gamma < ip \leq n'$.

We therefore have that

$$\mathbb{P}[G_{n'}] = O\left(n^{\gamma-\lambda\beta-\alpha(1-\gamma)}\right).$$

An easy calculation confirms that γ satisfying

$$\gamma < \frac{\alpha + \varepsilon}{1 + \alpha}$$

is sufficient to ensure

$$\mathbb{P}[G_{n'}] = o(n^{-\lambda\beta+\varepsilon}).$$

This completes our proof.

□

Lemma 17. *Let π , β , λ , and Q_n be as in Lemma 15, then*

$$(35) \quad \mathbb{P}[\exists d \in Q_n \text{ such that } d^2 \nmid \Phi(\pi) \text{ and } C_d = 1] = 1 - o(1).$$

Proof. We will show the stronger result that

$$\mathbb{P}[d^2 \mid \Phi(\pi) \text{ for some } d \in Q_n \text{ such that } C_d = 1] = o(1).$$

Denoted the above event by E_n so that

$$E_n = \bigcup_{d \in Q_n} \{C_d = 1 \text{ and } d^2 \mid \Phi(\pi)\}.$$

A union bound gives

$$\begin{aligned} \mathbb{P}[E_n] &\leq \sum_{d \in Q_n} \mathbb{P}[C_d = 1 \text{ and } d^2 \mid \Phi(\pi)] \\ &= \sum_{d \in Q_n} \mathbb{P}[d^2 \mid \Phi(\pi) \mid C_d = 1] \mathbb{P}[C_d = 1] \\ &\leq \max_{d \in Q_n} (\mathbb{P}[d^2 \mid \Phi(\pi) \mid C_d = 1]) \left(\sum_{d \in Q_n} \mathbb{P}[C_d = 1] \right). \end{aligned}$$

By Lemma 16, the left probability is $o(n^{-\lambda\beta+\varepsilon})$ for arbitrarily small $\varepsilon > 0$. The Feller coupling ensures that the right hand sum can be bounded by $\alpha\beta \log n$. It follows that $\mathbb{P}[E_n] = o(1)$, completing our proof.

□

Proof of Theorem 5. By Lemma 17 with probability $1 - o(1)$ there exists a cycle length of length d satisfying (32) and (33). When these conditions are met, for $K = \Phi(\pi)/d$, we have π^K displaces $d \leq n^\beta$ elements. Thus, with high probability π has minimal degree no more than n^β . This completes our proof.

□

Proof of Theorem 6. The theorem statement is such that if it holds for $\omega(n)$, then it also holds for any $\omega'(n) \geq \omega(n)$. So, it suffices to prove the statement for all $\omega(n) \leq \sqrt[4]{\log n}$.

Fix such an $\omega(n)$ and set $b_n = n \exp(-\omega(n)\sqrt{\log n})$. Let P be the set of all primes and define $P_n = P \cap [b_n, n]$. Given $p \in P_n$ let $I_p = [1, 2, \dots, \lfloor n/p \rfloor]$ be the set of all i for which pi is a multiple of p between p and n . We also introduce the entire collection of such multiples smaller than n

$$W_n = \bigcup_{p \in P_n} \{(p, i) : i \in I_p\}.$$

It suffices to show that the event

$$E_n = \{\exists (p, i) \in W_n : ip \mid \Phi(\pi)\}$$

occurs with high probability. We can write E_n as a union of events,

$$E_n = \bigcup_{(p, i) \in W_n} \{C_{ip}^{(n)} > 0\}.$$

Notice that any two numbers in the set of products

$$\{ip : (p, i) \in W_n\}$$

are different. To see this we argue by contradiction, suppose that $ip = i'p'$ with $p < p'$. Because of primality, this could only happen if $p' \mid i$. Thus $i \geq p'$. Also note that our assumption $\omega(n) \leq \sqrt[4]{\log n}$ ensures $b_n > \sqrt{n}$. Thus, $p, p' > \sqrt{n}$. These two inequalities give

$$ip \geq p'p > \sqrt{n}\sqrt{n} = n.$$

But $ip > n$ is a contradiction, since our construction of I_p ensures that $ip \leq n$.

We have now shown that the occurrence of E_n is equivalent to the sum of cycle counts being positive: $E_n = \{\sum_{W_n} C_{ip} > 0\}$. The Feller coupling (10) ensures that

$$(36) \quad \sum_{W_n} C_{ip} \geq -D_n + \sum_{W_n} Y_{ip},$$

with D_n the total number of deletions. It follows that

$$(37) \quad \mathbb{P}[E_n] \geq \mathbb{P}[-D_n + \sum_{W_n} Y_{ip} > 0].$$

Using additivity of independent Poisson random variables, the sum on the right of (36) is distributed as a Poisson with mean

$$\begin{aligned}
 \mu_{n,\alpha} &= \sum_{p \in P_n} \sum_{i=1}^{\lfloor n/(p) \rfloor} \alpha/ip \approx \alpha \sum_{p \in P_n} \frac{\log(n/(p))}{p} \\
 (38) \qquad &= \alpha \left(\log n \sum_{p \in P_n} \frac{1}{p} - \sum_{p \in P_n} \frac{\log p}{p} \right).
 \end{aligned}$$

Lemma 18 shows that $\mu_{n,\alpha} \rightarrow \infty$. To establish that (37) occurs with high probability, we will show that the number of deletions, D_n , in the Feller coupling is unlikely to remove every successful Y_{ip} in (36).

As $\mu_{n,\alpha} \rightarrow \infty$, we take any $h(n) = o(\mu_{n,\alpha})$ with $h(n) \rightarrow \infty$. By Lemma 11 (14) we know that $\mathbb{E}D_n \leq c$ for some $c = c(\alpha)$. It follows from Markov's inequality that

$$\mathbb{P}[D_n \geq h(n)] \leq \frac{c}{h(n)} = o(1).$$

Standard estimates on a Poisson random variable tell us that

$$\mathbb{P}[\text{Poi}(\mu_{n,\alpha}) \leq h(n)] = 1 - o(1).$$

So D_n is with high probability smaller than $\sum_{W_n} Y_{ip}$, and combined with (36) we have established that the right side of (37) occurs with high probability. \square

Lemma 18. *Let $\mu_{n,\alpha}$ be as in the proof of Theorem 6. It holds that $\mu_{n,\alpha} \rightarrow \infty$.*

Proof. It suffices to show that (38) tends to infinity. Two classical theorems attributed to Euler ([Eul37]) and Chebyshev ([Apo76]), respectively, state that for sums of primes smaller than n

$$\sum_{p \leq n} \frac{1}{p} = \log \log n + O(1) \quad \text{and} \quad \sum_{p \leq n} \frac{\log p}{p} = \log n + O(1).$$

Recalling that $P_n = P \cap [b_n, n]$, it follows that (38) is asymptotic to

$$\log n \log \frac{\log n}{\log b_n} - \log \frac{n}{b_n}.$$

We can express the second term precisely

$$(39) \qquad \log \frac{n}{b_n} = \log \exp \left(\omega(n) \sqrt{\log n} \right) = \omega(n) \sqrt{\log n}.$$

Similarly we compute the order of the first term. Notice

$$(40) \quad \begin{aligned} \log \frac{\log n}{\log b_n} &= \log \left(\frac{\log n}{\log n - \omega(n)\sqrt{\log n}} \right) \\ &= \log \left(1 + \frac{\omega(n)\sqrt{\log n}}{\log n - \omega(n)\sqrt{\log n}} \right). \end{aligned}$$

As $\omega(n) \leq \sqrt[4]{\log n}$ the above term is $\log(1 - o(1))$. Using the fact that $\log(1 - x) \approx -x$ as $x \rightarrow 0$ we have (40) is asymptotic to

$$\frac{\omega(n)\sqrt{\log n}}{\log n - \omega(n)\sqrt{\log n}}.$$

Reintroducing the $\log n$ factor, we now have

$$(41) \quad \log n \sum_{P_n} \frac{1}{p} \approx \log n \frac{\omega(n)\sqrt{\log n}}{\log n - \omega(n)\sqrt{\log n}}.$$

Combining (39) and (41) gives

$$\begin{aligned} \mu_{n,\alpha} &\approx \log n \frac{\omega(n)\sqrt{\log n}}{\log n - \omega(n)\sqrt{\log n}} - \omega(n)\sqrt{\log n} \\ &= \omega(n)^2 \frac{\log n}{\log n - \omega(n)\sqrt{\log n}} \\ &= \omega(n)^2(1 - o(1)) \end{aligned}$$

Since $\omega(n) \rightarrow \infty$ we conclude that $\mu_{n,\alpha} \rightarrow \infty$. □

Now we set our sights on proving Theorem 3. This will require Theorem 5 and Theorem 6, as well as a few additional lemmas.

Lemma 19. *Let $0 < \gamma < 1$ and $\psi(n) \rightarrow \infty$ such that $\psi(n) = o(\log n)$. The probability that simultaneously for all $2 \leq r \leq \exp(\psi(n))$ there exists $C_{j_r} > 0$ such that $j_r > n^\gamma$ and $r \nmid j_r$ is $1 - o(1)$.*

Proof. Let $r_0 = \exp(\psi(n))$ and consider the complement of the event in question:

$$E = \{\exists r < r_0: r \mid j \text{ for all } n^\gamma < j \text{ such that } C_j > 0\}.$$

It suffices to show that $\mathbb{P}[E] = o(1)$. In order to model the values C_j with independent Poisson random variables we will work with a larger event. For any $\gamma < \gamma' < 1$, we consider

$$E_{\gamma'} = \{\exists r < r_0: r \mid j \text{ for all } n^\gamma < j < n^{\gamma'} \text{ such that } C_j > 0\}.$$

A union bound leads to

$$\begin{aligned}
 \mathbb{P}[E_{\gamma'}] &\leq \sum_{r=2}^{r_0} \mathbb{P}[r \mid j \text{ for all } n^\gamma < j < n^{\gamma'} \text{ such that } C_j > 0] \\
 (42) \qquad &= \sum_{r=2}^{r_0} \mathbb{P}[C_j = 0 \text{ for all } n^\gamma < j < n^{\gamma'} \text{ such that } r \nmid j].
 \end{aligned}$$

Let A_r be the set of integers in $(n^\gamma, n^{\gamma'}]$ that r fails to divide. Formally,

$$A_r = [n^\gamma, n^{\gamma'}] \cap ([n] \setminus \{ir : i \leq \lfloor n/r \rfloor\}).$$

As $n^{\gamma'} = o(n)$ we can apply [ABT00, Theorem 3.1] to conclude that $C[n^\gamma, n^{\gamma'}] \xrightarrow{TV} Y[n^\gamma, n^{\gamma'}]$. Thus, the probabilities at (42) converge to

$$(43) \qquad \sum_{r=2}^{r_0} \prod_{j \in A_r} \mathbb{P}[Y_j = 0] = \sum_{r=2}^{r_0} \exp\left(-\sum_{j \in A_r} \frac{\alpha}{j}\right).$$

The quantity $\sum_{j \in A_r} \frac{\alpha}{j}$ is minimized at $r = 2$, and in this case $\sum_{j \in A_2} \frac{1}{j} \gg \log n$. We can apply this to (43) to obtain

$$\mathbb{P}[E_{\gamma'}] \leq \sum_{r=2}^{r_0} n^{-\delta} \leq r_0 n^{-\delta} = o(1),$$

for some $\delta > 0$. As $E \subseteq E_{\gamma'}$, this completes the proof. \square

Lemma 20. *Let $\delta = \delta(\varepsilon, \alpha) = 1 - (\alpha + \varepsilon) \log 2$. For any $\varepsilon > 0$, there exists a constant $c = c(\varepsilon, \alpha)$ such that the probability of π fixing a set of size k is bounded by $ck^{-\delta}$ uniformly for all $1 \leq k \leq n/2$.*

Proof. In Lemma 24 we show that $\mathbb{P}[k \in \mathcal{L}(\mathbf{X}(\alpha))] \leq Ck^{-\delta}$ for some C and all $k \leq n$. The Feller coupling lets us relate the probability of a fixing a set size to this quantity, the only complication is a possible contribution from $\mathbf{1}\{J_n = \ell\}$. However, the location of this extra cycle is unconcentrated enough to not alter the order of these probabilities.

Lemma 11 (11) ensures that $\mathbb{P}[J_n = \ell] \leq \frac{\alpha}{n-\ell}$. We make use of this as we condition on the value of J_n .

$$(44) \quad \mathbb{P}[\pi \text{ has a } k\text{-cycle}] \leq \mathbb{P}[k \in \mathcal{L}(\mathbf{X}(\alpha))] + \sum_{\ell \leq k} \mathbb{P}[k - \ell \in \mathcal{L}(\mathbf{X}(\alpha))] \mathbb{P}[J_n = \ell]$$

$$(45) \quad \leq Ck^{-\delta} + \sum_{\ell \leq k} C(k - \ell)^{-\delta} \frac{\alpha}{n - \ell}$$

$$(46) \quad \leq Ck^{-\delta} + \frac{2\alpha}{n} \sum_{\ell \leq k} C\ell^{-\delta} \\ \leq Ck^{-\delta} + C'n^{-\delta}.$$

Where at (45) we use our bound on $\mathbb{P}[k \in \mathcal{L}(\mathbf{X}(\alpha))]$ from Lemma 24 and bound $\mathbb{P}[J_n = \ell]$ with Lemma 11 (11). The next line, (46), reindexes the sum and uses that $\ell \leq n/2$ to bound $\frac{\alpha}{n-\ell} \leq 2\alpha/n$. Our claim then follows by bounding the leftmost sum by $n^{-\delta}$ time some constant C' , and the fact that $k^{-\delta} > n^{-\delta}$. \square

We now prove a result analogous to [LP93, Theorem 1]. Namely, that if a permutation sampled according to Ewen's sampling formula is in a transitive subgroup, this subgroup is likely to be either all of S_n or the alternating group, A_n . Though the lemmas culminating to this used rather different arguments, our proof from here closely follows that in [LP93]. Recall that a *primitive subgroup* is a transitive subgroup that does not fix any partition of $[n]$.

Proof of Theorem 3. Let $\pi \in G$ for a transitive subgroup G . According to [GM98], it follows from results of [Bab81] that the minimal degree of a primitive subgroup not containing A_n is at least $(\sqrt{n} - 1)/2$. On the other hand, by Theorem 5 with high probability the minimal degree of $\langle \pi \rangle$ is less than $n^{0.4}$. Thus, the probability that G is primitive is $o(1)$.

Assume G is a non-primitive, transitive subgroup. Then it must fix some partition $\{B_i\}_{i=1}^r$ of $[n]$. By transitivity every block in such a partition will be mapped to every other, thus the partition must be into blocks of equal size, that is

$$|B_i| = \frac{n}{r}.$$

Note that any cycle of a permutation that preserves this partition must have the same number of elements in each block it acts on. We will use this to establish that any $\pi \in G$ with high probability cannot fix this partition for any r .

We consider three cases for the possibly amount, r , of blocks:

Case 1: $2 \leq r \leq r_0 = \exp(\log \log n \sqrt{\log n})$

By Lemma 19 with $\psi(n) = \log \log n \sqrt{\log n}$, with high probability for every r there exists there exists $j_r > n^{.99}$ such that π contains a cycle of length j_r and $r \nmid j_r$. Fix some r and j_r , and let B be the union of all blocks this cycle acts upon. Because r does not divide j_r , B is a proper invariant subset of $[n]$ with size sn/r , where s is the number of blocks in B . By Lemma 20 this occurs with probability bounded by

$$\sum_{r=2}^{r_0} \sum_{s=1}^r c(sn/r)^{-\delta} \leq cr_0^2 (n/r_0)^{-\delta} = o(1).$$

Case 2: $r_0 \leq r \leq n/r_0$

By Theorem 6, with high probability there exists a prime $q > n/r_0$ that divides $\Omega(\pi)$. Thus there exists a cycle of length j such that $q \mid j$. Let $t \leq r$ be the number of blocks this cycle intersects. Because of the bounds on r we have

$$\frac{j}{t} \leq |B_i| = \frac{n}{r} \leq \frac{n}{r_0} < q.$$

Since q divides $j = t \frac{j}{t}$ and $\frac{j}{t} < q$ by the above inequality, we must have that $q \mid t$. However, the total number of blocks r satisfies

$$r < \frac{n}{r_0} < q$$

so this is impossible.

Case 3: $n/r_0 \leq r < n$

Let a be as in Proposition 14 and $\varepsilon > 0$ such that $a + \varepsilon < 1$. Using Lemma 19 with $\psi(n) = \log \log n \sqrt{\log n}$ one more time, we now find $j > n^{a+\varepsilon}$ such that π contains a cycle of length j and the blocks size, $s = n/r$ does not divide j . Again, let B be the union of blocks which intersect this cycle. Because the size of the blocks does not divide j , this intersection is a proper subset of B . We conclude that π must contain another cycle of length j' that intersects each block in B . Hence both cycle lengths j and j' are divisible by the number of blocks in B . We bound the number of blocks,

$$\frac{|B|}{s} > \frac{j}{s} > \frac{n^{a+\varepsilon}}{s} > n^a.$$

However, by Proposition 14 with probability $1 - o(1)$ no two cycle lengths have a common divisor greater than n^a . This completes our proof.

□

3.3. Proving Proposition 8. To rule out generating A_n we need an estimate that shows an $ESF(\alpha, n)$ permutation is odd with probability bounded away from zero. Of course this probability ought to converge to $1/2$, but we were unable to find a proof. We make do with an absolute lower bound.

Lemma 21. *Let $\pi \sim ESF(\alpha, n)$, then*

$$\inf_{n \geq 2} \min\{\mathbb{P}[\pi \text{ is odd}], \mathbb{P}[\pi \text{ is even}]\} \geq \min\left\{\frac{1}{\alpha+1}, \frac{\alpha}{\alpha+1}\right\}.$$

Proof. We will proceed by induction. It is straightforward to check via the Feller coupling that an $ESF(\alpha, 2)$ permutation is odd with probability $\frac{\alpha}{\alpha+1}$, and it is even with probability $1 - \frac{\alpha}{\alpha+1} = \frac{1}{\alpha+1}$.

Suppose that for $2 \leq m \leq n-1$, if $\pi_m \sim ESF(\alpha, m)$ then

$$\min\{\mathbb{P}[\pi_m \text{ is odd}], \mathbb{P}[\pi_m \text{ is even}]\} \geq \varepsilon,$$

for some yet to be determined $\varepsilon > 0$. Now, it suffices to prove that

$$(47) \quad \min\{\mathbb{P}[\pi \text{ is odd}], \mathbb{P}[\pi \text{ is even}]\} \geq \varepsilon.$$

Consider the sequence ξ_1, ξ_2, \dots of Bernoulli random variables for the Feller coupling. Recall that $J_n = (n+1) - R_n$, with R_n the index of the rightmost 1 in ξ_1, \dots, ξ_n . We will condition on the value of J_n , but need to also take into account the parity of n .

n is odd: We condition on the value of J_n . For convenience let $q_\ell = \mathbb{P}[J_n = \ell]$. If $J_n = \ell$, then $\xi_{n+1-\ell} = 1$ and we can view the beginning sequence $\xi_1, \dots, \xi_{n-\ell}, 1$ as corresponding to a random permutation sampled with distribution $ESF(\alpha, n-\ell)$. We denote such a permutation by $\pi_{n-\ell}$. Noting that when $J_n = n$, π is a single n -cycle and therefore even, it follows that

$$\mathbb{P}[\pi \text{ is odd}] = \sum_{\ell \text{ odd}}^{n-3} \mathbb{P}[\pi_{n-\ell} \text{ is even}]q_\ell + \sum_{\ell \text{ even}}^{n-2} \mathbb{P}[\pi_{n-\ell} \text{ is odd}]q_\ell + q_{n-1}.$$

By the inductive hypothesis

$$\mathbb{P}[\pi \text{ is odd}] \geq \sum_{\ell=1}^{n-2} \varepsilon q_\ell + q_{n-1}.$$

Since $\sum_1^n q_\ell = \sum_1^n \mathbb{P}[J_n = \ell] = 1$, it suffices to prove that $q_{n-1} \geq \varepsilon(q_{n-1} + q_n)$. Equivalently, this requires that $(1 - \varepsilon)q_{n-1} \geq \varepsilon q_n$. Notice that

$$q_{n-1} = \mathbb{P}[\xi_2 = 1] \prod_{i=3}^n \mathbb{P}[\xi_i = 0], \text{ and } q_n = \prod_{i=2}^n \mathbb{P}[\xi_i = 0].$$

Much of this cancels when we solve $(1 - \varepsilon)q_{n-1} \geq \varepsilon q_n$, so we arrive at the requirement

$$(1 - \varepsilon) \frac{\mathbb{P}[\xi_2 = 1]}{\mathbb{P}[\xi_2 = 0]} \geq \varepsilon.$$

The ratio $\frac{\mathbb{P}[\xi_2=1]}{\mathbb{P}[\xi_2=0]} = \alpha$, and thus we require

$$(48) \quad (1 - \varepsilon)\alpha \geq \varepsilon.$$

n even: The argument is similar. The key difference is that we instead need $q_n \geq \varepsilon(q_{n-1} + q_n)$.

Rewriting as before this is the same as

$$(49) \quad \varepsilon\alpha \leq (1 - \varepsilon).$$

After solving for ε in (48) and (49), we see that (47) holds whenever $0 < \varepsilon \leq \min\left\{\frac{1}{\alpha+1}, \frac{\alpha}{\alpha+1}\right\}$. \square

Remark 22. When $\alpha = 1$ this confirms the intuition that even and odd permutations occur with probability $1/2$ for all $n \geq 2$.

Corollary 23. *Let $\delta = \delta(\alpha, \varepsilon)$ and $c = c(\alpha, \varepsilon)$ be as in Lemma 20. It holds for all $k \leq n/2$ that*

$$\mathbb{P}[\pi \text{ fixes a } k\text{-set} \mid \pi \text{ is odd}] \leq ck^{-\delta} \cdot \max\left\{\frac{\alpha+1}{\alpha}, \frac{\alpha+1}{1}\right\}.$$

Proof. We start with the reformulation

$$\mathbb{P}[\pi \text{ fixes a } k\text{-set} \mid \pi \text{ is odd}] = \frac{\mathbb{P}[\pi \text{ fixes a } k\text{-set and } \pi \text{ is odd}]}{\mathbb{P}[\pi \text{ is odd}]}.$$

The numerator is bounded by $\mathbb{P}[\pi \text{ fixes a } k\text{-set}]$ which is less than $ck^{-\delta}$ by Lemma 20. We bound the denominator away from zero by Lemma 21, completing our proof. \square

4. THE UPPER BOUND FOR s_α

We carry out the plan described in Section 1.4. First we bound the quenched probability \tilde{p}_k defined formally just below here. The proof is similar to [PPR15, Lemma 3.1]. We repeat a selection of the details for clarity, and point out where the argument differs with general α . Subsequently, we mirror [PPR15, Theorem 3.2] as we apply the Borel–Cantelli lemma.

Recall the functions f_k and g_k from (5). Namely that f_k is the number of available summands smaller than k , and g_k is the largest sum attainable with them. Also, define

$$\tau_\varepsilon := \sup\{k: f_k \geq (\alpha + \varepsilon) \log k\},$$

$$\tau := \sup\{n: g_{\lambda_k} \geq k\},$$

where $\varepsilon > 0$ is yet to be chosen, but small, and $\lambda_k := \lfloor \frac{k}{\alpha \log k} \rfloor$. The proofs of [PPR15, Lemma 2.1, Lemma 2.2] generalize in a straightforward way, so that $T = T(\varepsilon, \delta) := \max\{\tau_\varepsilon, \tau\}$ is almost surely finite. As $\mathbb{P}[T > \lambda_k] \rightarrow 0$, we define the quenched probabilities to avoid this diminishing sequence of events

$$\tilde{p}_k := \mathbb{P}[T < \lambda_k \text{ and } k \in \mathcal{L}(\mathbf{X})].$$

Lemma 24. *Suppose that $\alpha < 1/\log 2$. For each $\varepsilon > 0$ there exists $C = C(\varepsilon)$ such that for all $k \geq 1$,*

$$\tilde{p}_k(\alpha) \leq Ck^{-1+(\alpha+\varepsilon)\log 2}.$$

Proof. Fix $\varepsilon > 0$. Let G_k be the event that $T < \lambda_k$ while also $k \in \mathcal{L}(\mathbf{X})$. Call a sequence $\mathbf{y} = (y_1, y_2, \dots)$ admissible if it is coordinate-wise less than or equal to \mathbf{X} . When G_k occurs, it is not possible for n to be a sum of $\sum_j j y_j$ for an admissible \mathbf{y} with y_j vanishing for $j > \lambda_k$. Indeed, on the event G_k , we have $\sum_{j \leq \lambda_k} j X_j < k$. The event satisfying the following three conditions contains G_k :

- (i) $f_{\lambda_k} \leq (\alpha + \varepsilon) \log \lambda_k$.
- (ii) $g_{\lambda_k} < k$.
- (iii) There is some k with $k = \sum_j (y'_j + y''_j)$ with \mathbf{y}' supported on $[0, \lambda_k]$ and \mathbf{y}'' nonzero and supported on $[\lambda_k + 1, k]$ with both \mathbf{y}' and \mathbf{y}'' admissible.

Define the probability $p'_k = \mathbb{P}[f_{\lambda_k} \leq (\alpha + \varepsilon) \log \lambda_k \text{ and } g_{\lambda_k} < n \text{ and } \ell = \sum_j j y'_j]$ where \mathbf{y}' is admissible and supported on $[1, \lambda_n]$. Also, define $p''_\ell = \mathbb{P}[\ell = \sum_j j y''_j]$ where \mathbf{y}'' is admissible and supported on $[\lambda_k + 1, k]$. Using independence we can obtain an upper bound on \tilde{p}_k by decomposing into these two admissible vectors:

$$\begin{aligned} \tilde{p}_k &\leq \sum_{k=\lambda_k+1}^k p'_{k-\ell} p''_\ell \\ (50) \quad &\leq \left(\sum_{\ell=\lambda_k+1}^k p'_{k-\ell} \right) \max_{\lambda_k \leq \ell \leq k} p''_\ell. \end{aligned}$$

The first factor above is (by Fubini's theorem) equal to $\mathbb{E}|\mathcal{L}(\mathbf{X}) \cap [0, \lambda_k]|$. Using the trivial bound where we assume each sum from \mathbf{X} is distinct, and that we are working on the event $f_{\lambda_k} \leq (\alpha + \varepsilon) \log \lambda_k$ we have

$$(51) \quad |\mathcal{L}(\mathbf{X}) \cap [0, \lambda_k]| \leq 2^{Z_m} \leq \lambda_k^{(\alpha+\varepsilon)\log 2}.$$

We now bound the second term. We will show there exists a constant C such that

$$(52) \quad p_\ell'' \leq C \frac{\log^2 k}{k}, \text{ for all } \ell \in [\lambda_k + 1, k].$$

Let H be the event that $X_j \geq 2$ for some $j \in [\lambda_k, k]$. As $\mathbb{P}[X_j \geq 2] \leq \alpha/j^2$ we can write

$$(53) \quad \mathbb{P}[H] \leq \sum_{j=\lambda_k}^k \alpha/j^2 \leq \alpha/\lambda_k \ll \frac{\log k}{k}.$$

The event that $\ell = \sum_j j y_j''$ for an admissible ℓ supported on $[\lambda_k, k]$, but that H does not occur is contained in the union of events E_j that $X_j = 1$ and $\ell = j + \sum_i i y_i''$ for some admissible \mathbf{y}'' supported on $[\lambda_k, k] \setminus \{j\}$. Using independence of X_j from the other coordinates of \mathbf{X} , along with our description of what must happen if H does not, we obtain

$$(54) \quad \begin{aligned} p_\ell'' &\leq \mathbb{P}[H] + \sum_{j=\lambda_k}^k \frac{1}{j} p_{\ell-j}'' \\ &\leq \mathbb{P}[H] + \frac{1}{\lambda_k} \sum_{j=\lambda_k}^k p_{\ell-j}'' \\ &\leq \frac{\log k}{k} \left(1 + \sum_{j=\lambda_k}^k p_{\ell-j}'' \right). \end{aligned}$$

It remains to bound the summation in the last factor. Here taking $\alpha \neq 1$ is a minor nuisance, but we circumvent any difficulties with the crude bound $\alpha \leq 2$. The idea is to use the generating function

$$F(z, \mathbf{X}) := \prod_{j \in [\lambda_k, k]: X_j=1} (1 + z^j).$$

Writing $[z^j]F$ for the j th coefficient of F . Observe that this corresponds to forming j using the $j \in [\lambda_k, k]$ for which $X_j = 1$. It follows that

$$\sum_{j=\lambda_k}^k p_{\ell-j}'' \leq \sum_{j=\lambda_k}^k \mathbb{E} \left([z^{\ell-j}] F(z, \mathbf{X}) \right).$$

Using that $[z^j]F(z, \mathbf{X}) \leq F(1, \mathbf{X})$ we can bound the above line by $\mathbb{E}F(1, \mathbf{X})$. This is tractable using independence of the X_j . Carrying the calculation out we obtain

$$\begin{aligned} \mathbb{E}F(1, \mathbf{X}) &= \prod_{j=\lambda_k}^n \mathbb{E}[1 + \mathbf{1}\{X_j = 1\}] \\ &= \prod_{j=\lambda_k}^k \left(1 + \frac{\alpha}{j}\right) \\ &\leq \prod_{j=\lambda_k}^n \left(1 + \frac{2}{j}\right). \end{aligned}$$

The last term can be expanded and written as

$$\frac{k^2 + 3k + 2}{\lambda_k^2 + \lambda_k} \ll \frac{k^2}{\lambda_k^2} \ll \log^2 k.$$

Combine this with (54) and we have

$$p'_\ell \ll \frac{\log k}{k} (1 + \log^2 k) \ll \frac{\log^3 k}{k}.$$

This yields (52). To conclude we plug (51) and (52) into (50) and arrive at

$$\tilde{p}_k \leq C(k - \lambda_k)^{(\alpha+\varepsilon)\log 2} \frac{\log^3 k}{k}.$$

This is bounded above by a constant multiple of $k^{-1+(\alpha+2\varepsilon)\log 2}$, which, after swapping all ε 's for $\varepsilon/2$, establishes the lemma. □

Proposition 25. *Let s_α be as in Theorem 7 and $h(\alpha)$ be as in (4). For continuity points of h it holds that $s_\alpha \leq h(\alpha)$, and $s_\alpha \leq h(\alpha) + 1$ for all α .*

Proof of Proposition 25. Fix a small $\varepsilon > 0$ and let $s_{\alpha,\varepsilon} = \lceil (1 - (\alpha - 2\varepsilon)\log 2)^{-1} \rceil$. Sample $s_{\alpha,\varepsilon}$ vectors $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(s_{\alpha,\varepsilon})}$ and let $T(\mathbf{X}^{(1)}), \dots, T(\mathbf{X}^{(s_{\alpha,\varepsilon})})$ be the a.s. random times distributed as T . Take $T^* = \max\{T(\mathbf{X}^{(i)})\}_1^{s_{\alpha,\varepsilon}}$. By independence of the $\mathbf{X}^{(i)}$ and the bound in Lemma 24 we have the probability that $T^* < \lambda_k$ and $k \in \mathcal{L}(\mathbf{X}^{(i)})$ for all $1 \leq i \leq s_{\alpha,\varepsilon}$ is at most a constant multiple of $k^{s_{\alpha,\varepsilon}(-1+(\alpha+\varepsilon)\log 2)}$. Our choice of $s_{\alpha,\varepsilon}$ ensures that the exponent is less than -1 . The series is thus summable. It follows that finitely many $k > T^*$ belong to the intersection of the $\mathcal{L}(\mathbf{X}^{(i)})$. As T^* is almost surely finite (the proof of this can be found in [PPR15, Section 2]), the result follows by letting $\varepsilon \rightarrow 0$ so that $s_{\alpha,\varepsilon} \rightarrow h(\alpha)$. □

5. THE LOWER BOUND FOR s_α

Here we establish the lower bounds on m_α .

Proposition 26. *Let s_α be as in Theorem 7 and $h(\alpha)$ be as in (4). For $0 \leq \alpha < 1/\log 2$, it holds that $s_\alpha \geq h(\alpha)$. Moreover, $\bigcap_{i=1}^{h(\alpha)} \mathcal{L}(\mathbf{X}^{(i)}(\alpha)) = \infty$ a.s. Lastly, when $\alpha \geq 1/\log 2$ it holds that $s_\alpha = \infty$.*

The proof is at the end of this section. The idea is to use Fournier analysis like in [EFG17]. We begin by considering continuity points α of h , and set $m = h(\alpha)$. If $\alpha \geq 1/\log 2$ so that $h(\alpha) = \infty$ we take m to be a fixed arbitrarily large integer. We then fix m i.i.d. Poisson vectors $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}$.

Lemma 27. *Let $I = (0, k]$ and $\varepsilon > 0$. Then the event*

$$A = \left\{ \bigcup_{i=1}^m \mathcal{L}(I, \mathbf{X}^{(i)}) \subset \left[0, \frac{mk}{\alpha\varepsilon}\right] \right\}$$

holds with probability $\mathbb{P}[A] \geq 1 - \varepsilon$.

Proof. This is a straightforward application of Markov's inequality. □

Recall our notation that $\mathbf{X}^{(i)} = (X_j^{(i)})_{j=1}^\infty$.

Lemma 28. *Fix $\varepsilon \in (0, 1/2)$. For any $\delta > 0$, there is a constant $C = C(\varepsilon, \delta, m, \alpha)$ such that the event*

$$(55) \quad E = E(\varepsilon, \delta, m, \alpha) := \left\{ \min_i \sum_{\ell < j \leq k} X_j^{(i)} \geq (1 - \delta)\alpha \log(k/\ell) - C \quad \forall 1 \leq \ell \leq k \right\}$$

holds with probability $\mathbb{P}[E] \geq 1 - \varepsilon$.

We remark that this is a straightforward generalization of [EFG17, Lemma 3.3].

Proof. In fact, we will take $C \gg \frac{1}{\alpha\delta^2} \log(m/\varepsilon)$ while ensuring that $C \geq 1$. Note that the claim holds vacuously when $\ell \geq e^{-C}k$, so we need only consider the case $\ell < e^{-C}k$.

Fix $1 \leq i \leq m$ and let B denote the event that $\sum_{\ell < j \leq k} X_j^{(i)} < (1 - \delta)\alpha \log(k/\ell) - 1$ for some $\ell \leq e^{-C}k$. Writing ℓ' for the smallest power of 2 with $\ell' > \ell$, we thus have

$$(56) \quad \sum_{\ell' < j \leq k} X_j^{(i)} \leq \alpha(1 - \delta) \log(k/\ell')$$

on B .

Let $\mathbb{D} = \{2^i\}_{i \geq 0}$ denote the set of non-negative powers of 2. From (56) it follows that

$$(57) \quad \mathbb{1}_B \leq \sum_{\substack{\ell' \leq 2e^{-C}k, \\ \ell' \in \mathbb{D}}} (1 - \delta)^{\sum_{\ell' < j \leq \ell} X_j^{(i)} - \alpha(1 - \delta) \log(k/\ell')}.$$

Each of the variables $X_j^{(i)}$ are independent and Poisson with mean $\mathbb{E}X_j^{(i)} = \frac{\alpha}{j}$. For a Poisson variable P with mean λ , we have $\mathbb{E}a^P = e^{\lambda(a-1)}$. Moreover, the sum $\sum_{\ell' < j \leq k} X_j^{(i)}$ is Poisson with mean $\alpha \log(k/\ell') + O(1)$. Taking expectations on both sides of (57) therefore yields

$$\mathbb{P}[B] \ll \sum_{\substack{\ell' \leq 2e^{-C}k, \\ \ell' \in \mathbb{D}}} \exp[-\alpha\delta - \alpha(1 - \delta) \log(1 - \delta)] \log(k/\ell')$$

Near zero, $-\delta - (1 - \delta) \log(1 - \delta) = -(1 + o(1))\delta^2$. Therefore

$$\mathbb{P}[B] \ll \sum_{\substack{\ell' \leq 2e^{-C}k, \\ \ell' \in \mathbb{D}}} (k/\ell')^{-\alpha\delta^2} \ll e^{-\alpha\delta^2 C},$$

since the sum of the geometric series is bounded by a constant multiple of its first term. Choosing $C \gg \frac{1}{\alpha\delta^2} \log(m/\varepsilon)$ ensures that $\mathbb{P}[B] \leq \frac{\varepsilon}{m}$, and now the claim follows by the union of events bound. \square

An insight in [EFG17] (to which they credit the idea to Maier and Tenenbaum from [MT84]) is that Fourier analysis may be used to obtain a lower bound on the size of the set

$$(58) \quad S(I; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}) := \left\{ (n_i - n_m)_{1 \leq i < m} : n_i \in \mathcal{L}(I, \mathbf{X}^{(i)}) \right\}.$$

We show that the technique extends to the present level of generality. Note that the analysis of [EFG17] may be regarded as the special case $m = 3$ and $\alpha = 1$ of our arguments.

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the unit torus. For $\theta \in \mathbb{T}$, let $\|\theta\|$ denote the distance to \mathbb{Z} and set $e(z) = e^{2\pi iz}$. Write $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m) \in \mathbb{T}^m$. Define

$$\mathbb{T}_0^m = \{\boldsymbol{\theta} \in \mathbb{T}^m : \theta_1 + \dots + \theta_m = 0\}.$$

Fix an interval I . We define the function $F(\boldsymbol{\theta}) = F_I(\boldsymbol{\theta}, \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)})$ via the formula

$$(59) \quad F(\boldsymbol{\theta}) := \prod_{j \in I} \prod_{i=1}^m \left(\frac{1 + e(j\theta_i)}{2} \right)^{X_j^{(i)}}.$$

Writing $\theta_m = -\theta_1 - \dots - \theta_{m-1}$, we regard $F(\boldsymbol{\theta})$ as a function on \mathbb{T}^{m-1} . It is straightforward to verify that its Fourier transform is a function $\hat{F}: \mathbb{Z}^{m-1} \rightarrow \mathbb{C}$ supported on the set $S(I; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)})$ defined above in (58).

The Cauchy-Schwarz inequality yields that

$$(60) \quad \sum_{a \in \mathbb{Z}^{m-1}} |\hat{F}(a)|^2 \sum_{a: \hat{F}(a) \neq 0} 1 \geq \left[\sum_{a \in \mathbb{Z}^{m-1}} \hat{F}(a) \right]^2.$$

Recognizing that $S(I; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}) \geq |\{a: \hat{F}(a) \neq 0\}|$, that $\sum_{a \in \mathbb{Z}^{m-1}} \hat{F}(a) = F(0) = 1$, and that by Parseval's identity $\sum_{a \in \mathbb{Z}^{m-1}} |\hat{F}(a)|^2 = \int_{\mathbb{T}_0^m} |F(\boldsymbol{\theta})|^2 d\boldsymbol{\theta}$, it follows from (60) that

$$(61) \quad |S(I; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)})| \geq \left(\int_{\mathbb{T}_0^m} |F(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} \right)^{-1}.$$

Our next task is to bound the integral appearing in (61). We will achieve this in two steps: first, by obtaining a pointwise bound with respect to $\boldsymbol{\theta}$ in Lemma 29, and second, by integrating out the $\boldsymbol{\theta}$ dependence in Corollary 30.

Let $\beta \in (0, 1)$ be a real parameter satisfying

$$\beta \alpha \log 2 = 1 - m^{-1} + \delta_2,$$

where $\delta_2 > 0$ and $\delta_2 = o_\delta(1)$ is the parameter occurring in (55). (Note that this is possible since α is a continuity point of h , so $m < (1 - \alpha \log 2)^{-1}$.) Recall the event $E = E(\varepsilon, \delta, m, \alpha)$ from Lemma 28.

Lemma 29. *Fix $\varepsilon \in (0, 1/2)$ and let I denote the interval $I := (k^{1-\beta}, k]$. For a tuple $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$, let $j(\boldsymbol{\theta})$ denote an index for which $\|\theta_j\|$ is maximal. Then*

$$(62) \quad \mathbb{E}_{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}} 1_E |F(\boldsymbol{\theta})|^2 \ll_{\varepsilon, m} \prod_{\substack{i: 1 \leq i \leq m, \\ i \neq j(\boldsymbol{\theta})}} (\|k \theta_i\| \vee 1)^{-1 + \delta_2}.$$

Proof. For each $1 \leq i \leq m$, set $t_i := k \theta_i$ and define the cutoff parameter

$$k_i := \begin{cases} k^\beta, & \|t_i\| \geq k^\beta \\ \|t_i\|, & 1 < \|t_i\| < k^\beta \\ 1, & \|t_i\| \leq 1. \end{cases}$$

Note that $k_i \geq 1$ and $k_i \geq \|t_i\|^\beta$. For each $1 \leq i \leq m$, let Y_i denote the expression

$$Y_i := \sum_{k_i < j \leq k} X_j^{(i)}.$$

By definition of the event E , when it occurs we have that

$$\sum_{i=1}^m Y_i \geq \alpha(1-\delta) \sum_{i=1}^m \log k_i - C(\varepsilon).$$

It will be convenient to rewrite this bound in the form

$$(63) \quad 1_E \ll_{\varepsilon} \prod_{i=1}^m k_i^{-\alpha(1-\delta) \log 2} \prod_{i=1}^m 2^{Y_i}.$$

We define the quantity

$$(64) \quad R = R(\theta, \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}, \varepsilon, \delta, \alpha) := |F(\theta)|^2 \prod_{i=1}^m 2^{Y_i}.$$

To establish the bound (62), we will first show that $\mathbb{E}_{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}} R \ll_{\varepsilon, m} 1$. Once this has been established, it will follow directly from (63) that

$$(65) \quad \mathbb{E} 1_E |F(\theta)|^2 \ll_{\varepsilon, m} \prod_{i=1}^m k_i^{-\alpha(1-\delta) \log 2}.$$

Obtaining (62) from (65) is a simple exercise in lower bounding the maximum of a sequence by its geometric mean. Recalling the definition of $j(\theta)$ above, we have that

$$k_{j(\theta)} \geq \prod_{\substack{i: 1 \leq i \leq m, \\ i \neq j(\theta)}} k_i^{\frac{1}{m-1}}.$$

Consequently

$$(66) \quad \prod_{i=1}^m k_i^{-\alpha(1-\delta) \log 2} \leq \prod_{\substack{i: 1 \leq i \leq m, \\ i \neq j(\theta)}} k_i^{-\frac{m}{m-1} \alpha(1-\delta) \log 2}.$$

Substituting the lower bound $k_i \geq (\|k\theta_i\| \vee 1)^\beta$ into (66), we see that (65) implies

$$\mathbb{E} 1_E |F(\theta)|^2 \ll_{\varepsilon, m} \prod_{\substack{i: 1 \leq i \leq m, \\ i \neq j}} (\|k\theta_i\| \vee 1)^{-\frac{m}{m-1} \alpha \beta (1-\delta) \log 2},$$

and since $\frac{m}{m-1} \alpha \beta (1-\delta) \log 2 = 1 + o_\delta(1)$ by definition of β , the desired bound follows.

Finally we justify that $\mathbb{E}_{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}} R \ll_{\varepsilon, m} 1$. Substituting (59) into (64) yields that

$$R = \prod_{i=1}^m \left[2^{Y_i} \prod_{j \in I} \left| \frac{1 + e(j\theta_i)}{2} \right|^{2X_j^{(i)}} \right].$$

Recall that $Y_i = \sum_{k_i < j \leq k} X_j^{(i)}$ and that $I = (k^{1-\beta}, k]$. Thus

$$R = \prod_{i=1}^m \left[\prod_{k^{1-\beta} < j \leq k_i} \left| \frac{1 + e(j\theta_i)}{2} \right|^{2X_j^{(i)}} \prod_{k_i < j \leq k} \left| \frac{1 + e(j\theta_i)}{\sqrt{2}} \right|^{2X_j^{(i)}} \right].$$

To compute $\mathbb{E}R$, we recall that if P is a Poisson random variable with mean λ , then $\log \mathbb{E}a^P = e^{\lambda(a-1)}$. Since the variables $X_j^{(i)}$ are independent Poisson variables with mean $\mathbb{E}X_j^{(i)} = \frac{\alpha}{j}$, it follows after some simplification that

$$(67) \quad \mathbb{E}R = \exp \sum_{i=1}^m \left[\sum_{k^{1-\beta} < j \leq k_i} \frac{\alpha(\cos(2\pi j\theta_i) - 1)}{2j} + \sum_{k_i < j \leq k} \frac{\alpha \cos(2\pi j\theta_i)}{j} \right].$$

Next we apply the following bound from [EFG17, Lemma 3.4]:

$$(68) \quad \sum_{j \leq k} \frac{\cos(2\pi j\theta)}{j} = \log \min(k, \|\theta\|^{-1}) + O(1).$$

Substituting into (67) yields that

$$\mathbb{E}R = \exp \alpha \sum_{i=1}^m \left[\frac{1}{2} \log \frac{k^\beta \vee \|k\theta_i\|}{k_i \vee \|k\theta_i\|} - \frac{1}{2} \log \frac{k^\beta}{k_i} + \log \frac{k_i \vee \|k\theta_i\|}{1 \vee \|k\theta_i\|} + O(1) \right].$$

By definition of k_i , each summand is $O(1)$. Therefore $\mathbb{E}R \ll_{\varepsilon, m} 1$, as desired. \square

Finally we integrate the bound (62) with respect to θ to obtain a bound on the integral appearing in (61).

Corollary 30. *Let $I = (k^{1-\beta}, k]$ and recall the function $F(\theta) = F_I(\theta)$ from (59). Then*

$$\int_{\mathbb{T}_0^m} \mathbb{E}1_E |F(\theta)|^2 d\theta \ll_{\varepsilon, m} k^{1-m}.$$

Proof. After the change of variables $\mathbf{t} = k\theta$, the bound reduces to verifying that

$$\int_{k \cdot \mathbb{T}_0^m} \mathbb{E}1_E |F(\theta)|^2 d\mathbf{t} \ll_{\varepsilon, m} 1.$$

Applying Lemma 29, we have

$$(69) \quad \int_{k \cdot \mathbb{T}_0^m} \mathbb{E}1_E |F(\theta)|^2 d\mathbf{t} \ll_{\varepsilon, m} \int_{k \cdot \mathbb{T}_0^m} \prod_{\substack{i: 1 \leq i \leq m, \\ i \neq j(\mathbf{t})}} (\|t_i\| \vee 1)^{-(1+\delta_2)} d\mathbf{t}.$$

By symmetry, we may upper bound the integral (69) by

$$(70) \quad m \int_{k \cdot \mathbb{T}_0^m} \prod_{i=1}^{m-1} (\|t_i\| \vee 1)^{-(1+\delta_2)} d\mathbf{t} = m \left(\int_{-k/2}^{k/2} (\|t_i\| \vee 1)^{-(1+\delta_2)} dt_i \right)^{m-1}.$$

The rightmost integral in (70) is bounded above by

$$m \left(2 \int_1^\infty t_i^{-(1+\delta_2)} dt_i + 2 \right)^{m-1}.$$

The desired result now follows since the latter integral is finite and independent of k . \square

Substituting the result of Corollary 30 into (61) lets us show $|S(I; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)})| \gg k^{m-1}$ with high probability.

Proposition 31. *Let $I = (k^{1-\beta}, k]$. With probability greater than $1/2$, $S(I; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}) \subset [-ck, ck]^{m-1}$ for some constant c and $|S(I; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)})| \gg k^{m-1}$.*

Proof. By Lemma 27 with $\varepsilon = 1/3$, with probability greater than $2/3$ we have $\mathcal{L}(I, \mathbf{X}^{(i)}) \subset [0, \frac{3mk}{\alpha}] \subset [0, ck]$ for some constant c and all i . It follows that $S(I; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}) \subset [-ck, ck]^{m-1}$.

By Lemma 28 with $\varepsilon = 1/20$, the event $E(\varepsilon, \delta, m, \alpha)$ holds with probability greater than $19/20$ and by Corollary 30,

$$\int_{\mathbb{T}_0^m} \mathbb{E} 1_E |F(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} \ll_{\varepsilon, m} k^{1-m}.$$

It follows by Markov's inequality that for some sufficiently large L ,

$$\begin{aligned} \mathbb{P} \left[\left\{ \int_{\mathbb{T}_0^m} |F(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} \ll k^{1-m} \right\}^c \right] &\leq \mathbb{P}[E^c] + \mathbb{P} \left[1_E \int_{\mathbb{T}_0^m} |F(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} > Lk^{1-m} \right] \\ &\leq \frac{1}{20} + \frac{\mathbb{E} 1_E \int_{\mathbb{T}_0^m} |F(\boldsymbol{\theta})|^2 d\boldsymbol{\theta}}{Lk^{1-m}} \\ &\leq \frac{1}{10}. \end{aligned}$$

It then follows from (61) that with probability at least $9/10$,

$$|S(I; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)})| \geq \left(\int_{\mathbb{T}_0^m} |F(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} \right)^{-1} \gg k^{m-1}.$$

Both events then occur simultaneously with probability at least $1/2$, completing our proof. \square

Proposition 32. *Let D be sufficiently large constant and $I = (k^{1-\beta}, Dk]$. With probability bounded away from zero there exists an array $(x_j^{(i)} : 1 \leq i \leq m, j \in I)$ for which*

$$0 \leq x_j^{(i)} \leq X_j^{(i)}$$

for all i and $j \in I$, and

$$\sum_{j \in I} j x_j^{(1)} = \dots = \sum_{j \in I} j x_j^{(m)}.$$

Proof. We define the interval $I' = (k^{1-\beta}, k]$. By Proposition 31 with probability greater than $1/2$

$$S(I'; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}) \subset [-ck, ck]^{m-1}$$

and $|S(I'; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)})| \gg k^{m-1}$.

Let $d = 2(5/2)^{1/\alpha}$. Straightforward calculations tell us that independently with probability greater than $1/2$ there exists $j_m \in (2ck, cdk]$ such that $X_{j_m}^{(m)} > 0$. Given this j_m , there are $\gg k^{m-1}$ collections of constants $\{j_\ell\}_1^{m-1} \in (ck, (d+1)ck]^{m-1}$ such that

$$(j_m - j_1, \dots, j_m - j_{m-1}) \in S(I'; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)}).$$

Therefore, independently with probability bounded away from zero, there is one such collection $\{j_\ell\}_1^{m-1}$ such that $X_{j_\ell}^{(\ell)} > 0$ for all ℓ .

By the definition of $S(I'; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(m)})$ we have a set of values $(n_i \in \mathcal{L}(I', \mathbf{X}^{(i)}))_{i=1}^m$ such that

$$(n_1 - n_m)_1^{m-1} = (j_m - j_\ell)_1^{m-1}.$$

Manipulating the terms coordinate-wise yields

$$j_1 + n_1 = \dots = j_m + n_m.$$

Our claim follows by setting $D = (d+1)c$. □

Corollary 33. $\bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)})$ is almost surely infinite.

Proof. We define k_1 to be sufficiently large for Proposition 32 to hold, then inductively define $k_{i+1} = (Dk_i)^{1/(1-\beta)}$. The intervals $I_i = (k_i^{1-\beta}, Dk_i]$ are pairwise disjoint and the set

$$\bigcap_{j=1}^m \mathcal{L}(I_i, \mathbf{X}^{(j)})$$

is non-empty with probability bounded away from zero. Since these events are independent for distinct values of i , the claim now follows from the second Borel-Cantelli lemma. □

Proof of Proposition 26. When $\alpha < 1/\log 2$, as we had fixed $m = h(\alpha)$, by Corollary 33 $s_\alpha \geq h(\alpha)$ holds for continuity points of h . For points of discontinuity, we fix $\alpha' < \alpha$ such that $h(\alpha') = h(\alpha) = m$. We couple two sets of independent realizations $\mathbf{X}^{(1)}(\alpha), \dots, \mathbf{X}^{(m)}(\alpha)$ and $\mathbf{X}^{(1)}(\alpha'), \dots, \mathbf{X}^{(m)}(\alpha')$ such that $X^{(i)}(\alpha') \subseteq X^{(i)}(\alpha)$ for all i . Now $\bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)}(\alpha'))$ is almost surely finite by Corollary 33. Our coupling ensures $\mathcal{L}(\mathbf{X}^{(i)}(\alpha')) \subseteq \mathcal{L}(\mathbf{X}^{(i)}(\alpha))$ for all i , so $\bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)}(\alpha')) \subseteq \bigcap_{i=1}^m \mathcal{L}(\mathbf{X}^{(i)}(\alpha))$. Thus the second intersection is almost surely infinite, and our claim follows. When $\alpha \geq 1/\log 2$ all of

the preliminary results in this section hold for arbitrarily large m , and thus every finite intersection is almost surely infinite. \square

Proof of Theorem 7. This now follows immediately from Proposition 25 and Proposition 26. \square

REFERENCES

- [ABT92] Richard Arratia, Andrew Barbour, and Simon Tavaré, *Poisson process approximations for the ewens sampling formula*, Ann. Appl. Probab. **2** (1992), no. 3, 519–535.
- [ABT00] ———, *Limits of logarithmic combinatorial structures*, Ann. Probab. **28** (2000), no. 4, 1620–1644.
- [ABT16] ———, *Exploiting the feller coupling for the ewens sampling formula*, Statist. Sci. **31** (2016), no. 1, 27–29.
- [Apo76] Tom M. Apostol, *Introduction to analytic number theory*, Undergraduate texts in mathematics, Springer, New York, Berlin, Heidelberg, 1976, Autres tirages : 1984, 1986, 1995, 1998, 2010.
- [AT92] Richard Arratia and Simon Tavaré, *The cycle structure of random permutations*, Ann. Probab. **20** (1992), no. 3, 1567–1591.
- [AT94] ———, *Independent process approximations for random combinatorial structures*, Adv. Math. **104** (1994), 90–154.
- [Bab81] László Babai, *On the order of uniprimitive permutation groups*, Annals of Mathematics **113** (1981), no. 3, 553–568.
- [Bov80] JD Bovey, *The probability that some power of a permutation has small degree*, Bulletin of the London Mathematical Society **12** (1980), no. 1, 47–51.
- [BUV11] Volker Betz, Daniel Ueltschi, and Yvan Velenik, *Random permutations with cycle weights*, Ann. Appl. Probab. **21** (2011), no. 1, 312–331.
- [Cra16a] Harry Crane, *Rejoinder: The ubiquitous ewens sampling formula*, Statist. Sci. **31** (2016), no. 1, 37–39.
- [Cra16b] ———, *The ubiquitous ewens sampling formula*, Statist. Sci. **31** (2016), no. 1, 1–19.
- [Dix92] John D. Dixon, *Random sets which invariably generate the symmetric group*, Discrete Mathematics **105** (1992), no. 1, 25 – 39.
- [DS00] J.H. Davenport and G.C. Smith, *Fast recognition of alternating and symmetric galois groups*, Journal of Pure and Applied Algebra **153** (2000), no. 1, 17 – 25.
- [EFG15] Sean Eberhard, Kevin Ford, and Ben Green, *Permutations fixing a k -set*, International Mathematics Research Notices (2015).
- [EFG17] ———, *Invariable generation of the symmetric group*, Duke Math. J. **166** (2017), no. 8, 1573–1590.
- [ET67] P Erdős and P Turán, *On some problems of a statistical group-theory. ii*, Acta Mathematica Hungarica **18** (1967), no. 1-2, 151–163.
- [EU14] Nicholas M. Ercolani and Daniel Ueltschi, *Cycle structure of random permutations with cycle weights*, Random Structures & Algorithms **44** (2014), no. 1, 109–133.
- [Eul37] Leonhard Euler, *Observationes circa series, infinitas*, Commentarii Academiae Scientarum Petropolitanae **9** (1737), 160–188.

- [Ewe72] W.J. Ewens, *The sampling theory of selectively neutral alleles*, Theoretical Population Biology **3** (1972), no. 1, 87 – 112.
- [Fel45] William Feller, *The fundamental limit theorems in probability*, Bulletin of the American Mathematical Society **51** (1945), no. 11, 800–832.
- [GM98] Robert Guralnick and Kay Magaard, *On the minimal degree of a primitive permutation group*, Journal of Algebra **207** (1998), no. 1, 127–145.
- [GP16] A. Gladkich and R. Peled, *On the Cycle Structure of Mallows Permutations*, ArXiv e-prints (2016).
- [Gra08] Andrew Granville, *The anatomy of integers and permutations*, preprint. (2008).
- [Hei86] Joos Heintz, *On polynomials with symmetric galois group which are easy to compute*, Theoretical Computer Science **47** (1986), 99 – 105.
- [KKRW15] R. Kenyon, D. Kral, C. Radin, and P. Winkler, *Permutations with fixed pattern densities*, ArXiv e-prints (2015).
- [LP93] Tomasz Luczak and László Pyber, *On random generation of the symmetric group*, Combinatorics, Probability and Computing **2** (1993), 505–512.
- [Mal57] C. L. Mallows, *Non-null ranking models. i*, Biometrika **44** (1957), no. 1/2, 114–130.
- [MT84] H. Maier and G. Tenenbaum, *On the set of divisors of an integer*, Inventiones mathematicae **76** (1984), no. 1, 121–128.
- [Muk16] Sumit Mukherjee, *Fixed points and cycle structure of random permutations*, Electron. J. Probab. **21** (2016), 18 pp.
- [Mus78] David R. Musser, *On the efficiency of a polynomial irreducibility test*, J. ACM **25** (1978), no. 2, 271–282.
- [PPR15] Robin Pemantle, Yuval Peres, and Igor Rivin, *Four random permutations conjugated by an adversary generate sn with high probability*, Random Structures & Algorithms (2015), n/a–n/a.
- [Wae34] B.L. van der Waerden, *Die seltenheit der gleichungen mit affekt*, Mathematische Annalen **109** (1934), 13–16.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON

E-mail address: gerandy@math.washington.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON

E-mail address: cff2008@math.washington.edu

DEPARTMENT OF MATHEMATICS, DUKE UNIVERSITY

E-mail address: jungem@math.duke.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON

E-mail address: avius@math.washington.edu

Chapter 3

**HOMOLOGY OF MULTI-PARAMETER RANDOM
SIMPLICIAL COMPLEXES**

HOMOLOGY OF MULTI-PARAMETER RANDOM SIMPLICIAL COMPLEXES

CHRISTOPHER F. FOWLER

Abstract

We consider a multi-parameter model for randomly constructing simplicial complexes that interpolates between random clique complexes and Linial-Meshulam random k -dimensional complexes. Unlike these models, multi-parameter complexes exhibit nontrivial homology in numerous dimensions simultaneously. We establish upper and lower thresholds for the appearance of nontrivial cohomology in each dimension and characterize the behavior at criticality.

1. INTRODUCTION

1.1. Background. Many problems in physics, economics, biology, and mechanics involve the modeling of extremely large and intricate systems. With such high levels of complexity, understanding these systems from their microscopic structure is often intractable. In such cases it may make more sense to view them as random topological spaces with certain probability parameters. This framework enables us to make a variety of powerful conclusions about how these systems will generally behave. Indeed, as mentioned in [Kah14b], the study of random geometric and topological spaces has on several occasions lent intuition to the extraordinary prevalence of certain properties amongst mathematical objects.

The purpose of this work is to understand the homological behavior of a generalized model for random simplicial complexes, mentioned in [Kah14b] and recently explored in [CF16]. We define $X(n, p_1, p_2, \dots)$ to be the probability distribution over simplicial complexes on vertex set $[n] = \{1, \dots, n\}$ whose distribution on 1-skeletons agrees with $G(n, p_1)$. The distribution on higher dimensional skeletons is constructed inductively: for an integer $k > 1$, any k -simplex whose boundary is contained in our complex is added with probability p_k . This provides a measure on all simplicial complexes on n vertices. Two well studied structures, random k -complexes (Linial–Meshulam complexes for $k = 2$ and Meshulam–Wallach complexes for higher d) and clique complexes, are realized as $X(n, 1, \dots, 1, p_k, 0, \dots)$ and $X(n, p, 1, \dots)$.

The study of random topological spaces began with random graphs, the seminal example of which is $G(n, p)$, the Erdős–Rényi model. Given a probability parameter $p \in (0, 1)$, typically a function of n , we consider a graph on n vertices where every edge between two vertices of G is added independently with probability p . This defines a probability measure on the set of all simple graphs on n vertices and we say $G(n, p)$ to indicate a random graph with law $G(n, p)$.

Most random topology results pertain to the asymptotic behavior of a model, ie. what happens as the number of vertices tends to infinity. Given some property \mathcal{A} of simplicial complexes, we say that $X \in \mathcal{A}$ *with high probability*, or *w.h.p.*, if

$$\lim_{n \rightarrow \infty} \mathbb{P}[X \in \mathcal{A}] = 1.$$

A formative result of random graph theory, proven by Erdős and Rényi in [ER59], was the sharp threshold of $p = \log n/n$ for connectivity in $G(n, p)$: if $p \geq (\log n + \omega(1))/n$ then $G(n, p)$ is w.h.p. connected, and if $p \leq (\log n - \omega(1))/n$ then $G(n, p)$ is w.h.p. disconnected.

Significant work has been done on the behavior of random graphs since [ER59]. Providing a higher dimension analog, recent study has been focused on several models for random simplicial complexes. One of the most natural questions to ask, results in this field often depict the homological or cohomological behavior of a complex. Even the connectivity threshold for $G(n, p)$ is a statement about the 0-homology of graphs: $H_0(G, \mathbb{Z}) = \mathbb{Z}^m$ where m is the number of connected components of G .

In this context there are two different types of phase transitions that occur in these models. For any given dimension, there can be a threshold at which homology or cohomology changes from trivial to nontrivial. Conversely, there can be a threshold at which it goes from nontrivial to trivial, or *vanishes*. Extensive work on has been done to establish lower bounds on the thresholds at which homology appears and upper bounds on the thresholds at which homology vanishes for various models.

A high-dimensional analog to $G(n, p)$ is $Y_k(n, p)$, the model for random k -dimensional simplicial complexes. We begin with a complex on n vertices and full $(k - 1)$ -skeleton, then add every possible k -face independently with probability p . Linial and Meshulam initially considered when $k = 2$ in [LM06], establishing a sharp threshold for when \mathbb{Z}_2 -homology disappears in the first dimension. Babson, Hoffman, and Kahle later looked at the fundamental group of this model in [BHK11], proving a threshold where $\pi_1(Y_2(n, p))$ transitions w.h.p. from hyperbolic to trivial.

Meshulam and Wallach in [MW09] extended the result in [LM06] to $H_{k-1}(Y_k(n, p), \mathbb{Z}_q)$ for any dimension k . Their work was followed by [HKP13], where Hoffman, Kahle, and Paquette demonstrated an upper bound for the vanishing of integer homology in this model. It is also natural to ask how $H_k(Y_k(n, p), \mathbb{Z})$ behaves in these complexes. Kozlov proved a threshold for the appearance of k -homology in [Koz10]. Aronshtam and Linial in [AL15], joined by Łuczak and Meshulam in [ALLM13], extended this work to bounds on when the top dimension of this complex is in fact collapsible.

Another model of interest is the random clique complex model, $X(n, p)$. Just as in our own model, the distribution of the 1-skeleton is identical to $G(n, p)$, but in this case the edges dictate the entire complex. Given some $X \stackrel{dist}{=} X(n, p)$, X contains the k -simplex spanned by a set of $k + 1$ vertices only if the vertices form a complete subgraph in X , called a $(k + 1)$ -clique. For any dimension k , Kahle established in [Kah09] and [Kah14a] sharp thresholds for p for which there will be nontrivial k -th cohomology. This shows that, outside the critical windows of these thresholds, cohomology will w.h.p. be nontrivial in just one dimension, the middle dimension of the complex. Kahle has proved numerous results concerning the behavior of $X(n, p)$, such as establishing a central limit theorem for the distribution of Betti numbers $\beta_k = \dim(H^k(X, \mathbb{Q}))$ with Meckes in [KM⁺13].

As we noted before, all these complexes are special cases of $X(n, p_1, p_2, \dots)$. The random graph model $G(n, p)$ is identical to $X(n, p, 0, \dots)$, $Y_k(n, p)$ corresponds to $X(n, 1, \dots, 1, p_k = p, 0, \dots)$, and clique complexes are the case $X(n, p_1, 1, \dots)$. In fact, many of our results are achieved through a reworking of frameworks laid down in [Kah09] and [Kah14a]. This appears to be the natural bridge between these models, and we show that often the results for specific models may be extended to this broader construction. Through this process we exhibit cohomological behavior unique to this model.

Significant work has been done on this model concurrently by Costa and Farber, where they introduce the additional parameter p_0 for adding vertices of $[n]$. In [CF16] they address the containment problem: given a r -dimensional subcomplex S , they define a convex set $\bar{\mathcal{M}}(S) \subset \mathbb{R}^{r+1}$ such that if $(\alpha_0, \dots, \alpha_k) \in \bar{\mathcal{M}}(S)$, then X w.h.p. contains a subcomplex isomorphic to S . In [CF17a] and [CF17b], they look at the fundamental group of these complexes, establishing regimes in which it w.h.p. trivial, nontrivial, and specifically has property (T). They also show cohomology is concentrated in a *critical dimension*, showing bounds on the Betti numbers in this and nearby dimensions, as well as bounding the size of possible cycles in higher dimensions.

1.2. Statement of Results. *Notation:* We write $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ to indicate that X is chosen from the distribution $X(n, p_1, p_2, \dots)$.

Our theorems deal with the $(k-1)$ -th homology or cohomology of $X(n, p_1, p_2, \dots)$. As mentioned above there are two types of phase transitions, and we work to develop bounds on the thresholds for both. Since the $(k-1)$ -th (co)homology of a simplicial complex depends only on its k -skeleton, these theorems only depend on probabilities p_1 through p_k . The primary open problem from this work concerns the $(k-1)$ -th homology of our complexes when $p_k = 1$, which we discuss following our statement of results.

As with clique complexes, the $(k-1)$ -cohomology of $X(n, p_1, p_2, \dots)$ has two phase transitions. We begin with no $(k-1)$ -simplices and trivial cohomology, as our probabilities increase cohomology appears, then eventually the p_i become too large and it disappears. This defines the two thresholds we wish to understand.

The following result establishes when the probabilities are sufficiently large that we will have trivial cohomology.

Theorem 1. *Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with $p_i = n^{-\alpha_i}$ and $\alpha_i \geq 0$ for all i . If*

$$(1) \quad \sum_{i=1}^k \alpha_i \binom{k}{i} < 1$$

then w.h.p. $H^{k-1}(X, \mathbb{Q}) = 0$.

We prove this threshold is essentially the best possible by establishing nontrivial cohomology on the other side of (1). Moreover, the second regime for which cohomology exists is establishes the potential for $H^k(X, \mathbb{Q}) \neq 0$ simultaneously for several k .

Theorem 2. *Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with $p_i = n^{-\alpha_i}$, $\alpha_i \geq 0$ for all i , and*

$$(2) \quad 1 \leq \sum_{i=1}^k \alpha_i \binom{k}{i}.$$

If

$$(3) \quad \sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i} < 1$$

then w.h.p. $H^{k-1}(X, \mathbb{Q}) \neq 0$. Furthermore, when $\alpha_k > 0$ we can relax the condition in (3) to

$$(4) \quad \sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1} < k+1.$$

A common question to ask concerning phase transitions is what happens at the boundary between phases. Given a complex X , we let β_k denote the k -th Betti number of X , given by $\beta_k := \dim(H^k(X, \mathbb{Q}))$. Allowing the p_i to be more varied functions of n , we identify this critical region and establish a limit theorem for the Betti number. Combined with Theorems 1 and 2, this proves a threshold for vanishing cohomology for all possible p_i .

Theorem 3. Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with

$$p_i = (\rho_1 \log n + \rho_2 \log \log n + c)^{\nu_i} n^{-\alpha_i}$$

such that

$$\rho_1 = k - \sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}, \quad \rho_2 = \sum_{i=1}^{k-1} \nu_i \binom{k}{i+1} \quad \text{and} \quad \sum_{i=1}^k \alpha_i \binom{k}{i} = \sum_{i=1}^k \nu_i \binom{k}{i} = 1.$$

Then β_{k-1} approaches a Poisson distribution

$$\beta_{k-1} \rightarrow \text{Poi}(\mu)$$

with mean

$$\mu = \frac{\rho_1^{\rho_2} e^{-c}}{k!}.$$

We also provide a lower bound on the threshold where homology first appears in our complex. This bound, combined with the second part of Theorem 2, is essentially the best possible when $\alpha_k > 0$.

Theorem 4. Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with $p_i = n^{-\alpha_i}$ and $\alpha_i \geq 0$ for all i . If

$$(5) \quad k+1 < \sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1}$$

then w.h.p. $H_{k-1}(X, \mathbb{Z}) = 0$.

Repeated application of our theorems for each dimension will often fully describe the cohomology of our random complex. Specifically when the set of $p_i = n^{-\alpha_i}$ parameters fall within the specified regimes of Theorems 1, 2, and 4.

Example 5. Consider $X \stackrel{\text{dist}}{=} X(n, p_1, p_2, \dots)$ with $p_i = n^{-\alpha_i}$ and

$$(\alpha_1, \alpha_2, \alpha_3, \dots) = \left(\frac{1}{5}, \frac{1}{5}, \frac{2}{5}, 0, 0, \dots\right).$$

Then

$$\sum_{i=1}^2 \alpha_i \binom{2}{i} = \frac{3}{5} < 1$$

so w.h.p. $H^1(X, \mathbb{Q}) = 0$ by Theorem 1.

$$\sum_{i=1}^3 \alpha_i \binom{3}{i} = \frac{7}{5} > 1 \text{ and } \sum_{i=1}^{3-1} \alpha_i \binom{3-1}{i} = \frac{3}{5} < 1.$$

so by Theorem 2 w.h.p. $H^2(X, \mathbb{Q}) \neq 0$. Finally,

$$\sum_{i=1}^{4-1} \alpha_i \binom{4+1}{i+1} = 6 > 4 + 1,$$

so by Theorem 4 $H_3(X, \mathbb{Z}) = 0$ w.h.p. Moreover, a simple bound using the last equation allows us to use the Theorem to deduce $H_k(X, \mathbb{Z}) = 0$ w.h.p. for all $k \geq 3$.

The proof of Theorem 1 is handled in Sections 3 and 4. The inequality (1) precisely ensures every $(k-1)$ -simplex of X is w.h.p. contained in a k -simplex, so no single face generates a non-trivial cocycle in $H^{k-1}(X)$. With this condition satisfied we prove the result by applying [BS97, Theorem 2.1], a result connecting spectral gap theory and the homology of simplicial complexes and presented in Section 2. Most of the work lies in showing the various hypotheses of the theorem are met by our complexes, for which we use [HKP12, Theorem 1.1], a tool for bounding the spectral gap of Erdős–Rényi random graphs.

Theorem 2 is proven in Sections 5 and 6. The statement for the range defined by (2) and (3) is shown by exhibiting that our complex will have far more $(k-1)$ -dimensional faces than those in adjacent dimensions, so the kernel of the coboundary map is very large. In fact, the second moment argument used in the proof yields the stronger result that within this range of values our Betti number β_{k-1} will grow polynomially in n . We write $X \sim Y$ with high probability if for all $\epsilon > 0$, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}[(1 - \epsilon) \leq Y/X \leq (1 + \epsilon)] \rightarrow 1.$$

Using this notation, we have a strong handle on the size of cohomology within this window.

Theorem 6. Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with $p_i = n^{-\alpha_i}$ and $\alpha_i \geq 0$ for all i and β_{k-1} be the $(k-1)$ -th Betti number. If

$$\sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i} < 1 < \sum_{i=1}^k \alpha_i \binom{k}{i},$$

then

$$g_{k-1}(n, \alpha_1, \dots, \alpha_k) = \binom{n}{k} n^{-\sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}}$$

is growing polynomially, and w.h.p.

$$\beta_{k-1} \sim g_{k-1}(n, \alpha_1, \dots, \alpha_k).$$

The result when (2) and (4) hold extends the argument presented in the example at the end of this section: showing our complex will w.h.p. contain certain subcomplexes that generate nontrivial homological cycles.

In Section 7 we prove Theorem 3. The strict requirements on our p_i define a range where we have a non-zero but finite expected number of maximal $(k-1)$ -faces. A factorial moment argument shows this number approaches a limiting distribution, a slight adaptation of the work in Section 4 then proves these faces generate the only nontrivial cocycles of dimension $k-1$.

Finally, the proof of Theorem 4 is found in Section 8. The subset (5) defines when our complex will w.h.p. not contain the boundary of a k -simplex. We show this is the most likely subcomplex to appear in X that generates a $(k-1)$ -cycle. Thus, when X w.h.p. does not contain the boundary of a k -simplex it will have no $(k-1)$ -cycles.

1.3. Discussion. Primarily our results concern when $p_i = n^{-\alpha_i}$ with $\alpha_i \geq 0$ or $p_i = 0$ (here we say $\alpha_i = \infty$). This was done to make the theorem statements as concise as possible. Our threshold results extend easily to when p_i are more varied functions of n . If $p_i = \omega_i n^{-\alpha_i}$ with $\omega_i(n) \rightarrow \infty$ and $\omega_i(n) = o(n^\epsilon)$ for all $\epsilon > 0$, then Theorems 1, 2, and 4 still hold provided the α_i do not lie on the boundary between two thresholds.

Our work on this multi-parameter model confirms it as the natural bridge between $X(n, p)$ and $Y_k(n, p)$. Our theorems imply the analogous results for the rational cohomology of these complexes. However, it's important to note for both these models there are results concerning the vanishing of homology over arbitrary field coefficients. The boundary between Theorems 1 and 2 is sharp when $p_i = n^{-\alpha_i}$, and combined with Theorem 3 establishes a sharp upper bound for vanishing

cohomology that encompasses the analogous results for clique complexes [Kah14a, Theorem 1.1] and Linial-Meshulam complexes [MW09, Theorem 1.1].

While our bounds on the threshold where homology vanishes are seen to be close to optimal, we have not fully characterized the threshold for appearing homology. Our bounds for when H_{k-1} first becomes nontrivial are good so long as $\alpha_k > 0$, and Kahle proved the correct bound for clique complex case in [Kah09], but we have been unable to generalize his arguments or find another method. For now we leave this as an open problem.

Open Problem 1 What is the correct threshold for the appearance of $H_{k-1}(X, \mathbb{Z})$ when $\alpha_k = 0$? I.e., what hyperplane for $\alpha_1, \dots, \alpha_{k-1}$ determines whether homology is trivial or nontrivial?

Note that $\alpha_k = 0$, or $p_k = 1$, implies X cannot contain the unfilled boundary of a k -simplex, the question likely reduces to understanding the smallest homological cycle that can appear in X . We suspect the answer is determined, perhaps uniquely, by the largest $l < k$, such that $p_l \neq 1$. Meanwhile, the Linial–Meshulam and Meshulam–Wallach models begin with nontrivial $(k-1)$ -homology, and $p_{k+1} = 0$ so our bounds for $H_k(Y_k(n, p))$ are roughly in line with the main results, though we only consider probability parameters that are powers of n .

Another open problem concerns when integer homology vanishes in a specific dimension.

Open Problem 2 Does (1) in Theorem 1 imply that w.h.p. $H_{k-1}(X, \mathbb{Z}) = 0$?

We understand the phase transition for $H^{k-1}(X, \mathbb{Q})$ and have reason to believe our results should hold for integer homology, but our present arguments are insufficient. We note this question is also currently unsolved for $X(n, p)$.

Although $X(n, p)$ and $Y_k(n, p)$ seem like quite different instances of $X(n, p_1, p_2, \dots)$, they do not fully characterize our model. We often observe asymptotic behavior dramatically different from either one. In fact, for any fixed integer l we can find some k such that the range of values for p_i defined by applications of Theorem 2 in dimensions k through $k+l$ is nontrivial. This yields a result exemplifying the differences in this model.

Corollary 7. *Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with $p_i = n^{-\alpha_i}$, for any integer l there exists an integer k and an open set of α_i for which X w.h.p. has non-trivial cohomology in dimensions k through $k+l$.*

Proof. The result follows directly from Theorem 2. Considering (2), if

$$1 \leq \sum_{i=1}^k \alpha_i \binom{k}{i}$$

for some k , then

$$1 \leq \sum_{i=1}^j \alpha_i \binom{j}{i}$$

for all $j > k$. Similarly, considering (4), if

$$\sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1} < k+1,$$

then

$$\sum_{i=1}^{j-1} \alpha_i \binom{j+1}{i+1} < j+1$$

for all $j < k$.

We will construct a simple, far from optimized open set. We fix our l and let k be sufficiently large such that $k > l + 2$. If

$$\frac{1}{k+1} < \alpha_1,$$

then

$$1 < \sum_{i=1}^{k+1} \alpha_i \binom{k+1}{i}.$$

Moreover, if

$$\alpha_1 < \frac{1}{k},$$

then

$$\alpha_1 \binom{k+l+2}{2} < k+l+2.$$

Thus, so long as $p_i \neq 1$ for $k+1 \leq i \leq k+l+1$ and

$$\sum_{i=2}^{k+l} \alpha_i \binom{k+l+2}{i+1} < 1,$$

then our result follows from Theorem 2. □

1.4. Low-dimensional example. We present a low-dimensional example to give some intuition for where the inequalities in our theorems come from, as well as illustrate the potential for non-trivial cohomology in multiple dimensions simultaneously.

Example 8. Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$, if

$$\alpha_2, \alpha_3 > 0, 6\alpha_1 + 4\alpha_2 < 4, \text{ and } 1 \leq 2\alpha_1 + \alpha_2$$

then w.h.p. $H^1(X, \mathbb{Q}) \neq 0$ and $H^2(X, \mathbb{Q}) \neq 0$.

Proof. Within this proof, and later in Section 5, we consider the appearance of certain subcomplexes in X . First, we establish the presence of triangles with an unfilled 2-face whose first edge, determined lexicographically, is not part of any 2-face in X . Our complex is defined on the vertex set $[n]$, and for any $j \in \binom{[n]}{3}$ we let A_j denote the event that the vertex set corresponding to j forms such a subcomplex. Using independence, this has probability

$$\mathbb{P}[A_j] = p_1^3 (1 - p_2) (1 - p_1^2 p_2)^{n-3}.$$

The first two terms require the three edges are in X while the 2-simplex itself is not present. The last term ensures our first edge is *maximal*, i.e. does not form a 2-simplex with any of the $n - 2$ remaining vertices.

Letting M_1 denote the number of such subcomplexes in X , by linearity of expectation

$$\mathbb{E}[M_1] = \sum_{j \in \binom{[n]}{3}} \mathbb{P}[A_j] = \binom{n}{3} p_1^3 (1 - p_2) (1 - p_1^2 p_2)^{n-3}.$$

Using standard first moment techniques we see, for large enough n ,

$$\begin{aligned} \mathbb{E}[M_1] &\approx \frac{n^3}{6} n^{-3\alpha_1} (1 - p_2) \left(1 - n^{-(2\alpha_1 + \alpha_2)}\right)^n \\ &\approx \frac{1}{6} n^{3-3\alpha_1} (1 - p_2) e^{-n^{1-(2\alpha_1 + \alpha_2)}}. \end{aligned}$$

Since $\alpha_2 > 0$, we have that $1 - p_2 > 0$. The last two terms are therefore $\Theta(1)$ when $1 \leq 2\alpha_1 + \alpha_2$, then $\alpha_1 < 1$ implies $\mathbb{E}[M_1] \rightarrow \infty$. Second moment arguments, detailed in Appendix A, then show that w.h.p. $M_1 > 0$.

We now show the existence of tetrahedrons with unfilled 3-face, the first 2-face of which is maximal. For each $l \in \binom{[n]}{3}$, let B_l be the event that the vertices l form such a subcomplex in X .

Similar considerations show

$$\mathbb{P}[B_l] = p_1^6 p_2^4 (1 - p_3) (1 - p_1^3 p_2^3 p_3)^{n-4}.$$

Letting M_2 denote the total number of such subcomplexes in X , linearity of expectation shows

$$\mathbb{E}[M_2] = \sum_{l \in \binom{[n]}{4}} \mathbb{P}[B_l] = \binom{n}{4} p_1^6 p_2^4 (1 - p_3) (1 - p_1^3 p_2^3 p_3)^{n-4}.$$

It follows that if $\alpha_3 > 0$, $6\alpha_1 + 4\alpha_2 < 4$, and $1 \leq 3\alpha_1 + 2\alpha_2 + \alpha_3$, then $\mathbb{E}[M_2] \rightarrow \infty$. Second moment calculations establish that w.h.p. $M_2 > 0$.

Combining the two sets of requirements on p_i yields that whenever $p_2, p_3 \neq 1$, $1 \leq 2\alpha_1 + \alpha_2$, and $6\alpha_1 + 4\alpha_2 < 4$ w.h.p. $M_1, M_2 > 0$. Each such subcomplex can be seen to generate a non-trivial \mathbb{Z} -summand in the 1 or 2-homology, respectively. Thus w.h.p. $H_1(X, \mathbb{Z}) \neq 0$ and $H_2(X, \mathbb{Z}) \neq 0$, and our result follows by the Universal Coefficients Theorem, covered in the next section. \square

2. TOPOLOGICAL PRELIMINARIES

2.1. Basic definitions. Before proceeding, we lay out the definitions and theorems critical to our work. For further reference, specifically regarding the homology and cohomology of simplicial complexes, we direct the reader to [Hat02].

A crucial definition for our work is the link of a subcomplex. Given a simplicial complex X and a k -dimensional simplex σ in X , we define the *link of σ in X* , denoted $\text{lk}_X(\sigma)$, to be a new simplicial complex with vertex set corresponding to the vertices of X that form an $(k+1)$ -face with σ . We then construct the new simplicial complex by adding the $(l-1)$ -face corresponding to a set of vertices v_1, \dots, v_l precisely when the vertices $\sigma \cup \{v_1, \dots, v_l\}$ comprise a $(k+l)$ -face in X .

A simplicial complex X is *pure k -dimensional* if every face of X is contained in a k -dimensional face.

Finally, let G be some graph of ordered vertices with minimum degree at least 1. Define D and A to be the associated degree and adjacency matrices of G , respectively. We define the *normalized Laplacian of G* , denoted \mathcal{L} , by

$$\mathcal{L} = I - D^{-1/2} A D^{-1/2}.$$

For our work we look at the *spectral gap of G* (denoted $\lambda_2[G]$), which is the absolute value of the smallest non-zero eigenvalue of the normalized Laplacian of G .

2.2. Useful Theorems. There are several established theorems we use in our work.

While not explicitly used in this work, the Universal Coefficient Theorem provides the link between the homology and cohomology over \mathbb{Z} and various finite fields. Any statement about rational homology can be extended to cohomology, and vice versa. Moreover, a \mathbb{Z} -summand of $H_k(X, \mathbb{Z})$ necessarily corresponds to a \mathbb{Q} -summand of $H_k(X, \mathbb{Q})$, and any torsion will correspond to nontrivial homology of the finite field with the same number of elements. Within our work, the language of a theorem statement primarily corresponds to whichever group we worked with in the proof. Finally, we note that the vanishing of integer homology is a much stronger statement than the vanishing of rational homology.

With the definitions established we introduce the first of the two theorems instrumental in our proof of Theorem 1. We use a special case of Theorem 2.1 in a paper by Ballmann and Świątkowski [BS97].

Cohomology Vanishing Theorem. To prove vanishing cohomology we employ a result of Garland in [Gar73]. Paraphrasing [BS97, Theorem 2.1], let X be a pure D -dimensional finite simplicial complex such that for every $(D-2)$ -dimensional face σ , the link $\text{lk}_X(\sigma)$ is connected and has spectral gap

$$\lambda_2[\text{lk}_X(\sigma)] > 1 - \frac{1}{D}$$

Then $H^{D-1}(X, \mathbb{Q}) = 0$.

We note that since X is stipulated to be pure D -dimensional, the link of any $(D-2)$ -face will be of dimension 1. The spectral gaps of these link complexes are therefore well-defined.

To produce the necessary estimates on these gaps we then need the help of the main result in [HKP12], established by Hoffman, Kahle, and Paquette. We present it here as a concise statement sufficient for our needs, noting the actual result yields more general and precise results.

Spectral Gap Theorem. [HKP12, Theorem 1.1] Fix a $\delta > 0$ and let $G \stackrel{\text{dist}}{\equiv} G(n, p)$ with $p \geq \frac{(1+\delta)\log n}{n}$. Then G is connected and

$$\lambda_2(G) > 1 - o(1)$$

with probability $1 - o(n^{-\delta})$.

3. CALCULATING MAXIMAL FACES

We call a $(k - 1)$ -face in a simplicial complex *maximal* if it is not contained in any k -simplex. These subcomplexes naturally play an important role in homology, their characteristic functions generate $(k - 1)$ -cocycles. We let N_{k-1} denote the number of maximal $(k - 1)$ -faces in X . Recall our complex has vertex set $[n]$, we use $j \in \binom{[n]}{k}$ to denote a set of k vertices of $[n]$. Letting C_j be the event that the vertices of j span a maximal $(k - 1)$ -simplex, it follows that

$$N_{k-1} = \sum_{j \in \binom{[n]}{k}} 1_{C_j}.$$

Lemma 9. For any $j \in \binom{[n]}{k}$,

$$(6) \quad \mathbb{P}[C_j] = \left(\prod_{i=1}^{k-1} p_i^{\binom{k}{i+1}} \right) \left(1 - \prod_{i=1}^k p_i^{\binom{k}{i}} \right)^{n-k}.$$

Proof. The left parenthetical calculates the probability that j is in our complex. For any $1 \leq i \leq k - 1$ we need the $\binom{k}{i+1}$ possible i -faces on the vertices of j to be contained in X . Proceeding inductively, the $(i - 1)$ -skeleton of each face is already contained in X and each i -face is added independently with probability p_i . The right parenthetical calculates the probability these vertices do not form a k -face with one of the other $n - k$ vertices. For a fixed vertex v , this happens when every face of dimension $1, \dots, k$ involving v and vertices of j is contained in our complex. This event that we wish to avoid occurs independently for each vertex with probability $\prod_{i=1}^k p_i^{\binom{k}{i}}$, and our result follows. \square

We now establish the threshold where these subcomplexes do not appear in our complex.

Lemma 10. Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with $p_i = n^{-\alpha_i}$, if

$$\sum_{i=1}^k \binom{k}{i} \alpha_i < 1$$

then X w.h.p. contains no maximal $(k - 1)$ -faces.

Proof. Recall N_{k-1} counts the maximal faces in X , by (6) and linearity of expectation we have

$$\begin{aligned}
\mathbb{E}[N_{k-1}] &= \sum_{j \in \binom{[n]}{k}} \mathbb{E}[1_{C_j}] = \sum_{j \in \binom{[n]}{k}} \mathbb{P}[C_j] \\
&= \sum_{j \in \binom{[n]}{k}} \left(\prod_{i=1}^{k-1} p_i^{\binom{k}{i+1}} \right) \left(1 - \prod_{i=1}^k p_i^{\binom{k}{i}} \right)^{n-k} \\
&= \binom{n}{k} \left(\prod_{i=1}^{k-1} p_i^{\binom{k}{i+1}} \right) \left(1 - \prod_{i=1}^k p_i^{\binom{k}{i}} \right)^{n-k} \\
&\leq \frac{n^k}{k!} \left(\prod_{i=1}^{k-1} n^{-\alpha_i \binom{k}{i+1}} \right) \left(e^{-(n-k) \left(\prod_{i=1}^k n^{-\alpha_i \binom{k}{i}} \right)} \right).
\end{aligned}$$

Then for some $D > 0$,

$$\begin{aligned}
\mathbb{E}[N_{k-1}] &\leq D \frac{n^k}{k!} \left(n^{-\sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}} \right) \left(e^{-n \left(n^{-\sum_{i=1}^k \alpha_i \binom{k}{i}} \right)} \right) \\
&= \frac{D}{k!} \left(n^{k - \sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}} \right) \left(e^{-n^{1 - \sum_{i=1}^k \alpha_i \binom{k}{i}}} \right).
\end{aligned}$$

By hypothesis

$$\sum_{i=1}^{k-1} \alpha_i \binom{k}{i} < 1,$$

so the right parenthetical of our last term is e^{-n^ϵ} for some $\epsilon > 0$. This term asymptotically dominates the rest of the expression and $\mathbb{E}[N_{k-1}] \rightarrow 0$ exponentially. Markov's inequality tells us

$$\mathbb{P}[N_{k-1} \geq 1] \leq \mathbb{E}[N_{k-1}] = o(1),$$

completing our proof. □

So in this regime w.h.p. every $(k-1)$ -face of our complex is contained in a k -face, a fact necessary to utilize [BS97, Theorem 2.1] and prove that $H^{k-1}(X, \mathbb{Q}) = 0$ in this range.

4. TRIVIAL COHOMOLOGY

In this section we prove Theorem 1, the upper threshold for vanishing cohomology, with [BS97, Theorem 2.1] and [HKP12, Theorem 1.1] crucial to our argument.

To understand the $(k-1)$ -th cohomology of a complex we need only consider its k -skeleton, ie. the subcomplex of X induced by its faces of dimension k and lower. We use X_k to denote the

k -skeleton of X , observing $H^{k-1}(X_k) = H^{k-1}(X)$. The following lemma provides the first step to invoking the [BS97, Theorem 2.1].

Lemma 11. *Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ such that*

$$\sum_{i=1}^k \alpha_i \binom{k}{i} < 1$$

and X_k be its k -skeleton. Then X_k is w.h.p. pure k -dimensional.

Proof. This implies $\alpha_1 < \frac{1}{k} < 1$, so w.h.p. every vertex has degree greater than 0. Fixing a $1 \leq j \leq k-1$ we have

$$\sum_{i=1}^{j+1} \alpha_i \binom{j+1}{i} \leq \sum_{i=1}^k \alpha_i \binom{k}{i} < 1,$$

so by Lemma 10 w.h.p. every j -face of X_k is contained in a $(j+1)$ -face. Our claim follows immediately. \square

Thus X_k satisfies the first hypothesis of [BS97, Theorem 2.1]. To establish trivial cohomology we must bound the spectral gaps of the links of X_k .

4.1. Using the Spectral Gap Theorem. We wish to understand the structure of the links of the $(k-2)$ -faces in our complex. Given a $(k-2)$ -face $\sigma \in X_k$, we let L_σ denote the number of vertices in $\text{lk}_{X_k}(\sigma)$. We also define

$$\bar{p} = \prod_{i=1}^{k-1} p_i^{\binom{k-1}{i}}$$

and

$$p' = \prod_{i=1}^k p_i^{\binom{k-1}{i-1}}$$

Lemma 12. *For any $(k-2)$ -face $\sigma \in X$, L_σ has the same distribution as $\text{Bin}(n-k+1, \bar{p})$. Moreover, conditioning on the value of L_σ , $\text{lk}_{X_k}(\sigma)$ has the same distribution as $G(L_\sigma, p')$.*

Proof. Fixing a $(k-2)$ -face σ , a vertex v will be in $\text{lk}_{X_k}(\sigma)$ if X_k contains every possible face spanned by v and some subset of the vertices of σ . In dimension $1 \leq i \leq k-1$ there are $\binom{k-1}{i}$ such faces, each present with probability p_i . Distinct vertices appearing in the link are statements about disjoint sets of faces, so these events are independent with probability \bar{p} and our statement about L_σ follows.

Similarly, after conditioning upon the number of vertices in the link, the edge between two vertices of $\text{lk}_{X_k}(\sigma)$ is included when X_k contains every face of dimension $1, \dots, k$ involving those two vertices and vertices of σ . This occurs with probability p' , and the inclusion of distinct edges are again independent events. Thus $\text{lk}_{X_k}(\sigma)$ has the same distribution as $G(L_\sigma, p')$ as desired. \square

So the link of a $(k-2)$ -face behaves like an Erdős-Rényi random graph, but before applying the Spectral Gap Theorem we must bound L_σ .

Lemma 13. *Let $X \stackrel{\text{dist}}{=} X(n, p_1, p_2 \dots)$ with*

$$\sum_{i=1}^k \alpha_i \binom{k}{i} < 1,$$

then w.h.p. $n\bar{p}/2 \leq L_\sigma$ for every $(k-2)$ -face $\sigma \in X$.

Proof. For any specific $(k-2)$ -face σ and n large enough that

$$n\bar{p}/2 < 4(n-k+1)\bar{p}/7,$$

if μ denotes the mean of L_σ , then Chernoff bounds on binomial random variables give us that

$$\begin{aligned} \mathbb{P}(L_\sigma < n\bar{p}/2) &\leq \mathbb{P}(L_\sigma < 4(n-k+1)\bar{p}/7) \\ &= \mathbb{P}\left(L_\sigma < \left(1 - \frac{3}{7}\right)\mu\right) \\ &\leq e^{-\left(\frac{3}{7}\right)^2 \frac{\mu}{2}} \\ (7) \qquad &= e^{-\frac{9\mu}{98}}. \end{aligned}$$

However, these probabilities are not independent for each $(k-2)$ -face. Defining J_σ to be the indicator random variable for $\{L_\sigma < n\bar{p}/2\}$ for a $(k-2)$ -face σ , Markov's Inequality tells us

$$\mathbb{P}\left[\sum_{\sigma \in \binom{[n]}{k-1}} J_\sigma \geq 1\right] \leq \mathbb{E}\left[\sum_{\sigma \in \binom{[n]}{k-1}} J_\sigma\right] = \sum_{\sigma \in \binom{[n]}{k-1}} \mathbb{E}[J_\sigma].$$

There are at most $\binom{n}{k-1}$ faces of dimension $(k-2)$ in X and by construction $\mathbb{E}[J_\sigma] = \mathbb{P}(L_\sigma < n\bar{p}/2)$, so

$$\begin{aligned} \mathbb{P}\left[\sum_{\sigma \in \binom{[n]}{k-1}} J_\sigma \geq 1\right] &\leq \binom{n}{k-1} \mathbb{P}(L_\sigma < n\bar{p}/2) \text{ (for some fixed } \sigma) \\ &\leq \binom{n}{k-1} e^{-\frac{9\mu}{98}} \text{ (by (7))} \\ &= \binom{n}{k-1} e^{-\frac{9(n-k+1)\bar{p}}{98}} \\ &= \binom{n}{k-1} e^{-\frac{9n\bar{p}}{98}} e^{\frac{(k-1)\bar{p}}{98}}. \end{aligned}$$

Since $\alpha_i \geq 0$ for all i , we know

$$\sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i} < \sum_{i=1}^k \alpha_i \binom{k}{i} < 1$$

and so for some $\epsilon > 0$

$$\bar{p} = \prod_{i=1}^{k-1} p_i^{\binom{k-1}{i}} = n^{-\sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i}} = n^{\epsilon-1}.$$

Since $\frac{(k-1)\bar{p}}{98} \rightarrow 0$ so long as $p_i < 1$ for some $1 \leq i \leq k-1$, we may bound $e^{\frac{(k-1)\bar{p}}{98}}$ above by a constant $C > 0$. It follows that

$$\mathbb{P}\left[\left(\sum_{\sigma \in \binom{[n]}{k-1}} J_\sigma\right) \geq 1\right] \leq C \binom{n}{k-1} e^{-\frac{9}{98}n^\epsilon} = o(1).$$

Thus w.h.p. $L_\sigma = 0$ for every $(k-2)$ -face σ , completing our proof. \square

We require one last lemma before proving our main result.

Lemma 14. *Let $X \stackrel{\text{dist}}{=} X(n, p_1, p_2, \dots)$ and fix $\delta > 0$. If*

$$\sum_{i=1}^k \alpha_i \binom{k}{i} < 1$$

then w.h.p.

$$(8) \quad \frac{(1+\delta) \log L_\sigma}{L_\sigma} \leq p'$$

for all $(k-2)$ -faces σ in X .

Proof. We let $L = n\bar{p}/2$. Straightforward calculus shows $f(x) = \frac{(1+\delta)\log(x)}{x}$ is monotonically decreasing on $[e, \infty)$. For large n we have $e < L$, so if $f(L) < p'$ then by Lemma 13 $f(L_\sigma) < p'$ for all σ w.h.p. We let $\epsilon = 1 - \sum_{i=1}^k \alpha_i \binom{k}{i}$, noting $\epsilon > 0$ by hypothesis. Then

$$\begin{aligned} \frac{f(L)}{p'} &= (1 + \delta) \frac{\log L}{Lp'} \\ &\leq (2 + 2\delta) \frac{\log n}{n\bar{p}p'} \\ &= (2 + 2\delta) \frac{\log n}{n^{1 - \sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i} - \sum_1^k \alpha_i \binom{k-1}{i-1}}} \\ &= (2 + 2\delta) \frac{\log n}{n^{1 - \sum_{i=1}^k \alpha_i \binom{k}{i}}} \\ &= (2 + 2\delta) \frac{\log n}{n^\epsilon} \\ &= o(1). \end{aligned}$$

Thus w.h.p. $f(L_\sigma) < f(L) < p'$ for all $(k-2)$ -faces σ . \square

4.2. The Main Result. We now have the machinery to prove a main theorem.

Proof of Theorem 1. We begin by fixing the $\delta > 0$ we will use in the Spectral Gap Theorem in Section 2:

$$(9) \quad \delta = \frac{k - \sum_{i=1}^{k-2} \alpha_i \binom{k-1}{i+1}}{1 - \sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i}}.$$

A standard second moment technique, detailed in Section 6, tells us that if f_{k-2} denotes the number of $(k-2)$ -faces in X , or X_k , then w.h.p.

$$(10) \quad f_{k-2} \leq (1 + o(1))\mathbb{E}[f_{k-2}] = (1 + o(1)) \binom{n}{k-1} \prod_{i=1}^{k-2} p_i \binom{k-1}{i+1}.$$

By Lemma 12 each of these faces has link with distribution $G(L_\sigma, p')$, and by Lemma 14 w.h.p.

$$\frac{(1 + \delta) \log L_\sigma}{L_\sigma} < p'$$

for all $(k-2)$ -faces σ of X . Thus by the Spectral Gap Theorem the probability P_σ that

$$\lambda_2[\text{lk}_{X_k}(\sigma)] < 1 - 1/k$$

is $o(L_\sigma^{-\delta})$. Let P_X denote the probability there exists any $(k-2)$ -face whose link in X_k has spectral gap less than $1 - 1/k$. We apply a union bound over all $(k-2)$ -faces to see

$$\begin{aligned} P_X &\leq \sum_{\sigma \in \binom{[n]}{k-1}} P_\sigma \\ &= \sum_{\sigma \in \binom{[n]}{k-1}} o(L_\sigma^{-\delta}) \\ &\leq \sum_{\sigma \in \binom{[n]}{k-1}} o\left(\left(\frac{n\bar{p}}{2}\right)^{-\delta}\right). \end{aligned}$$

The last line holds since w.h.p. $n\bar{p}/2 < L_\sigma$, so $L_\sigma^{-\delta} < (n\bar{p}/2)^{-\delta}$. By (10),

$$\begin{aligned} P_X &\leq (1 + o(1)) \binom{n}{k-1} \left(\prod_{i=1}^{k-2} p_i^{\binom{k-1}{i+1}} \right) o\left(2^\delta (n\bar{p})^{-\delta}\right) \\ &\leq (1 + o(1)) \frac{n^{k-1}}{(k-1)!} \left(\prod_{i=1}^{k-2} p_i^{\binom{k-1}{i+1}} \right) o\left(2^\delta (n\bar{p})^{-\delta}\right) \\ &= O\left(2^\delta n^{k-1} n^{-\sum_{i=1}^{k-2} \alpha_i \binom{k-1}{i+1}} \left(n \cdot n^{-\sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i}}\right)^{-\delta}\right) \\ &= O\left(2^\delta n^{k-1 - \sum_{i=1}^{k-2} \alpha_i \binom{k-1}{i+1}} n^{-\delta(1 - \sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i})}\right). \end{aligned}$$

By our choice of δ in (9),

$$\begin{aligned} P_X &= O\left(n^{k-1 - \sum_{i=1}^{k-2} \alpha_i \binom{k-1}{i+1}} n^{-(k - \sum_{i=1}^{k-2} \alpha_i \binom{k-1}{i+1})}\right) \\ &= O(n^{-1}) \\ &= o(1). \end{aligned}$$

Thus w.h.p.

$$\lambda_2[\text{lk}_{X_k}(\sigma)] > 1 - \frac{1}{k}$$

for every $(k-2)$ -face σ of X . Combining this with Lemma 10, we may apply the Cohomology Vanishing Theorem in Section 2 on X_k to conclude that w.h.p. $H^{k-1}(X_k, \mathbb{Q}) = 0$. Noting that $H^{k-1}(X_k, \mathbb{Q}) \cong H^{k-1}(X, \mathbb{Q})$ completes our proof. \square

5. NONTRIVIAL HOMOLOGY: BOUNDARIES OF SIMPLICES

In this Section we consider the case

$$1 \leq \sum_{i=1}^k \alpha_i \binom{k}{i}, \sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1} < k+1, \text{ and } \alpha_k > 0, \text{ so } p_k \neq 1$$

to prove the second half of Theorem 2.

In [ALLM13], the threshold for the appearance of nontrivial k -homology in $Y_k(n, p)$ was studied with a stochastic k -face adding process. It was shown that under this process of adding faces uniform randomly, the first type of homological k -cycle to appear was either an *empty k -simplex*, the $(k-1)$ -skeleton of a k -simplex, or a cycle supported on a positive fraction of the total number of k -faces. In this section we consider the first case to $X(n, p_1, p_2, \dots)$, when $p_k \neq 1$ and the presence of an empty k -simplex is possible. If X contains an empty k -simplex with at least one maximal $(k-1)$ -face, then it generates a \mathbb{Z} -summand in $H_{k-1}(X, \mathbb{Z})$. For a set of $k+1$ vertices $j \in \binom{[n]}{k+1}$, we define A_j as the event j corresponds to an empty k -simplex with first $(k-1)$ -face, determined by lexicographic order, maximal in X . Letting M_{k-1} denote the total number of such subcomplexes in X , it follows that

$$M_{k-1} = \sum_{j \in \binom{[n]}{k+1}} 1_{A_j}.$$

We then calculate the probability of A_j .

Lemma 15. *For any $j \in \binom{[n]}{k+1}$,*

$$(11) \quad \mathbb{E}[1_{A_j}] = \mathbb{P}[A_j] = \left(\prod_{i=1}^{k-1} p_i \binom{k+1}{i+1} \right) (1-p_k) \left(1 - \prod_{i=1}^k p_i \binom{k}{i} \right)^{n-k-1}.$$

Proof. The first term calculates the probability that X contains the necessary i -faces for $i < k$: we need every subset of $i+1$ vertices of j to form an i -simplex. The second term is the requirement that the associated k -simplex is empty. The last term is ensuring our first $(k-1)$ -face is maximal, or does not form a k -simplex with any of the remaining $n-k-1$ vertices. This occurs independently with probability $\prod_{i=1}^k p_i \binom{k}{i}$ for each vertex. \square

We note that narrowing our consideration to when the first $(k-1)$ -face is maximal simplifies the calculations without altering the relevant probability thresholds. Recall that we say $X \sim Y$

with high probability if for all $\epsilon > 0$, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}[(1 - \epsilon) \leq Y/X \leq (1 + \epsilon)] \rightarrow 1.$$

Lemma 16. *Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with $p_i = n^{-\alpha_i}$ and M_{k-1} count the number of empty k -simplices in X with maximal first $(k-1)$ -face. If*

$$1 \leq \sum_{i=1}^k \alpha_i \binom{k}{i}, \quad \sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1} < k+1, \quad \text{and } p_k \neq 1$$

then w.h.p. $M_{k-1} > 0$ and $M_{k-1} \sim \mathbb{E}[M_{k-1}]$.

Proof. By linearity of expectation we have

$$\begin{aligned} \mathbb{E}[M_{k-1}] &= \binom{n}{k+1} \left(\prod_{i=1}^{k-1} p_i^{\binom{k+1}{i+1}} \right) (1 - p_k) \left(1 - \prod_{i=1}^k p_i^{\binom{k}{i}} \right)^{n-k-1} \\ &\approx \frac{1 - p_k}{(k+1)!} \left(n^{k+1 - \sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1}} \right) \left(e^{-n^{1 - \sum_{i=1}^k \alpha_i \binom{k}{i}}} \right). \end{aligned}$$

By the first inequality in our hypothesis,

$$1 - \sum_{i=1}^k \alpha_i \binom{k}{i} \leq 0,$$

which implies that

$$e^{-n^{1 - \sum_{i=1}^k \alpha_i \binom{k}{i}}} = \Theta(1).$$

We therefore have that

$$(12) \quad \mathbb{E}[M_{k-1}] = \Theta \left(n^{k+1 - \sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1}} \right),$$

hence $\mathbb{E}[M_{k-1}] \rightarrow \infty$.

This, along with a straightforward second moment argument (see e.g. [Kah14a]) which is detailed in Appendix A, allow us to use Chebyshev's inequality to conclude that w.h.p. $M_{k-1} \sim \mathbb{E}[M_{k-1}]$. \square

Proof of the second part of Theorem 2. It follows from Lemma 16 that w.h.p. $M_{k-1} > 0$. Consider such a subcomplex σ . As the boundary of a k -simplex, a signed sum of its $(k-1)$ -faces is in the kernel of the $(k-1)$ -boundary map. Since one of these faces, τ , is maximal, it is not contained in a k -face of X . Thus no $(k-1)$ -chain with a non-zero coefficient of τ can be a $(k-1)$ -boundary

of X . Thus we have a non-trivial cycle no multiple of which is a boundary, contributing a \mathbb{Z} -summand to $H_{k-1}(X, \mathbb{Z})$ and a \mathbb{Q} -cycle to $H_{k-1}(X, \mathbb{Q})$. By vector space duality we conclude $H^{k-1}(X, \mathbb{Q}) \cong H_{k-1}(X, \mathbb{Q}) \neq 0$. \square

6. NONTRIVIAL COHOMOLOGY: BETTI NUMBERS ARGUMENT

We now consider when

$$1 < \sum_{i=1}^k \alpha_i \binom{k}{i} \text{ and } \sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i} < 1,$$

proving the other half of Theorem 2.

Proof of the first part of Theorem 2. For $X \stackrel{\text{dist}}{=} X(n, p_1, p_2, \dots)$, with the aforementioned conditions on p_i , we let f_i denote the number of i -simplices in X and $\beta_i = \dim H^i(X, \mathbb{Q})$. Linear algebra considerations tell us

$$(13) \quad f_{k-1} \geq \beta_{k-1} \geq f_{k-1} - f_k - f_{k-2}.$$

Thus showing that w.h.p. $f_{k-1} > f_k + f_{k-2}$ implies $\beta_{k-1} > 0$. We begin by calculating the expected number of faces in these dimensions:

$$\begin{aligned} \mathbb{E}[f_{k-2}] &= \binom{n}{k-1} \prod_{i=1}^{k-2} p_i^{\binom{k-1}{i+1}} \\ \mathbb{E}[f_{k-1}] &= \binom{n}{k} \prod_{i=1}^{k-1} p_i^{\binom{k}{i+1}} \\ \mathbb{E}[f_k] &= \binom{n}{k+1} \prod_{i=1}^k p_i^{\binom{k+1}{i+1}}. \end{aligned}$$

By linearity of expectation

$$\mathbb{E}[f_{k-1}] \geq \mathbb{E}[\beta_{k-1}] \geq \mathbb{E}[f_{k-1}] - \mathbb{E}[f_{k-2}] - \mathbb{E}[f_k].$$

Comparing the expectations in different dimensions we see

$$(14) \quad \frac{\mathbb{E}[f_k]}{\mathbb{E}[f_{k-1}]} = \frac{n-k}{k+1} \prod_{i=1}^k p_i^{\binom{k+1}{i+1} - \binom{k}{i+1}} = \frac{n-k}{k+1} \prod_{i=1}^k p_i^{\binom{k}{i}} \leq n \prod_{i=1}^k p_i^{\binom{k}{i}} = o(1),$$

because

$$\prod_{i=1}^k p_i^{\binom{k}{i}} < n^{-1}$$

by hypothesis. Similarly, since

$$\prod_{i=1}^{k-1} p_i^{\binom{k-1}{i}} = n^{c-1}$$

for some $c > 0$, we have

$$(15) \quad \frac{\mathbb{E}[f_{k-2}]}{\mathbb{E}[f_{k-1}]} = \frac{k}{n-k+1} \prod_{i=1}^{k-1} p_i^{\binom{k-1}{i+1} - \binom{k}{i+1}} = \frac{k}{n-k+1} \prod_{i=1}^{k-1} p_i^{-\binom{k-1}{i}} = \frac{kn^{1-c}}{n-k+1} = o(1).$$

Thus $\mathbb{E}[f_{k-1}]$ asymptotically dominates the other two terms.

Letting

$$\tilde{f}_{k-1} := f_{k-1} - f_k - f_{k-2},$$

it follows from (14) and (15) that

$$(16) \quad \mathbb{E}[\tilde{f}_{k-1}] \sim \mathbb{E}[\beta_{k-1}] \sim \mathbb{E}[f_{k-1}].$$

To prove stronger statements about β_{k-1} we again make use of Chebyshev's Inequality. That is, if Z is a random variable with $\mathbb{E}[Z] \rightarrow \infty$ and $\text{Var}[Z] = o(\mathbb{E}[Z]^2)$, then w.h.p. $Z \sim \mathbb{E}[Z]$.

Now

$$\begin{aligned} \text{Var}[f_{k-1}] &= \mathbb{E}[f_{k-1}^2] - \mathbb{E}[f_{k-1}]^2 \\ &= \mathbb{E}[f_{k-1}^2] - \binom{n}{k}^2 \left(\prod_{i=1}^{k-1} p_i^{2\binom{k}{i+1}} \right). \end{aligned}$$

It remains to calculate $\mathbb{E}[f_{k-1}^2]$. For any $j \in \binom{[n]}{k}$ let E_j be the event that the vertices of j span a $(k-1)$ -face in X . Then

$$\mathbb{E}[f_{k-1}^2] = \sum_{j,l \in \binom{[n]}{k}} \mathbb{P}[E_j \cap E_l] = \binom{n}{k} \sum_{l \in \binom{[n]}{k}} \mathbb{P}[E_j \cap E_l].$$

The second equality follows by symmetry and fixing some set of vertices j , say $\{1, \dots, k\}$. We proceed by grouping the l according to $|j \cap l|$. Through this approach we see

$$\begin{aligned} \mathbb{E}[f_{k-1}^2] &= \binom{n}{k} \sum_{l \in \binom{[n]}{k}} \mathbb{P}[A_j \cap A_l] \\ &= \binom{n}{k} \sum_{m=0}^k \binom{k}{m} \binom{n-k}{k-m} \left(\prod_{i=1}^{k-1} p_i^{2\binom{k}{i+1} - \binom{m}{i+1}} \right) \\ &= \binom{n}{k} \prod_{i=1}^{k-1} p_i^{2\binom{k}{i+1}} \left(\sum_{m=0}^k \binom{k}{m} \binom{n-k}{k-m} \prod_{i=1}^{m-1} p_i^{-\binom{m}{i+1}} \right). \end{aligned}$$

We pull the $m = 0$ term out of the summation and use $\binom{n-k}{k} < \binom{n}{k}$ to see

$$\mathbb{E}[f_{k-1}^2] \leq \mathbb{E}[f_{k-1}]^2 + \binom{n}{k} \prod_{i=1}^{k-1} p_i^{2\binom{k}{i+1}} \left(\sum_{m=1}^k \binom{k}{m} \binom{n-k}{k-m} \prod_{i=1}^{m-1} p_i^{-\binom{m}{i+1}} \right).$$

We observe

$$\begin{aligned} \frac{\text{Var}[f_{k-1}]}{\mathbb{E}[f_{k-1}]^2} &\leq \frac{\binom{n}{k} \prod_{i=1}^{k-1} p_i^{2\binom{k}{i+1}} \left(\sum_{m=1}^k \binom{k}{m} \binom{n-k}{k-m} \prod_{i=1}^{m-1} p_i^{-\binom{m}{i+1}} \right)}{\binom{n}{k}^2 \left(\prod_{i=1}^{k-1} p_i^{2\binom{k}{i+1}} \right)} \\ &= \frac{\sum_{m=1}^k \binom{k}{m} \binom{n-k}{k-m} \prod_{i=1}^{m-1} p_i^{-\binom{m}{i+1}}}{\binom{n}{k}} \\ &= \sum_{m=1}^k O \left(n^{-m} \prod_{i=1}^{m-1} p_i^{-\binom{m}{i+1}} \right) \\ &= o(1). \end{aligned}$$

The final line holds from our hypotheses since

$$\sum_{i=1}^{m-1} \alpha_i \binom{m}{i+1} \leq \frac{m}{k} \left(\sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1} \right) \leq \frac{m}{k} \left(k \cdot \sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i} \right) < m,$$

so for $m = 1, \dots, k$,

$$\prod_{i=1}^{m-1} p_i^{-\binom{m}{i+1}} = n^{\sum_{i=1}^{m-1} \alpha_i \binom{m}{i+1}} = o(n^m).$$

We conclude $f_{k-1} \sim \mathbb{E}[f_{k-1}]$.

We note that nothing in the above argument is unique to f_{k-1} , so w.h.p. $-f_{k-2} \sim \mathbb{E}[-f_{k-2}]$ and $-f_k \sim \mathbb{E}[-f_k]$. By linearity of expectation $\tilde{f}_{k-1} \sim \mathbb{E}[\tilde{f}_{k-1}]$, then from (13) and (16) we conclude that w.h.p. $\beta_{k-1} \sim \mathbb{E}[\beta_{k-1}] \sim f_{k-1}$. Thus $\beta_{k-1} = \dim(H^{k-1}(X, \mathbb{Q})) \neq 0$ w.h.p., which completes our proof. \square

Under these conditions we have proven a stronger result than nontrivial homology.

Proof of Theorem 6. From the above,

$$\beta_{k-1} \sim f_{k-1} \sim \mathbb{E}[f_{k-1}] = g_{k-1}(n, \alpha_1, \dots, \alpha_k).$$

The result follows. \square

Our proof also shows that allowing

$$\sum_{i=1}^k \alpha_i \binom{k}{i} = 1$$

still ensures nontrivial cohomology.

Lemma 17. *If*

$$\sum_{i=1}^k \alpha_i \binom{k}{i} = 1,$$

then w.h.p. $H^{k-1}(X, \mathbb{Q}) \neq 0$ and

$$\beta_{k-1} \geq \left(\frac{k}{k+1} \right) f_{k-1}.$$

Proof. We first calculate

$$\begin{aligned} \frac{\mathbb{E}[f_k]}{\mathbb{E}[f_{k-1}]} &= \frac{n-k}{k+1} \prod_{i=1}^k p_i^{(k)} \\ &= \frac{n-k}{n(k+1)} \\ &\approx \frac{1}{k+1}. \end{aligned}$$

The machinery established in the previous section then does the work for us. Since β_{k-1} is bounded between f_{k-1} and $f_{k-1} - f_k - f_{k-2}$, with $f_{k-1} \sim \mathbb{E}[f_{k-1}]$ and $(f_{k-1} - f_k - f_{k-2}) \sim \mathbb{E}[(f_{k-1} - f_k - f_{k-2})] \sim \left(\frac{k}{k+1} \right) \mathbb{E}[f_{k-1}]$, our result follows immediately. \square

7. BEHAVIOR AT THE BOUNDARY

In this section we explore the behavior of the $(k-1)$ -th cohomology of $X(n, p_1, p_2, \dots)$ at the upper threshold line. Specifically, we refine the parameters of our p_i to elicit some interesting behavior and prove Theorem 3.

7.1. Maximal faces. To get the threshold for maximal faces, and thus trivial cohomology, we must slightly refine our model. Unfortunately there is no concise way to categorize these p_i . We consider when

$$p_i = (\rho_1 \log n + \rho_2 \log \log n + c)^{\nu_i} n^{-\alpha_i}$$

for some constants ν_i, ρ_1, ρ_2 , and c , with

$$\sum_{i=1}^k \alpha_i \binom{k}{i} = 1.$$

It follows that

$$\begin{aligned}
\mathbb{E}[N_{k-1}] &\approx \frac{n^k}{k!} \left(\prod_{i=1}^{k-1} p_i^{\binom{k}{i+1}} \right) \left(e^{-n \left(\prod_{i=1}^k p_i^{\binom{k}{i}} \right)} \right) \\
&= \frac{n^{k - \sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}}}{k!} \left[\prod_{i=1}^{k-1} ((\rho_1 + o(1)) \log n)^{\nu_i \binom{k}{i+1}} \right] e^{-\prod_{i=1}^k (\rho_1 \log n + \rho_2 \log \log n + c)^{\nu_i \binom{k}{i}}} \\
&= \frac{n^{k - \sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}}}{k!} ((\rho_1 + o(1)) \log n)^{\sum_{i=1}^{k-1} \nu_i \binom{k}{i+1}} e^{-(\rho_1 \log n + \rho_2 \log \log n + c)^{\sum_{i=1}^k \nu_i \binom{k}{i}}}.
\end{aligned}$$

Letting

$$\sum_{i=1}^k \nu_i \binom{k}{i} = 1,$$

we have

$$\begin{aligned}
\mathbb{E}[N_{k-1}] &\approx \frac{n^{k - \sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}}}{k!} ((\rho_1 + o(1)) \log n)^{\sum_{i=1}^{k-1} \nu_i \binom{k}{i+1}} e^{-(\rho_1 \log n + \rho_2 \log \log n + c)} \\
&= \frac{n^{k - \sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}}}{k!} ((\rho_1 + o(1)) \log n)^{\sum_{i=1}^{k-1} \nu_i \binom{k}{i+1}} n^{-\rho_1} (\log n)^{-\rho_2} e^{-c}.
\end{aligned}$$

If we set

$$\rho_1 = k - \sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}$$

and

$$\rho_2 = \sum_{i=1}^{k-1} \nu_i \binom{k}{i+1},$$

then

$$(17) \quad \mathbb{E}[N_{k-1}] \rightarrow \frac{\rho_1^{\rho_2} e^{-c}}{k!}$$

as $n \rightarrow \infty$. We then establish the following result.

Lemma 18. *Let $X \stackrel{dist}{=} X(n, p_1, p_2, \dots)$ with*

$$p_i = (\rho_1 \log n + \rho_2 \log \log n + c)^{\nu_i} n^{-\alpha_i}$$

such that

$$\rho_1 = k - \sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}$$

and

$$\rho_2 = \sum_{i=1}^{k-1} \nu_i \binom{k}{i+1}.$$

If

$$\sum_{i=1}^k \alpha_i \binom{k}{i} = 1 = \sum_{i=1}^k \nu_i \binom{k}{i},$$

then N_{k-1} the number of maximal $(k-1)$ -faces in X approaches a Poisson distribution

$$N_{k-1} \rightarrow \text{Poi}(\mu)$$

with mean

$$\mu = \frac{\rho_1^{\rho_2} e^{-c}}{k!}.$$

Proof. We prove this with through a standard factorial moment argument, found in Appendix B. \square

7.2. Betti Numbers. At criticality, if we condition on the event $N_{k-1} = 0$ then slightly modifying our arguments in Section 4 will show $H^{k-1}(X, \mathbb{Q}) = 0$ w.h.p. This enables us to use the limiting distribution of N_{k-1} to prove an identical result for β_{k-1} .

Proof of Theorem 3. From Lemma 18 we know that given these hypotheses $N_{k-1} \rightarrow \text{Poi}(\mu)$. We suppose $N_{k-1} = m$ for some $m \in \mathbb{Z}$. The characteristic functions of these m maximal faces are $(k-1)$ -cocycles. We show these cocycles are not coboundaries, and in fact constitute the only cohomological cocycles of dimension $k-1$ in X .

We label these faces $\sigma_1, \dots, \sigma_m$ and their respective characteristic functions ϕ_1, \dots, ϕ_m . Letting R_{k-2} count the number of $(k-2)$ -faces of X contained in m or fewer $(k-1)$ -faces, we have

$$\begin{aligned} \mathbb{E}[R_{k-2}] &= \binom{n}{k-1} \prod_{i=1}^{k-2} p_i^{\binom{k-1}{i+1}} \left(\sum_{j=0}^m \binom{n-k+1}{j} \left(\prod_{i=1}^{k-1} p_i^{\binom{k-1}{i}} \right)^j \left(1 - \prod_{i=1}^{k-1} p_i^{\binom{k-1}{i}} \right)^{n-k+1-j} \right) \\ &= o(e^{-n^{-\epsilon}}) \text{ for some } \epsilon > 0. \end{aligned}$$

This holds since by our hypotheses

$$n \left(\prod_{i=1}^{k-1} p_i^{\binom{n-1}{i}} \right) \rightarrow \infty,$$

so the right-most term is exponentially decaying and dominates the expression.

Therefore w.h.p. X contains no $(k-2)$ -face contained solely in some combination of our σ_i . We now suppose there exists some $(k-2)$ -cochain λ such that $\delta^{k-2}(\lambda) = \sum_{i=1}^m a_i \phi_i$ with $a_i \neq 0$ for some i . It follows that λ is not a $(k-2)$ -coboundary. We now consider the subcomplex $X' = X - \{\sigma_1, \dots, \sigma_m\}$, and observe $R_{k-2} = 0$ implies that X' has no maximal $(k-2)$ -faces. Since $\sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i} < 1$, it follows Theorem 1 that w.h.p. $H^{k-2}(X', \mathbb{Q}) = 0$. But $\delta^{k-2}(\lambda) = 0$ in X' and λ

isn't a coboundary in X or X' , yielding a contradiction. Therefore no such λ exists and we conclude each ϕ_i generates a unique nontrivial cocycle in $H^{k-1}(X, \mathbb{Q})$.

To show these cochains are the only contributors to cohomology we again consider X' . By construction X' has no maximal $(k-1)$ -faces, and a reworking of our proof of Theorem 1 (primarily refining our estimate in Lemma 13 to show Lemma 14 still holds) tells us $H^{k-1}(X', \mathbb{Q}) = 0$ w.h.p. It follows that $H^{k-1}(X, \mathbb{Q}) \cong \mathbb{Q}^m$. \square

Implicit in our proof is the result that when

$$\sum_{i=1}^k \alpha_i \binom{k}{i} = 1,$$

the presence of maximal $(k-1)$ -faces is a necessary and sufficient condition for $H^{k-1}(X, \mathbb{Q}) \neq 0$.

8. THE PHASE TRANSITION FOR HOMOLOGY APPEARING

In this section we prove Theorem 4. The requirement

$$k+1 < \sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1}$$

is exactly the condition that our complex will w.h.p. not contain the boundary of a k -simplex. Logic dictates that, as the first $(k-1)$ -cycle to appear, the threshold for the presence of these subcomplexes should provide a lower bound for trivial homology. We proceed by verifying this intuition, using the fact that minimal homological cycles have bounded vertex support. After establishing these points we may apply a union bound to conclude our result.

8.1. Cycles of small vertex support. We begin with a few definitions identical to those in Section 5 of [Kah09]. For a $(k-1)$ -chain C the *support* of C is the union of $(k-1)$ -faces with non-zero coefficients in C , while the *vertex support* is the underlying vertex set of the support. A pure $(k-1)$ -dimensional subcomplex K is *strongly connected* if every pair of $(k-1)$ -faces $\sigma, \tau \in K^{k-1}$ can be connected by a sequence of faces $\sigma = \sigma_0, \sigma_1, \dots, \sigma_j = \tau$ such that $\dim(\sigma_i \cap \sigma_{i+1}) = k-2$ for $0 \leq i \leq j-1$. Every $(k-1)$ -cycle is a linear combination of $(k-1)$ -cycles with strongly connected support.

Lemma 19. *Let*

$$1 < \sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i}$$

and fix D such that

$$\frac{k - \sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}}{\sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i} - 1} < D.$$

Then w.h.p. all strongly connected pure $(k-1)$ -dimensional subcomplexes of X have fewer than $D+k$ vertices in their support.

Proof. Let K be such a subcomplex, since it is strongly connected we may order its faces f_1, f_2, \dots, f_m where each face f_j , for $j > 1$, has $(k-2)$ -dimensional intersection with at least one f_l with $l < j$. This induces an ordering on the supporting vertices v_1, \dots, v_s by looking at the vertex supports of $f_1, f_1 \cup f_2, f_1 \cup f_2 \cup f_3, \dots$. Thus each vertex after v_k corresponds to the addition of a $(k-1)$ -face f_j , along with the $\binom{k-1}{i}$ i -dimensional faces of f_j that include this vertex (and hence were not contained in $f_1 \cup \dots \cup f_{j-1}$).

If K has $D+k$ vertices, it follows that there are at least

$$\binom{k}{i+1} + D \binom{k-1}{i}$$

i -dimensional faces for each $1 \leq i \leq k-1$. Now either X w.h.p. contains no $(k-1)$ -simplices, in which case the result is trivial, or

$$\prod_{i=1}^{k-1} p_i \binom{k}{i+1} = \prod_{i=1}^{k-1} n^{-\alpha_i \binom{k}{i+1}} = n^{-k+\beta}$$

for some $\beta > 0$. By hypothesis

$$\prod_{i=1}^{k-1} p_i \binom{k-1}{i} = \prod_{i=1}^{k-1} n^{-\alpha_i \binom{k-1}{i}} = n^{-1-\epsilon}$$

for some $\epsilon > 0$. We choose D such that $\beta < D\epsilon$ and let A_K denote the event that X contains a subcomplex isomorphic to K . We apply a union bound on the probability of A_K as follows,

$$\begin{aligned} \mathbb{P}(A_K) &\leq (D+k)! \binom{n}{D+k} \prod_{i=1}^{k-1} p_i^{\binom{k}{i+1} + D \binom{k-1}{i}} \\ &= (D+k)! \binom{n}{D+k} n^{(-k+\beta) - D(1+\epsilon)} \\ &\leq n^{D+k} n^{-(D+k)} n^{\beta - D\epsilon} \\ &= n^{\beta - D\epsilon} \\ &= o(1). \end{aligned}$$

The last line holds by our choice of D .

As there are finitely many isomorphism classes of strongly connected $(k-1)$ -complexes on $D+k$ vertices, a union bound shows that w.h.p. none of them are subcomplexes of X . We complete our proof by observing that any such complex with more vertices must contain a strongly connected subcomplex on $D+k$ vertices. For example, under the ordering of the faces and vertices defined at the beginning of the proof, the subcomplex induced by the first $D+k$ vertices must be strongly connected. \square

8.2. The threshold for a simplex boundary. Here we prove our lower threshold for vanishing homology, which is sharp when $p_k \neq 1$.

Proof of Theorem 4. We consider some non-trivial $(k-1)$ -cycle γ with strongly connected support and K , its induced subcomplex in X . By our hypothesis we have that

$$k+1 < \sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1},$$

and either X will w.h.p. contain no $(k-1)$ -faces, making the result trivial, or

$$\sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1} < k.$$

Moreover,

$$\begin{aligned} \sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i} &= \sum_{i=1}^{k-1} \alpha_i \frac{i+1}{k} \binom{k}{i+1} \\ &= \sum_{i=1}^{k-1} \alpha_i \frac{i+1}{k} \frac{k-i}{k+1} \binom{k+1}{i+1} \\ &\geq \frac{1}{k+1} \sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1} > 1. \end{aligned}$$

Thus we may invoke Lemma 19 to conclude K is w.h.p. supported on less than $D+k$ vertices. As in that proof, we may order the vertices v_1, \dots, v_{k+m} for some $m < D$. We prove our result by removing one vertex at a time from K and counting the faces containing it that must also be removed.

Since we have a non-trivial cycle every vertex is contained in **at least** k faces of dimension $k-1$. Removing v_{k+m} first, we observe the fewest faces are removed if v_{k+m} is contained in exactly k faces of dimension $k-1$. In this case we then remove $\binom{k}{i}$ i -dimensional faces for each i . We then

remove vertices $v_{k+m-1}, \dots, v_{k+1}$, and by construction each one was contained in a $(k-1)$ -face comprised exclusively of vertices before it, so at each removal step we remove at least that simplex. Thus at each removal we account for at least $\binom{k-1}{i}$ i -faces for each i . The last k vertices correspond to our initial $(k-1)$ -simplex. Putting this together we get a lower bound on the probability of a subcomplex isomorphic to K appearing. Letting B_K denote the probability X contains such a subcomplex,

$$\begin{aligned} \mathbb{P}(B_K) &\leq (k+m)! \binom{n}{k+m} \left(\prod_{i=1}^{k-1} p_i^{\binom{k}{i}} \right) \left(\prod_{i=1}^{k-1} p_i^{\binom{k-1}{i}} \right)^{m-1} \left(\prod_{i=1}^{k-1} p_i^{\binom{k}{i+1}} \right) \\ &\leq n^{k+m} \left(\prod_{i=1}^{k-1} p_i^{\binom{k+1}{i+1}} \right) \left(\prod_{i=1}^{k-1} p_i^{\binom{k-1}{i}} \right)^{m-1} \\ &\leq \left(n^{k+1} \prod_{i=1}^{k-1} p_i^{\binom{k+1}{i+1}} \right) \left(n^{m-1} \prod_{i=1}^{k-1} p_i^{\binom{k-1}{i}} \right)^{m-1} \\ &= o(1). \end{aligned}$$

The last line holds since

$$k+1 < \sum_{i=1}^{k-1} \alpha_i \binom{k+1}{i+1} \text{ and } 1 < \sum_{i=1}^{k-1} \alpha_i \binom{k-1}{i}.$$

As there are finitely many isomorphism types of strongly connected $(k-1)$ -complexes on less than $D+k$ vertices, we may apply this argument to each of them and apply a union bound to conclude that w.h.p. none of them are subcomplexes of X . Thus we w.h.p. have no non-trivial $(k-1)$ -cycles, and $H_{k-1}(X, \mathbb{Z}) = 0$. \square

APPENDIX A. BOUNDARIES OF SIMPLICES

Proof of Lemma 16. We consider the case

$$1 \leq \sum_{l=1}^k \binom{k}{l} \alpha_l$$

where (from (12) in Section 5) we have that $\mathbb{E}[M_{k-1}] \rightarrow \infty$. By Chebyshev's inequality,

$$\mathbb{P}\left[|M_{k-1} - \mathbb{E}[M_{k-1}]| \geq \mathbb{E}[M_{k-1}]\right] \leq \frac{\text{Var}[M_{k-1}]}{\mathbb{E}[M_{k-1}]^2}.$$

Thus if we can show $\text{Var}[M_{k-1}] = o(\mathbb{E}[M_{k-1}]^2)$, then we may conclude

$$\mathbb{P}[M_{k-1} > 0] \rightarrow 1.$$

Considering M_{k-1} as a sum of indicator random variables,

$$\begin{aligned} \text{Var}[M_{k-1}] &\leq \mathbb{E}[M_{k-1}] + \sum_{i,j \in \binom{[n]}{k}} \text{Cov}[1_{A_i}, 1_{A_j}] \\ &= \mathbb{E}[M_{k-1}] + \sum_{i,j \in \binom{[n]}{k}} (\mathbb{P}[A_i \cap A_j] - \mathbb{P}[A_i]\mathbb{P}[A_j]). \end{aligned}$$

Clearly $\mathbb{E}[M_{k-1}] = o(\mathbb{E}[M_{k-1}]^2)$, to handle the sum we consider pairs $i, j \in \binom{[n]}{k+1}$ and break them into 3 cases depending on $I = |i \cap j|$. To make the calculations more readable we introduce some useful notation, defining η_k to be

$$\eta_k = (1 - p_k) \prod_{l=1}^{k-1} p_l^{\binom{k+1}{l+1}},$$

the probability that our complex contains the unfilled boundary of a specific k -simplex. We define γ_k as

$$\gamma_k = \prod_{l=1}^k p_l^{\binom{k}{l}},$$

the probability that a fixed $(k-1)$ -face and vertex form a k -simplex.

A.1. $I = 0$. We begin by calculating $\mathbb{P}[A_i \cap A_j]$. The probability that both boundaries are in our complex but unfilled is η_k^2 . By inclusion-exclusion principles the probability that neither σ_i nor σ_j , the associated first $(k-1)$ -faces of these subcomplexes, form a k -simplex with a vertex outside of $i \cup j$ is $1 - 2\gamma_k + \gamma_k^2$, and there are $n - 2k - 2$ such vertices. Finally, we must have that no k -face is formed between one subcomplex and a single vertex of the other. While this probability can be explicitly calculated, every term that isn't 1 will contain a copy of γ_k , so this probability is $1 - O(\gamma_k)$. Thus

$$\mathbb{P}[A_i \cap A_j] = \eta_k^2 (1 - 2\gamma_k + \gamma_k^2)^{n-2k-2} (1 - O(\gamma_k)),$$

and by (11) in Section 5 we know

$$\begin{aligned}
\mathbb{P}[A_i]\mathbb{P}[A_j] &= \left(\eta_k (1 - \gamma_k)^{n-k-1} \right)^2 \\
&= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2)^{n-k-1} \\
&= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2)^{n-2k-2} (1 - 2\gamma_k + \gamma_k^2)^{k+1} \\
&= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2)^{n-2k-2} (1 - O(\gamma_k)).
\end{aligned}$$

Thus

$$\mathbb{P}[A_i \cap A_j] - \mathbb{P}[A_i]\mathbb{P}[A_j] = \eta_k^2 (1 - 2\gamma_k + \gamma_k^2)^{n-2k-2} O(\gamma_k)$$

and there are $O(n^{2k+2})$ such pairs i, j , so the overall contribution of these pairs to our sum is

$$\begin{aligned}
S_0 &= O\left(n^{2k+2} \eta_k^2 (1 - 2\gamma_k + \gamma_k^2)^{n-2k-2} \gamma_k \right) \\
&= O\left(n^{2k+2} \eta_k^2 (1 - 2\gamma_k + \gamma_k^2)^{n-k-1} \gamma_k \right).
\end{aligned}$$

The second equality holds by restricting our consideration to $n > k$, then $\gamma_k \leq n^{-1} < k^{-1}$ and there is some $C > 0$ such that

$$(1 - 2\gamma_k + \gamma_k^2)^{k+1} > (1 - 2\gamma_k)^{k+1} > (1 - 2k^{-1})^k > C,$$

so removing this term does not affect our big- O calculations.

Since

$$\mathbb{E}[M_{k-1}]^2 = \binom{n}{k+1}^2 \eta_k^2 (1 - \gamma_k)^{2(n-k-1)} = O\left(n^{2k+2} \eta_k^2 (1 - 2\gamma_k + \gamma_k^2)^{n-k-1} \right)$$

and $\gamma_k \rightarrow 0$ we conclude

$$\frac{S_0}{\mathbb{E}[M_{k-1}]^2} = O(\gamma_k) = o(1).$$

Hence the contribution of these pairs to the variance is seen to be $o(\mathbb{E}[M_{k-1}]^2)$.

A.2. $I = 1$. The probability of both subcomplexes being in X is again η_k^2 since the two don't share a face of dimension greater than 0. We again use inclusion-exclusion to calculate the probability that σ_i and σ_j don't form k -simplices with another vertex. However, these faces may or may not both contain the shared vertex: if they don't then the calculations are identical to above, so we assume the alternative. In this case the two k -faces formed with some new vertex would share a common edge. So the probability is $(1 - 2\gamma_k + \gamma_k^2 p_1^{-1})$ for each of the $n - 2k - 1$ remaining vertices.

Similarly, the probability we don't have a k -face consisting of σ_i or σ_j and a vertex in $i \triangle j$ is $1 - O(\gamma_k p_1^{-1})$. We then calculate $\mathbb{P}[A_i \cap A_j]$ to be

$$\mathbb{P}[A_i \cap A_j] = \eta_k^2 (1 - 2\gamma_k + \gamma_k^2 p_1^{-1})^{n-2k-1} (1 - O(\gamma_k p_1^{-1})).$$

Before calculating $\mathbb{P}[A_i \cap A_j] - \mathbb{P}[A_i]\mathbb{P}[A_j]$, we observe

$$\begin{aligned} 1 - 2\gamma_k + \gamma_k^2 &= (1 - 2\gamma_k + \gamma_k^2 p_1^{-1}) \frac{1 - 2\gamma_k + \gamma_k^2}{1 - 2\gamma_k + \gamma_k^2 p_1^{-1}} \\ &= (1 - 2\gamma_k + \gamma_k^2 p_1^{-1}) \left(1 - \frac{\gamma_k^2 (p_1^{-1} - 1)}{1 - 2\gamma_k + \gamma_k^2 p_1^{-1}} \right) \\ &= (1 - 2\gamma_k + \gamma_k^2 p_1^{-1}) (1 - O(\gamma_k^2 p_1^{-1})). \end{aligned}$$

The last equality holds by an identical argument to the one in the first case: we can bound $1 - 2\gamma_k + \gamma_k^2 p_1^{-1}$, and consequently its inverse, from above and below by constants. We use this to calculate

$$\begin{aligned} \mathbb{P}[A_i]\mathbb{P}[A_j] &= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2)^{n-k-1} \\ &= \eta_k^2 [(1 - 2\gamma_k + \gamma_k^2 p_1^{-1}) (1 - O(\gamma_k^2 p_1^{-1}))]^{n-k-1} \\ &= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2 p_1^{-1})^{n-k-1} (1 - O(\gamma_k^2 p_1^{-1}))^{n-k-1}. \end{aligned}$$

But since $\gamma_k < n^{-1}$ we have

$$\begin{aligned} (1 - O(\gamma_k^2 p_1^{-1}))^{n-k-1} &= 1 - O(n\gamma_k^2 p_1^{-1}) \\ &= 1 - O(\gamma_k p_1^{-1}). \end{aligned}$$

We calculate

$$\begin{aligned} \mathbb{P}[A_i]\mathbb{P}[A_j] &= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2 p_1^{-1})^{n-k-1} (1 - O(\gamma_k p_1^{-1})) \\ &= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2 p_1^{-1})^{n-2k-1} (1 - 2\gamma_k + \gamma_k^2 p_1^{-1})^k (1 - O(\gamma_k p_1^{-1})) \\ &= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2 p_1^{-1})^{n-2k-1} (1 - O(\gamma_k)) (1 - O(\gamma_k p_1^{-1})) \\ &= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2 p_1^{-1})^{n-2k-1} (1 - O(\gamma_k p_1^{-1})). \end{aligned}$$

Therefore

$$\mathbb{P}[A_i \cap A_j] - \mathbb{P}[A_i]\mathbb{P}[A_j] = \eta_k^2 (1 - 2\gamma_k + \gamma_k^2 p_1^{-1})^{n-2k-1} O(\gamma_k p_1^{-1})$$

with $O(n^{2k+1})$ such pairs i, j , making the total contribution of these pairs to the variance

$$\begin{aligned} S_1 &= O\left(n^{2k-1}\eta_k^2(1-2\gamma_k+\gamma_k^2p_1^{-1})^{n-2k+1}\gamma_kp_1^{-1}\right) \\ &= O\left(n^{2k-1}\eta_k^2(1-2\gamma_k+\gamma_k^2p_1^{-1})^{n-k-1}\gamma_kp_1^{-1}\right). \end{aligned}$$

As before, the second equality follows from bounding $(1-2\gamma_k+\gamma_k^2p_1^{-1})^{k-1}$ by constants on either side.

Since

$$\mathbb{E}[M_{k-1}]^2 = O\left(n^{2k+2}\eta_k^2(1-2\gamma_k+\gamma_k^2)^{n-k-1}\right)$$

it follows that

$$\begin{aligned} \frac{S_1}{\mathbb{E}[M_{k-1}]^2} &= O\left(\frac{(1-2\gamma_k+\gamma_k^2p_1^{-1})^{n-k-1}\gamma_kp_1^{-1}}{n(1-2\gamma_k+\gamma_k^2)^{n-k-1}}\right) \\ &= O\left(\frac{\gamma_kp_1^{-1}}{n}\left(1+\frac{\gamma_k^2(p_1^{-1}-1)}{1-2\gamma_k+\gamma_k^2}\right)^{n-k-1}\right) \\ &= O\left(\frac{\gamma_kp_1^{-1}}{n}\left(1+\frac{\gamma_k^2p_1^{-1}}{1-2\gamma_k}\right)^{n-k-1}\right). \end{aligned}$$

We proceed by bounding the right term by a constant.

$$\begin{aligned} \left(1+\frac{\gamma_k^2p_1^{-1}}{1-2\gamma_k}\right)^{n-k-1} &\leq \exp\left((n-k-1)\frac{\gamma_k^2p_1^{-1}}{1-2\gamma_k}\right) \\ &\leq \exp\left(\frac{n\gamma_k^2p_1^{-1}}{1-k}\right) \\ &\leq e^{1/(1-k)}. \end{aligned}$$

Then

$$\begin{aligned} \frac{S_1}{\mathbb{E}[M_{k-1}]^2} &= O\left(\frac{\gamma_kp_1^{-1}}{n}\right) \\ &= O\left(\frac{1}{n}\right) \\ &= o(1). \end{aligned}$$

Thus the contribution of these pairs is also $o(\mathbb{E}[N_{k-1}]^2)$, as desired.

A.3. $2 \leq I \leq k$. In this final case the probability of the two subcomplexes being contained is $\eta_k^2 \eta_I^{-1}$ where $\eta_I := \prod_{l=1}^{I-1} p_l^{\binom{I}{l+1}}$. The η_I^{-1} accounts for all faces common to i and j , which would otherwise be counted twice. We note σ_i and σ_j share between $I - 2$ and I vertices, and assuming maximal overlap provides an upper bound on $\mathbb{P}[A_i \cap A_j]$. Hence the probability that neither will form a k -simplex with some other vertex is at most $(1 - 2\gamma_k + \gamma_k^2 \gamma_I^{-1})^{n-2k-2+I}$ with $\gamma_I := \prod_{l=1}^I p_l^{\binom{I}{l}}$. The probability of one not forming a k -simplex with one vertex of the other is $1 - O(\gamma_k \gamma_I^{-1})$. We see

$$\mathbb{P}[A_i \cap A_j] = \eta_k^2 \eta_I^{-1} (1 - 2\gamma_k + \gamma_k^2 \gamma_I^{-1})^{n-2k-2+I} (1 - O(\gamma_k \gamma_I^{-1})).$$

Just as in the previous case,

$$1 - 2\gamma_k + \gamma_k^2 = (1 - 2\gamma_k + \gamma_k^2 \gamma_I^{-1}) (1 - O(\gamma_k^2 \gamma_I^{-1})).$$

We now calculate

$$\begin{aligned} \mathbb{P}[A_i] \mathbb{P}[A_j] &= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2)^{n-k-1} \\ &= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2 \gamma_I^{-1})^{n-k-1} (1 - O(\gamma_k^2 \gamma_I^{-1}))^{n-k-1} \\ &= \eta_k^2 (1 - 2\gamma_k + \gamma_k^2 \gamma_I^{-1})^{n-2k-2+I} (1 - O(\gamma_k \gamma_I^{-1})). \end{aligned}$$

It then follows that

$$\begin{aligned} \frac{\mathbb{P}[A_i] \mathbb{P}[A_j]}{\mathbb{P}[A_i \cap A_j]} &= \frac{\eta_k^2 (1 - 2\gamma_k + \gamma_k^2 \gamma_I^{-1})^{n-2k-2+I} (1 - O(\gamma_k \gamma_I^{-1}))}{\eta_k^2 \eta_I^{-1} (1 - 2\gamma_k + \gamma_k^2 \gamma_I^{-1})^{n-2k-2+I} (1 - O(\gamma_k \gamma_I^{-1}))} \\ &= O(\eta_I). \end{aligned}$$

Thus if $\eta_I \neq 1$ then $\mathbb{P}[A_i \cap A_j] - \mathbb{P}[A_i] \mathbb{P}[A_j] = (1 - o(1)) \mathbb{P}[A_i \cap A_j]$, and otherwise $\mathbb{P}[A_i \cap A_j] - \mathbb{P}[A_i] \mathbb{P}[A_j] = \eta_k^2 (1 - 2\gamma_k + \gamma_k^2 \gamma_I^{-1})^{n-2k+I} O(\gamma_k \gamma_I^{-1})$. There are $O(n^{2k+2-I})$ such pairs, so their total contribution to the variance is either

$$\begin{aligned} S_I &= O\left(n^{2k+2-I} \eta_k^2 \eta_I^{-1} (1 - 2\gamma_k + \gamma_k^2 \gamma_I^{-1})^{n-2k-2+I}\right) \\ &= O\left(n^{2k+2-I} \eta_k^2 \eta_I^{-1} (1 - 2\gamma_k + \gamma_k^2 \gamma_I^{-1})^{n-k-1}\right), \end{aligned}$$

or

$$S_I = O\left(n^{2k+2-I} \eta_k^2 (1 - 2\gamma_k + \gamma_k^2 \gamma_I^{-1})^{n-k-1} \gamma_k \gamma_I^{-1}\right).$$

In the first case we have

$$\begin{aligned} \frac{S_I}{\mathbb{E}[M_{k-1}]^2} &= O\left(\frac{n^{2k+2-I}\eta_k^2\eta_I^{-1}(1-2\gamma_k+\gamma_k^2\gamma_I^{-1})^{n-k-1}}{n^{2k+2}\eta_k^2(1-2\gamma_k+\gamma_k^2)^{n-k-1}}\right) \\ &= O\left(\frac{\eta_I^{-1}}{n^I}\left(\frac{1-2\gamma_k+\gamma_k^2\gamma_I^{-1}}{1-2\gamma_k+\gamma_k^2}\right)^{n-k-1}\right). \end{aligned}$$

Just as before, the right-most term can be bounded above by a constant. We note

$$\sum_{l=1}^{I-1} \alpha_l \binom{I}{l+1} < \frac{I}{k} \sum_{l=1}^{k-1} \alpha_l \binom{k}{l+1} < \frac{I}{k} k = I$$

and conclude

$$\begin{aligned} \frac{S_I}{\mathbb{E}[M_{k-1}]^2} &= O\left(\frac{\eta_I^{-1}}{n^I}\right) \\ &= O\left(n^{-I+\sum_{l=1}^{I-1} \alpha_l \binom{I}{l+1}}\right) \\ &= o(1). \end{aligned}$$

In the second case we have

$$\begin{aligned} \frac{S_I}{\mathbb{E}[M_{k-1}]^2} &= O\left(\frac{\gamma_k\gamma_I^{-1}}{n^I}\left(\frac{1-2\gamma_k+\gamma_k^2\gamma_I^{-1}}{1-2\gamma_k+\gamma_k^2}\right)^{n-k-1}\right) \\ &= O\left(\frac{\gamma_k\gamma_I^{-1}}{n^I}\right) \\ &= O(n^{-I}) \\ &= o(1). \end{aligned}$$

Thus $S_I = o(\mathbb{E}[M_{k-1}]^2)$ for $2 \leq I \leq k$. We therefore have that $\mathbb{E}[M_{k-1}^2] = o(\mathbb{E}[M_{k-1}]^2)$, and our result that $M_{k-1} \sim \mathbb{E}[M_{k-1}]$ follows by Chebyshev's Inequality.

□

APPENDIX B. FACTORIAL MOMENTS OF MAXIMAL FACES

Proof of Lemma 18. Similar to previous second moment calculations:

$$\mathbb{E}[N_{k-1}^2] = \binom{n}{k} \sum_{m=0}^k \left[\binom{k}{m} \binom{n-k}{k-m} \left(\prod_{i=1}^{k-1} p_i^{2\binom{k}{i+1} - \binom{m}{i+1}} \right) \right. \\ \left. \left(1 - 2 \prod_{i=1}^k p_i^{\binom{k}{i}} + \prod_{i=1}^k p_i^{2\binom{k}{i} - \binom{m}{i}} \right)^{n-2k+m} (1 - o(1)) \right].$$

We can simplify this slightly to

$$\mathbb{E}[N_{k-1}^2] = (1 + o(1)) \binom{n}{k} \left(\prod_{i=1}^{k-1} p_i^{2\binom{k}{i+1}} \right) \sum_{m=0}^k \left[\binom{k}{m} \binom{n-k}{k-m} \left(\prod_{i=1}^{m-1} p_i^{-\binom{m}{i+1}} \right) \right. \\ \left. \left(1 - 2 \prod_{i=1}^k p_i^{\binom{k}{i}} + \prod_{i=1}^k p_i^{2\binom{k}{i} - \binom{m}{i}} \right)^{n-2k+m} \right]$$

Pulling out the $m = 0$ summand, asymptotically

$$\binom{n}{k} \binom{n-k}{k} \left(\prod_{i=1}^{k-1} p_i^{2\binom{k}{i+1}} \right) \left(1 - \prod_{i=1}^k p_i^{\binom{k}{i}} \right)^{2(n-2k)} \\ = (1 + o(1)) \left(\binom{n}{k} \left(\prod_{i=1}^{k-1} p_i^{\binom{k}{i+1}} \right) \left(1 - \prod_{i=1}^k p_i^{\binom{k}{i}} \right)^{n-k} \right)^2 \\ = (1 + o(1)) \mathbb{E}[N_{k-1}]^2.$$

Meanwhile, the $m = k$ term is seen to be $\mathbb{E}[N_{k-1}]$. We claim the $k - 1$ other summands do not contribute in the limit. For a fixed $m = 1, \dots, k - 1$ let $d_m < 1$ be some constant value such that

$$d_m > \max \left\{ 1 - \frac{m(m-1)}{k(k-1)}, 1 - \frac{m - \sum_{i=1}^{m-1} \alpha_i \binom{m}{i+1}}{k - \sum_{i=1}^{k-1} \alpha_i \binom{k}{i+1}} \right\}.$$

Both fraction terms are between 0 and 1, so such a d_m exists. For sufficiently large n we have

$$1 - 2 \prod_{i=1}^k p_i^{\binom{k}{i}} + \prod_{i=1}^k p_i^{2\binom{k}{i} - \binom{m}{i}} = 1 - \left(2 - \prod_{i=1}^k p_i^{\binom{k}{i} - \binom{m}{i}} \right) \prod_{i=1}^k p_i^{\binom{k}{i}} \\ \leq 1 - (1 + d_m) \prod_{i=1}^k p_i^{\binom{k}{i}}.$$

Thus there exists a constant D such that

$$\begin{aligned}
\left(1 - 2 \prod_{i=1}^k p_i^{\binom{k}{i}} + \prod_{i=1}^k p_i^{2\binom{k}{i} - \binom{m}{i}}\right)^{n-2k+m} &\leq \left(1 - (1+d_m) \prod_{i=1}^k p_i^{\binom{k}{i}}\right)^{n-2k+m} \\
&\leq D e^{-n(1+d_m) \left(\prod_{i=1}^k p_i^{\binom{k}{i}}\right)} \\
&= D e^{-(1+d_m)(\rho_1 \log n + \frac{k-1}{2} \log \log n + c)} \\
&= D n^{-(1+d_m)\rho_1} (\log n)^{-(1+d_m)\frac{k-1}{2}} e^{-(1+d_m)c}.
\end{aligned}$$

Then by our construction of d_m ,

$$n^{-(1+d_m)\rho_1} = o\left(n^{-2k+m+\sum_{i=1}^{k-1} \alpha_i \binom{k}{i} - \sum_{i=1}^{m-1} \alpha_i \binom{m}{i}}\right)$$

and

$$(\log n)^{-(1+d_m)\frac{k-1}{2}} = o\left((\log n)^{-(k-1) + \frac{m(m-1)}{2k}}\right).$$

It follows that the corresponding summand is bounded by

$$D n^{2k-m} \prod_{i=1}^{k-1} p_i^{2\binom{k}{i+1} - \binom{m}{i+1}} n^{-(1+d_m)\rho_1} (\log n)^{-(1+d_m)\frac{k-1}{2}} = o(1),$$

thereby contributing nothing as $n \rightarrow \infty$. Therefore,

$$\mathbb{E}[(N_{k-1})_2] = \mathbb{E}[N_{k-1}^2] - \mathbb{E}[N_{k-1}] = \mathbb{E}[N_{k-1}]^2(1 - o(1)) \rightarrow \mathbb{E}[N_{k-1}]^2$$

as $n \rightarrow \infty$. We will now establish a similar result for each factorial moment.

We direct our attention to the l -th factorial moment of N_{k-1} , assuming that $\mathbb{E}[(N_{k-1})_j] \rightarrow \mathbb{E}[N_{k-1}]^j$ for all $j < l$. Using the notation a Section 3 we have

$$\begin{aligned}
\mathbb{E}[N_{k-1}^l] &= \mathbb{E}\left[\left(\sum_{\sigma \in \binom{[n]}{k}} 1_{C_\sigma}\right)^l\right] \\
&= \sum_{\sigma_1, \dots, \sigma_l \in \binom{[n]}{k}} \mathbb{P}[C_{\sigma_1} \cap \dots \cap C_{\sigma_l}].
\end{aligned}$$

We break up this sum into two parts: where no two σ_i 's are identical and where such two σ_i are the same. Considering the first case, an identical argument for $l = 2$ tells us the only summand contributing in the limit corresponds to when no two faces share any vertices, and this term converges to $\mathbb{E}[N_{k-1}]^l$.

Moving on to the second case, we let $s(l, j)$ and $S(l, j)$ denote Stirling numbers of the first and second kind, respectively. There are $S(l, j)$ ways to break our σ_i up into j groups where all faces in a group are the same. Moreover, for any such configuration into j groups, the corresponding contribution to $\mathbb{E}[N_{k-1}^l]$ would be $\mathbb{E}[N_{k-1}^j]$. We begin by pulling out $S(l, l-1) = -s(l, l-1)$ copies of $\mathbb{E}[N_{k-1}^{l-1}]$. However, the number of partitions of σ_i into $k-2$ groups has now been overcounted. There should only be $S(l, l-2)$ such configurations, but we have just counted $-s(l, l-1)S(l-1, l-2)$ of them, so we add $S(l, l-2) + s(l, l-1)S(l-1, l-2) = -s(l, l-2)$ copies of $\mathbb{E}[N_{k-1}^{l-2}]$.

Fixing some $j < l-1$, we now assume attaching a coefficient of $-s(l, m)$ to $\mathbb{E}[N_{k-1}^m]$ for all $m > j$ ensures every partition of the σ_i into $j+1, \dots, l-1$ sets is properly counted. Then for each $m > j$, the $-s(l, m)$ copies of $\mathbb{E}[N_{k-1}^m]$ count $-s(l, m)S(m, j)$ partitions into just j groups. Meanwhile we know there are actually only $S(l, j)$ distinct partitions, so we must add:

$$\begin{aligned} S(l, j) + \sum_{m=j+1}^{l-1} s(l, m)S(m, j) &= \sum_{m=j+1}^l s(l, m)S(m, j) \\ &= \sum_{m=j}^l s(l, m)S(m, j) - s(l, j)S(j, j) \\ &= \delta_{l,j} - s(l, j) = -s(l, j). \end{aligned}$$

The last line follows from a well known Stirling number identity. We use induction to conclude that as $n \rightarrow \infty$,

$$\mathbb{E}[N_{k-1}^l] \rightarrow \mathbb{E}[N_{k-1}]^l - \sum_{j=1}^{l-1} s(l, j)\mathbb{E}[N_{k-1}^j],$$

thus

$$\mathbb{E}[(N_{k-1})_j] = \sum_{j=1}^l s(l, j)\mathbb{E}[N_{k-1}^j] \rightarrow \mathbb{E}[N_{k-1}]^l$$

for any fixed l . It follows that N_{k-1} converges in distribution to $\text{Poi}(\mu)$ with $\mu = \mathbb{E}[N_{k-1}]$, completing our proof. □

ACKNOWLEDGEMENTS

We would like to thank Christopher Hoffman, Matthew Junge, and Matthew Kahle for helpful conversations on this subject and readings of early versions of this manuscript.

REFERENCES

- [AL15] Lior Aronshtam and Nathan Linial, *When does the top homology of a random simplicial complex vanish?*, *Random Structures & Algorithms* **46** (2015), no. 1, 26–35.
- [ALLM13] Lior Aronshtam, Nathan Linial, Tomasz Łuczak, and Roy Meshulam, *Collapsibility and vanishing of top homology in random simplicial complexes*, *Discrete & Computational Geometry* **49** (2013), no. 2, 317–334.
- [BHK11] Eric Babson, Christopher Hoffman, and Matthew Kahle, *The fundamental group of random 2-complexes*, *Journal of the American Mathematical Society* **24** (2011), no. 1, 1–28.
- [BS97] Werner Ballmann and J Świątkowski, *On l_2 -cohomology and property (t) for automorphism groups of polyhedral cell complexes*, *Geometric & Functional Analysis GAFA* **7** (1997), no. 4, 615–645.
- [CF16] Armindo Costa and Michael Farber, *Random simplicial complexes*, 129–153.
- [CF17a] A Costa and M Farber, *Large random simplicial complexes, ii; the fundamental group*, *Journal of Topology and Analysis* **9** (2017), no. 03, 441–483.
- [CF17b] ———, *Large random simplicial complexes, iii the critical dimension*, *Journal of Knot Theory and Its Ramifications* **26** (2017), no. 02, 1740010.
- [ER59] P Erdős and A Rényi, *On random graphs i.*, *Publ. Math. Debrecen* **6** (1959), 290–297.
- [Gar73] Howard Garland, *p -adic curvature and the cohomology of discrete subgroups of p -adic groups*, *Annals of Mathematics* (1973), 375–423.
- [Hat02] Allen Hatcher, *Algebraic topology*, Cambridge University Press, 2002.
- [HKP12] Christopher Hoffman, Matthew Kahle, and Elliot Paquette, *Spectral gaps of random graphs and applications to random topology*, arXiv preprint arXiv:1201.0425 (2012).
- [HKP13] ———, *The threshold for integer homology in random d -complexes*, arXiv preprint arXiv:1308.6232 (2013).
- [Kah09] Matthew Kahle, *Topology of random clique complexes*, *Discrete Mathematics* **309** (2009), no. 6, 1658–1671.
- [Kah14a] ———, *Sharp vanishing thresholds for cohomology of random flag complexes*, *Ann. of Math.* **179** (2014), 1085–1107.
- [Kah14b] ———, *Topology of random simplicial complexes: a survey*, *AMS Contemp. Math* **620** (2014), 201–222.
- [KM⁺13] Matthew Kahle, Elizabeth Meckes, et al., *Limit theorems for betti numbers of random simplicial complexes*, *Homology, Homotopy and Applications* **15** (2013), no. 1, 343–374.
- [Koz10] Dmitry Kozlov, *The threshold function for vanishing of the top homology group of random d -complexes*, *Proceedings of the American Mathematical Society* **138** (2010), no. 12, 4517–4527.
- [LM06] Nathan Linial and Roy Meshulam, *Homological connectivity of random 2-complexes*, *Combinatorica* **26** (2006), no. 4, 475–487.
- [MW09] Roy Meshulam and Nathan Wallach, *Homological connectivity of random k -dimensional complexes*, *Random Structures & Algorithms* **34** (2009), no. 3, 408–417.