1-1-2013

The 2-Core of a Random Inhomogeneous Hypergraph

Omar Abuzzahab
University of Pennsylvania, omarabuzzahab.math@gmail.com

The 2-Core of a Random Inhomogeneous Hypergraph

Abstract
The k-core of a hypergraph is the unique subgraph where all vertices have degree at least k and which is the maximal induced subgraph with this property. We study the 2-core of a random hypergraph by probabilistic analysis of the following edge removal rule: remove any vertices with degree less than 2, and remove all hyperedges incident to these vertices. This process terminates with the 2-core. The hypergraph model studied is an inhomogeneous model --- where the expected degrees are not identical. The main result we prove is that as the number of vertices n tends to infinity, the number of hyperedges R in the 2-core obeys a limit law: R/n converges in probability to a non-random constant.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Mathematics

First Advisor
Robin Pemantle

Subject Categories
Mathematics

THE 2-CORE OF A RANDOM INHOMOGENEOUS HYPERGRAPH

Omar Abuzzahab

A DISSERTATION

in

Mathematics

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2013

Robin Pemantle, Merriam Term Professor of Mathematics
Supervisor of Dissertation

Jonathan Block, Professor of Mathematics
Graduate Group Chairperson

Dissertation Committee:
Robin Pemantle, Merriam Term Professor of Mathematics
J. Michael Steele, C.F. Koo Professor of Statistics
Sampath Kannan, Henry Salvatori Professor of Computer and Information Science
Andre Scedrov, Professor of Mathematics

Acknowledgments

Dedicated to Mom and Dad, whom I love dearly and have supported me throughout life. This support has enabled all of my own accomplishments, including this dissertation.

I also have heartfelt thanks to those who have acted as very special teachers and mentors to me: Robert Cassola, Victor Reiner, and Robin Pemantle. In a very real way, their wisdom and enthusiasm has positively shaped who I am. I aspire to achieve the same talent and ability.

ABSTRACT

THE 2-CORE OF A RANDOM INHOMOGENEOUS HYPERGRAPH

Omar Abuzzahab

Robin Pemantle

The k-core of a hypergraph is the unique subgraph where all vertices have degree at least k and which is the maximal induced subgraph with this property. We study the 2-core of a random hypergraph by probabilistic analysis of the following edge removal rule: remove any vertices with degree less than 2, and remove all hyperedges incident to these vertices. This process terminates with the 2-core. The main result we prove is that as the number of vertices n tends to infinity, the number of hyperedges R in the 2-core obeys a limit law: $\frac{1}{n}R$ converges in probability to a non-random constant.

More explicitly, given $a > 0$ we consider a hypergraph model with m independent hyperedges on n vertices where the jth vertex is incident to each hyperedge with probability asymptotically $\frac{a}{j}$. We also fix an overall density $c_{\text{den}} > 0$ and take limits $n \to \infty$ with the ratio m/n tending to c_{den}.

The result we prove is that $R = \beta m + o_p(n)$, where $\beta = \beta(a, c_{\text{den}})$ denotes the largest solution to the equation

$$\log \beta = -a \int_{ac_{\text{den}}\beta}^{\infty} \frac{e^{-t}}{t}\, dt$$

when there is at least one solution, and $\beta = 0$ otherwise. For $a \geq 1$, define c_* by

$$c_* = \frac{\log a}{a} \exp\left(a \int_{\log a}^{\infty} \frac{e^{-t}}{t}\, dt \right),$$

and for $a < 1$, let $c_* = 0$. The size of the 2-core exhibits a phase transition from $\beta = 0$ to $\beta > 0$ as c_{den} varies from $c_{\text{den}} < c_*$ to $c_{\text{den}} > c_*$. This transition is continuous

across $c = c_*$ when $a = 1$, and discontinuous when $a > 1$.

Contents

Chapter 1

Motivation

The motivation for this thesis comes from answering a conjecture in probabilistic number theory. The setting is studying the effectiveness of random algorithms for integer factorization, such as Dixon's random squares algorithm, the quadratic sieve algorithm, and related algorithms which operate by finding perfect squares within certain sequences of integers. In broadest terms, the conjecture asserts a sharp probability estimate for the appearance of these perfect squares.

The run time of these algorithms is influenced by this probability. Discoveries on the run time of factoring algorithms holds considerable interest due to both factoriza-

tion's close historical relationship with studying number theory itself and its modern relevance for encrypted communication. On this later point, encryption schemes dependent on computationally difficult problems have enabled much of the digital communication that has become ubiquitous today. And in practice, the most widely used computational problem is prime factorization. Factorization is regarded as a computationally difficult problem, and the unproven belief that this is so constitutes the assurance of security to the encryption.

Let us describe the algorithms of interest. The general strategy is based on congruences $X^2 \equiv Y^2 \pmod{n}$ between integers X and Y. Any such congruence might lead to discovering a factor of n, as one has

$$0 \equiv X^2 - Y^2 \equiv (X - Y)(X + Y) \pmod{n},$$

and assuming $X \neq \pm Y \pmod{n}$ we have that n does not divide $X - Y$ or $X + Y$. In this case, $\gcd(n, X - Y)$ will be a proper divisor of n. This factors n into a product of smaller integers, each of which we could then to attempt to factor further. To define a complete factorization algorithm we must only decide on how to generate (random) congruences. Both the run time of this generation process and the chance that $\gcd(n, X - Y)$ is nontrivial determines the run time of our algorithm, although here it is the former that we are interested in.

The generation process can be described as follows: generate a sequence of pseudorandom positive integers $a_1, a_2, \ldots$ such that for each i there is some integer b_i with $a_i \equiv b_i^2 \pmod{n}$. Generating these a_i's is the first part towards creating a single congruence $X^2 \equiv Y^2 \pmod{n}$. We do so by selecting a subsequence of the a_i's whose product is a perfect square, $a_{i_1} \cdot a_{i_2} \cdots a_{i_k} = Y^2$ for some Y. We call such a subsequence a square dependence. Then Y^2 is congruent to $X^2 \coloneqq (b_{i_1} \cdot b_{i_2} \cdots b_{i_k})^2$, as desired.

This same general process is used in a family of related algorithms: Dixon's random squares algorithm [7], the quadratic sieve [14], the multiple polynomial sieve [17], and the number field sieve [2] (which uses a close facsimile of this process). Dixon's random squares algorithm corresponds to choosing each integer b_i independently and uniformly over $\{1, 2, \ldots, n\}$, and then defining the integer a_i by reducing b_i^2 to the smallest positive residue modulo n. The distribution of a_i is thus uniform over the set of quadratic residues modulo n.

In analysis of the expected run time of Dixon's random squares, it is typically assumed that the distribution of each a_i is essentially equivalent to being uniform on the set $\{1, 2, \ldots, n\}$ in that the prime number theorem may be used to estimate its statistical properties. See for example Dixon [7], who attributes this heuristic to Richard Schroeppel. In the quadratic sieve and the other algorithms, the integers

b_i are in fact chosen deterministically and not randomly. Nevertheless the same assumption that each integer a_i is an independent, uniformly distributed integer is the assumption by which one creates a heuristic analysis of the expected run times. Observing this, Pomerance in [15, 16] formalized it as a problem worth studying in its own right.

Pomerance's Problem: *Given an integer $n > 0$ and an iid sequence $a_1, a_2, \ldots, a_m$ of positive integers chosen uniformly from $\{1, 2, \ldots, n\}$, how large must $m = m(n)$ be so that there is a subset of these integers whose product is a perfect square?*

Due to an observation by David Moulton, a small variation on finding a square dependence in Pomerance's problem — where the a_i's are distributed uniformly — can transfer results on Pomerance's problem back to the original situation where the a_i's are distributed uniformly on the quadratic residues. This gives a rigorous treatment which completely avoids the unproven heuristic that integers a_i distributed over the quadratic residues have essentially the same statistical properties as those distributed uniformly. Moulton's observation is explained more fully in [5], but the key idea is that a random variable a_i distributed uniformly on $G = (\mathbb{Z}/n\mathbb{Z})^*$ (an assumption which changes the probabilities in Pomerance's problem by only $o(1)$) can be considered a

random variable $b_i^2 g_i$ where b_i^2 is uniformly distributed on the subgroup Q of quadratic residues of G and where g_i is an independent random variable uniformly distributed on a set of representatives for the cosets G/Q. It is known that the time in Pomerance's problem for reaching the first square dependence is asymptotically the same as the time for having many independent square dependences. This knowledge is enough to imply that with high probability there is not only a square product of the a_i's, but one for which the product of the g_i's multiplies to a quadratic residue as well.

To proceed further we need to describe the specific, intelligent method used by algorithms to search for a square dependence. Ultimately, forming a product which is a perfect square amounts to arranging for the primes dividing one a_i with odd multiplicity to be paired with like prime factors in other a_j's. But since this relies on having a factorization of the integers a_i's at hand, some more subtlety is required; if we consider only those a_i's whose prime factors are all relatively small (so called smooth numbers [1]) then the extra factorization work will not defeat the purpose of the algorithm, with the tradeoff that square dependences involving the a_i's with large prime factors will no longer be discovered. Pomerance's question asks in a sense for the more liberal answer of when there is a square dependence at all, although as we

[1]The term appears to be coined by Leonard Adleman. This technique of employing smooth numbers for effective number theory algorithms occurs throughout the subject. The books [3] and [18] provide good references.

will see they are quite related.

Finding a square dependence has an equivalent formulation as a linear dependence problem: associate to each integer a_i the vector v_i whose components are the multiplicities of each prime in the factorization a_i. In other words, $a_i = \Pi_j p_j^{v_{i,j}}$, where p_j denotes the jth prime number and $v_{i,j}$ denotes the jth component of vector v_i. A product of integers is a perfect square if and only if their associated prime multiplicity vectors sum to zero mod 2. That is, a square dependence is equivalent to a linear dependence over $\mathbb{F}_2$.

When one is working in the vector space $\mathbb{F}_2^n$, a simple but effective linear algebra idea (both as an algorithmic strategy and as a probability estimate) is to consider instead the event that there are $k+1$ vectors all of which have their non-zero components residing in k components of $\mathbb{F}_2^n$. Since these vectors reside in a subspace of dimension k, such a set must necessarily be dependent.

Let $\pi(x)$ denote the number of primes $\leq x$. A number is said to be y-smooth if every prime factor is $\leq y$. Let $\Psi(x, y)$ denote the number of y-smooth integers $\leq x$. In the factorization algorithms, the fruitful linear algebra idea is employed by fixing $y > 0$ and considering only the a_i which are y-smooth. These a_i's are filtered from the rest of the sequence by a process called sieving (after which several of the algorithms

are named). As a row vector, all non-zero components of v_i reside in the first $\pi(y)$ columns. Once we obtain more than $\pi(y)$ many y-smooth numbers, we know we will have a square dependence. To find one such square dependence, Gaussian elimination (or a more specialized algorithm such as Wiedemann's sparse matrix method) may be used on the matrix whose rows consist of the vectors v_i.

In Pomerance's Problem, let T denote the first time m for which $a_1, a_2, \ldots, a_m$ contains a square dependence. A priori, Pomerance's Problem is not concerned with the effort required to factor each a_i. Nevertheless, Schroeppel gave the following simple argument based on the smooth number approach in the late 1970's (unpublished, but referenced secondhand in papers such as [13]) that gives a good upper bound for T. Each a_i is y-smooth with probability $\Psi(n, y)/n$. Pick $y_0 > 0$ which maximizes this probability. The number of y_0-smooth numbers in the sequence $a_1, a_2, \ldots, a_m$ is binomially distributed with mean $\frac{\Psi(n,y_0)m}{n}$. If this mean is at least $(1 + o(1))\,\pi(y_0)$ — which is to say if $m \geq (1 + o(1))\,J_0(n)$ where $J_0(n) := \frac{\pi(y_0)n}{\Psi(n,y_0)}$ — then it follows (from the concentration of the binomial distribution) that with high probability the number of y_0-smooth integers is at least $\pi(y_0) + 1$. Thus $T \leq J_0(n)$ with high probability.

Estimates for the non-random quantity $J_0(n)$ are known [5] and in the limit $n \to$

∞, it admits the estimate

$$J_0(n) = \exp\left(\sqrt{\big(2 + o\,(1)\big) \log n \log\, \log n}\right).$$

In recent work published in 2012, Croot, Granville, Pemantle and Tetali [5] proved that T satisfies, with high probability,

$$\frac{\pi}{4} e^{-\gamma}\big(1 + o\,(1)\big)\; J_0(n) \leq\; T\; \leq e^{-\gamma}\big(1 + o\,(1)\big)\; J_0(n)\,.$$

This theorem gives the best known bounds on T. Pemantle et al. conjectured that the threshold for T is sharp in the sense that the constant $\frac{\pi}{4} e^{-\gamma}$ in lower bound could be increased to match the constant $e^{-\gamma}$ in the upper bound.

Conjecture 1.1. For every $\epsilon > 0$,

$$P\left(T \in \left[\big(1 - \epsilon\big)\; e^{-\gamma} J_0, \big(1 + \epsilon\big)\; e^{-\gamma} J_0\right]\right) = 1 - o\,(1)$$

as $n \to \infty$.

Resolving this conjecture is the main motivation referred to earlier. This is an interesting problem to solve not only because it would indicate a sharp threshold for Pomerance's Problem, but because it would also inform a great deal about the best techniques for designing these algorithms.

To explain, the proof of the upper bound in [5] considers a larger event—compared to Schroeppel's proof—which implies a square dependence. The idea is that considering the numbers a_i which are not y-smooth is still useful for finding $\pi(y)+1$ many y-smooth numbers, one just needs for these non-smooth numbers to be multiplied together appropriately so that their large prime factors with odd multiplicities are paired to become even. In terms of the row vectors v_i, we must form appropriate linear combinations so as to cancel (mod 2) all non-zero components in columns whose index is greater than $\pi(y)$. Such a linear combination corresponds to creating an additional y-smooth number (essentially y-smooth at least — a y-smooth number times a perfect square), which speeds the search for $\pi(y)+1$ such numbers.

The details of their argument in fact narrow this event in a couple ways: (1) by limiting which nonsmooth a_i's are considered to only those with large primes below some threshold My (as the larger M is, the less useful it is for creating additional y-smooth numbers), and (2) by limiting the manner by which one attempts to form the appropriate combinations to produce the additional y-smooth numbers. A proof of Conjecture 1.1 would tell us that this specific, relatively narrow event for a square dependence in fact asymptotically captures the full event. When designing an algorithm then, there would be little purpose in casting a wider search for square dependences that arise from unusual combinations of a_i's. See [4] for a fuller discussion on practical

considerations.

Chapter 2

Scope of This Work

2.1 Hypergraphs and 2-cores

Formally, a hypergraph on a vertex set V is a collection E of subsets of V. The elements $e \in E$ are called hyperedges. The degree of a vertex is the number of hyperedges which contain it. The k-core of a hypergraph is the unique subgraph where all vertices have degree at least k and which is the maximal induced subgraph with this property.

The k-core of any hypergraph can be obtained by iterating the following edge

removal rule: remove any vertices v with degree less than k, and remove all hyperedges incident to these vertices. Since the graph is finite this must eventually terminate with the unique maximal subgraph whose vertices all have degree at least k.

The scope of this thesis concerns the size of the 2-core for a particular random hypergraph model related to Pomerance's Problem. Compared to existing literature [12, 10, 9], the new wrinkle captured by the random hypergraph model studied is inhomogeneity—where the degree distribution on vertices is not identical. Rather, the expected degree sequence has a power law tail. Let n denote the size of the vertex set V. A summary of the main result we prove, Theorem (2.1), is that as n tends to infinity, the number of hyperedges R in the 2-core obeys a limit law: $\frac{1}{n}R$ converges in probability to an explicit, non-random constant. Further, the theorem details the value of this constant and threshold for it being nonzero through explicit expressions of the parameters of the model.

Before going into more detail, let us illustrate the relationship with the subject of the previous chapter. Denote by V the set of vertices $\{1, 2, \ldots, n\}$, which may be viewed as the index set for the n scalar components of a vector in $\mathbb{F}_2^n$. A bijective correspondence from sequences $v_1, v_2, \ldots, v_m$ of vectors in $\mathbb{F}_2^n$ to hypergraphs on V is given by turning each vector v_i into a hyperedge e_i consisting of the components of v_i which are nonzero.

A vector which, in the language of hyperedges, contains a degree 1 vertex is necessarily independent to all other vectors. Passing to the 2-core of the original hypergraph recursively strips away all such vectors. Any linearly dependent set of vectors must therefore also be a subset of the vectors in the 2-core.

Moreover, there is strong reason to believe that Conjecture 1.1 can be resolved by understanding relatively simple properties of the 2-core of the associated random vectors. This reasoning is discussed in the next section, and for the present discussion we will summarize: what has been shown in many other models [8, 6, 11, 1] (properly translated to the current setting) is that with high probability the dependence of the 2-core coincides with the event that the number of hyperedges, m', in the 2-core, exceeds the number of vertices, n', in the 2-core. This is another instance of the fruitful linear algebra idea of chapter 1, but where we are considering a distinctly different set of vectors (the vectors in the 2-core here, as opposed to the y-smooth numbers and any additional y-smooth numbers which are reachable from combinations).

Returning to Pomerance's Problem, the sequence of integers $a_1, a_2, \ldots, a_m$ gives rise to a sequence of vectors in $\mathbb{F}_2^n$, $v_1, v_2, \ldots, v_m$ via their prime exponents. In turn these vectors define a random hypergraph. The random hypergraph model we study in this paper is an approximate version of this distribution.

Let us demonstrate how one might arrive at an approximate model: when n is large, the probability that a prime p divides a_i is approximately $1/p$, and this probability is approximately independent of another prime q dividing a_i. If the prime p is large then $1/p$ is also the approximate probability that p divides a_i with odd multiplicity (and therefore would represent a nonzero component in v_i). So the associated random hypergraph is one where the vertices are indexed by prime numbers, have degree distributions which are weakly dependent, and have expected degree decaying roughly as $\frac{m}{p}$.

The model we study (to be defined in section 2.3) simplifies this by first treating the vertices as having independent degree distributions, and second by smoothing and simplifying the rate at which this expected degree $\frac{m}{p}$ decreases. If we were to index our vertices sequentially $j = 1, 2, \ldots$ then we have the expected degree as $\frac{m}{p_j}$ for vertex j, where p_j is the jth prime number. By the prime number theorem, the jth prime is asymptotically $j \log j$, and so this expectation is asymptotically $\frac{m}{j \log j}$. The simplification taken in our model is to consider expected degrees asymptotically equal to $\frac{m}{j}$.

2.2 The Size of the 2-core as a Proxy for Dependence

What is remarkable is that in many cases a relatively simple property of the size of the 2-core determines with high probability (whp) whether a sequence of vectors $v_1, v_2, \ldots, v_m$ are linearly independent over $\mathbb{F}_2$. This is best understood by considering the following dual satisfiability problem. Denote by A the $m \times n$ matrix whose rows are the vectors v_i and let $b \in \mathbb{F}_2^m$ be a random vector chosen uniformly and independently. If the random system of linear equations $Ax = b$ is unsatisfiable then the vectors are surely dependent. Conversely, if the vectors are dependent then A has rank $m - s$ for $s > 0$. The probability of b lying in the column space of A is thus reduced to 2^{-s}. From these two observations it is easy to see that if m crosses a threshold for which the satisfiability problem transitions from being satisfiable whp to unsatisfiable whp, then the associated vector dependence problem also transitions from being independent whp to dependent whp.

In the dual satisfiability problem, it is not difficult to show that passing to the 2-core (also known as pure literal elimination) does not affect satisfiability of the system but will tend to decrease the dimension of the kernel. So when there is at least one solution there will be fewer of them. Probabilistically, this means the

expected number of solutions is brought closer to the probability that there is at least one solution. One therefore expects that the second moment method applied to the 2-core will yield sharper threshold bounds than the original system. In many models this has been carried out rigorously (although not yet in our inhomogeneous model), as we now discuss.

Consider choosing vectors randomly according to the uniform distribution over vectors with a fixed number $k \geq 3$ of nonzero components. The main result in [8] (phrased as the equivalent 3-XORSAT problem) considers $k = 3$ and shows that whp the satisfiability of the 2-core coincides with the event that the number of hyperedges, m', in the 2-core, exceeds the number of vertices, n', in the 2-core. This result is also believed to hold for $k > 3$ as demonstrated in [6, 11] (whose complete proof is subject to a small analytic conjecture). One half of these results is immediate: if $m' > n'$ then in terms of vectors there are more vectors than nonzero components—they are surely dependent. The nontrivial part is that when $m' \leq (1 - \epsilon)\, n'$ the 2-core is satisfiable whp. Finally, the satisfiability threshold of random 3-SAT was established in [1], and again this was done by proving the satisfiability of the 2-core coincides with this same size threshold. While 3-SAT is not a linear system, the result does further strengthen the belief that simple properties of the 2-core will capture the satisfiability threshold in many models.

As discussed at the end of the previous section, the hypergraph model we will consider serves as an approximation to Pomerance's problem. The sharp thresholds given in our main result is a major first step towards resolving the threshold for dependence and ultimately resolving Pomerance's problem. It is interesting to note that in Theorem (2.1), the parameter c_{den} of our model has the critical value $c_{\text{den}} = e^{-\gamma}$ for the threshold of the 2-core's size, which is already quite suggestive of the connection to Conjecture 1.1. In fact, the story so far is even more telling: in this hypergraph, m'/n' transitions from less than 1 to greater than 1 as c_{den} crosses this threshold.

2.3 Main Results

We consider a probability space Ω with measure P whose elements are hypergraphs on the n element set $V = \{0, 1, \ldots, n-1\}$ with at most m hyperedges. The probability measure for the hypergraph is given by generating m iid subsets of V, denoted as $e_1, e_2, \ldots, e_m$, representing the potential hyperedges (there will be strictly less than m hyperedges if any subset e_i is empty). The distribution of a single subset e is given by deciding independently whether each vertex $j \in V$ will be a member of e, and importantly this probability is not the same for each vertex. Instead, we'd like

to adjoin vertex j to e with probability asymptotically equal to $\frac{a}{j}$ where $a > 0$ is a constant. To ensure this is a proper probability (i.e. between 0 and 1), we take the probability to be $\frac{a}{2a+j} = \frac{1}{2+\frac{j}{a}}$.

The edge density of a hypergraph is the number of hyperedges divided by the number of vertices. We will consider the size n of our vertex set to tend to infinity with $m/n \longrightarrow c_{\text{den}}$ so that the expected density (proportional to m/n) is tending to a limit. Formally, $P = P_{n,a,c_{\text{den}}}$ and when we take limits $n \to \infty$ we do so with a and c_{den} fixed, and with $m = m(n)$ a fixed function of n. We say $X_n = o_p(f(n))$ if for all $\epsilon > 0$, $P(|X_n| > \epsilon f(n)) \to 0$ as $n \to \infty$.

Theorem 2.1 (Main Theorem)**.** Let R denote the number of edges in the 2-core, and let $\beta = \beta(a, c_{\text{den}})$ denote the largest solution to the equation

$$\log \beta = -a \int_{ac_{\text{den}}\beta}^{\infty} \frac{e^{-t}}{t}\, dt$$

when there is at least one solution, and define $\beta = 0$ otherwise. Then

$$R = \beta m + o_p(n)\,,$$

excluding the case $c_{\text{den}} = c_*$ when $a > 1$ (see below). Furthermore, there are 3 distinct cases for how β behaves:

1. If $a < 1$, then $\beta > 0$ for all $c_{\text{den}} > 0$.

2. For $a = 1$:

   Subcritical case: If $c_{\text{den}} \leq e^{-\gamma}$ then $\beta = 0$, and so $R = o_p(n)$.

   Supercritical case: If $c_{\text{den}} > e^{-\gamma}$ then $\beta > 0$. Here, $\beta \downarrow 0$ as $c_{\text{den}} \downarrow e^{-\gamma}$.

3. For $a > 1$:

   Define $c_* > 0$ by

$$c_* = \frac{\log a}{a} \exp\left(a \int_{\log a}^{\infty} \frac{e^{-t}}{t}\, dt \right).$$

   Subcritical case: If $c_{\text{den}} < c_*$ then $\beta = 0$, and so $R = o_p(n)$.

   Supercritical case: If $c_{\text{den}} > c_*$ then $\beta > 0$. Here, $\beta \downarrow \frac{\log a}{a c_*} > 0$ as $c_{\text{den}} \downarrow c_*$.

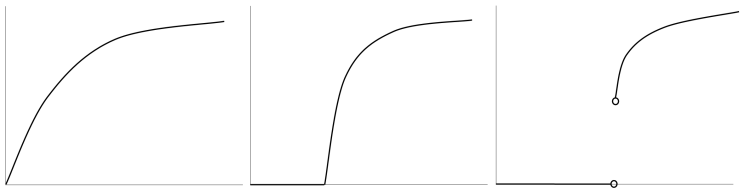

Figure 2.1: Plot of β versus c_{den}, showing transition behavior at c_*. From left to right: $a < 1$, $a = 1$, and $a > 1$.

The theorem can be organized into a phase diagram (Figure 2.3) where we plot β as a function of c_{den}.

To summarize, $\frac{1}{m}R$ represents the fractional size of the 2-core compared to original hypergraph, and this converges in probability to a constant β. As the edge parameter

c_{den} is increased from 0 there are one of three situations depending on a. When $a < 1$ the positivity of β says that the 2-core is always a positive fraction of the graph, whereas when $a \geq 1$ the 2-core represents a vanishing fraction of the graph until c_{den} crosses a threshold c_*, after which it is a positive fraction. The behavior near the threshold is different still for $a = 1$ versus $a > 1$ with the later having a discontinuous jump from a small 2-core to a 2-core that is not only giant, but also already as large as $\frac{\log a}{ac_*}m$.

Informal Discussion of the Degree Distribution and the Removal Map

The degree distribution of a vertex inside some interval $[xn, (x+\epsilon)\,n]$ converges to a Poisson random variable with mean $\frac{ac_{\text{den}}}{x}$. In general, for any intensity function $\lambda(x) : [0,1] \longrightarrow \mathbb{R}$ we may consider a random hypergraph on $\left\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$ where the degree distribution of vertex x is an independent Poisson with mean $\lambda(x)$. With $\lambda(x) = \frac{ac_{\text{den}}}{x}$ we expect that the trajectory of the process on the Poisson graph to approximate the process on the original graph.

Consider now how the removing edges map affects the degree distribution of the vertices. At each step, first a random set B of degree 1 vertices is removed. Since the hypergraph has iid hyperedges with independent vertices, each $v \in B$ lies in

a uniformly random hyperedge. So conditional on B, each hyperedge is incident to a binomially distributed number of degree 1 vertices (with success probability $|B|$ divided by the number of hyperedges). This gives some random fraction p of hyperedges that will survive (those incident to zero such vertices). We can summarize the randomness thusly: at each step a random set B of vertices is removed, and from $|B|$ a random edge survival fraction p is generated. The vertices outside of B have a degree distribution that has first been truncated to be at least 2, and then independently thinned with retention probability p.

It is well known that the distribution of a thinned Poisson with mean λ is again a Poisson with mean $p\lambda$. In a similar vein, if we take a Poisson random variable and apply a sequence of truncations and thinnings (each with its own retention probability) then the resulting distribution is easy to describe: when this distribution is truncated once more it will be that of a truncated Poisson with mean $p\lambda$, where p is the product of the retention probabilities used in the sequence. Therefore we can summarize the randomness of the Poisson graph after i steps as choosing a single random total thinning parameter p. All vertices have their degree distribution truncated to be at least 2, and then independently thinned with retention probability p.

This nice compatibility of Poisson vertex degrees with truncation and thinning makes it tempting to formalize a simple deterministic approximation to the removal

process. For now, a rough description will suffice. Condition on the number b_i of degree 1 vertices at step i as well as the number m_i of hyperedges remaining after i steps. The distribution of hyperedges removed in the next step is modeled by placing b_i independent balls into m_i boxes. We remove a hyperedge as part of the removal step if its box contains at least one ball. In the situation where b_i/m_i tends to a limit then the number of balls in each box is an (unrelated) Poisson random variable with mean $\lambda_i = b_i/m_i$. So the fraction of surviving hyperedges is tending to the probability e^{-b_i/m_i} that this Poisson variable is zero. Assuming some sort of concentration around the expectations involved we get a deterministic real sequence modeling the number of hyperedges: $m_{i+1} = m_i E_i e^{-b_i/m_i}$ where the expectation operator uses the distribution of the process on Poisson graph with total thinning parameter m_i/m.

Chapter Organization

The organization of the rest of this paper is as follows. First, in chapter 3 we prove that the removal process is described by a Markov chain. We note the removal map is a deterministic map, and thus the sequence of hypergaphs produced by the removal rule is a process with deterministic transitions starting from one random initial state. If instead we observe random variables at each step which do not reveal the entire graph, then the sequence of successive values of these variables would have random

transitions. For variables with Markovian transitions, this becomes a feasible way to study the process.

Following chapter 3, the main argument can be summarized as a series of successive arguments: (1) that the Markov chain has transitions approximated by modeling the vertices as having Poisson degrees distributions which are truncated and thinned, (2) that for $\log n$ many steps the Markov chain has a trajectory which fluctuates around the trajectory of a deterministically evolving process, (3) the deterministic process tends to a limiting 2-core whose size is as described by the Main Theorem, and finally (4) the size of the hypergraph after only $\log n$ steps faithfully represents the size of limiting 2-core. This is the section-by-section content of the main chapter, chapter 5. The prior chapter, chapter 4 contains supporting Lemmas of probability estimates for the first section of chapter 5.

Chapter 3

Properties of the Removal Map

Let $r : \Omega \longrightarrow \Omega$ denote the map on hypergraphs which removes all hyperedges containing a degree 1 vertex. The purpose of this chapter is to relate the probability measure P and the pushforward measure $P \circ r^{-1} = P(r^{-1}\_)$. The result is a generalization what is referred to as "maintenance of uniformity" in [8] and the Markov degree sequence in [12, 9].

Notationally, P will refer to a probability measure on Ω, with extra hypotheses on P introduced as needed. So in this chapter only, the specific probability measure P from chapter 2.3 will not be referenced by P, although the results in this chapter

can be applied to it as a special case.

As should be expected from working in a rather general setting, the proofs in this chapter use elementary arguments. The more complicated analysis involved in theorem 2.1 are in the following chapters.

It is preferable to represent hypergraphs as incidence matrices for this chapter. A binary matrix with columns indexed by V and rows indexed by E can be viewed as a collection of (possibly empty) subsets of V: each row $e \in E$ represents the set of vertices whose column has a non-zero entry in row e. The collection of such subsets which are non-empty gives a hypergraph on V (whose hyperedges additionally come equipped with distinct labels from E [2])

In this chapter we let Ω denote the set of all $m \times n$ binary matrices, and P will be a measure on Ω directly. It is important that we have chosen to represent hypergraphs as incidence matrices with possibly empty rows so that the removal map $r : \Omega \longrightarrow \Omega$ defines a pushforward measure $P \circ r^{-1}$ that correctly represents the distribution after one removal step of a random initial hypergraph.

[2]One would usually consider the set of all hypergaphs on V as just the plain collection of all subsets of V, and using incidence matrices is similar to considering an ordered collection of subsets of V. Any single hypergraph is represented in multiple ways according to how to the hyperedges are ordered as rows of the matrix.

3.1 General Properties

We begin with a section that is nonprobabilistic, and therefore requires no assumption on P.

For any $\omega \in \Omega$ define $A(\omega) \subseteq [n]$ to be the set of column indices whose columns contain at least one non-zero entry, define $B(\omega) \subseteq [n]$ to be set of column indices whose columns contain exactly one non-zero entry, and define $R(\omega) \subseteq [m]$ to be the set of row indices whose rows contain at least one entry.

For any $\omega_1, \omega_2 \in \Omega$ define $\omega_1 \sqcup \omega_2$ to be the matrix sum $\omega_1 + \omega_2$ if $R(\omega_1)$ and $R(\omega_2)$ are disjoint, and undefined otherwise. In terms of hypergraphs, $\omega_1 \sqcup \omega_2$ represents a union of two hypergraphs, but is undefined if there would be two hyperedges sharing the same label. Statements involving $\omega_1 \sqcup \omega_2$ implicitly assume the quantity is defined. This notation is useful for referring to elements in the inverse image $r^{-1}\omega$, since if $\omega_1 \in r^{-1}(\omega_2)$ then there is a unique decomposition $\omega_1 = s \sqcup \omega_2$.

Theorem 3.1 (Functional Properties)**.** Let $\omega_1, \omega_2 \in \Omega$. Then:

1. $A(\omega_1 \sqcup \omega_2) = A(\omega_1) \cup A(\omega_2)$.

2. $B(\omega_1 \sqcup \omega_2) = [B(\omega_1) \setminus A(\omega_2)] \cup [B(\omega_2) \setminus A(\omega_1)]$.

3. $R(\omega_1 \sqcup \omega_2) = R(\omega_1) \cup R(\omega_2)$.

Proof. The proof is straightforward and immediate after expanding definitions. □

Theorem 3.2 (Characterization of the Inverse Image)**.** Let $k \geq 0$. For $\omega \in \Omega$ define $S(\omega) \subseteq \Omega$ as the unique set such that $\{s \sqcup \omega : s \in S(\omega)\} = r^{-k}\omega$. Then $S(\omega)$ depends only on $A(\omega)$, $B(\omega)$, and $R(\omega)$.

Let $\mathfrak{m}_1 \subseteq \mathfrak{m}_2 \subseteq [m]$. For $\omega \in \Omega$ such that $R(\omega) \subseteq \mathfrak{m}_1$, define $r^{-k}_{\mathfrak{m}_2\mathfrak{m}_1}\omega = \{s \sqcup \omega : s \in S_{\mathfrak{m}_2\mathfrak{m}_1}(\omega)\}$ where $S_{\mathfrak{m}_2\mathfrak{m}_1}(\omega) = \{s \in S(\omega) : R(s) = \mathfrak{m}_2 \setminus \mathfrak{m}_1\}$. Then $S_{\mathfrak{m}_2\mathfrak{m}_1}(\omega)$ depends only on $A(\omega)$ and $B(\omega)$.

Comments. The notation $r^{-k}_{\mathfrak{m}_2\mathfrak{m}_1}$ defines a mapping $r^{-k}_{\mathfrak{m}_2\mathfrak{m}_1} : \Omega \longrightarrow 2^\Omega$ named by the symbol $r^{-k}_{\mathfrak{m}_2\mathfrak{m}_1}$, but it is not the inverse image of a map $\Omega \longrightarrow \Omega$. The side condition $R(s) = \mathfrak{m}_2 \setminus \mathfrak{m}_1$ in $S_{\mathfrak{m}_2\mathfrak{m}_1}(\omega)$ indicates that we have restricted the inverse image $r^{-k}\omega$ so that only a prespecified set of rows, $\mathfrak{m}_2 \setminus \mathfrak{m}_1$, is removed by r^k in $r^k(s \sqcup \omega)$.

Some motivation is in order. The first part of this theorem is the key property for proving the triple (A, B, R) is a Markov chain (case (1) of Theorem 3.4 from the next section). However in the probability space we intend to work with in chapter 5, there is something inconvenient about a Markov chain involving R: conditioning that some row is not the zero row introduces dependence among the entries of that row.

Contrast with conditioning that some row is not a *deleted* row—the difference being that an initially zero row no longer counts—which does not introduce dependence. The second part of Theorem 3.2 is the key property for proving we do in fact get a Markov chain using the non-deleted rows in place of the non-zero rows in our Markov chain triple (case (2) of Theorem 3.4).

Proof of Theorem 3.2. Begin with $k = 1$. Let $\omega \in \Omega$ be given. $s \in S(\omega)$ if and only if all of the following hold:

1. $R(s)$ and $R(\omega)$ are disjoint, ensuring $s \sqcup \omega$ is defined.
2. $B(\omega) \subseteq A(s)$, ensuring no row of ω is removed by r in $r(s \sqcup \omega)$.
3. Every row of $R(s)$ contains an non-zero entry in some column from $B(s) \setminus A(\omega)$, ensuring every row of $R(s)$ is removed by r in $r(s \sqcup \omega)$.

This is evidently determined by $A(\omega)$, $B(\omega)$, and $R(\omega)$ alone.

Refer to conditions (1)-(3) as $\Gamma(s, \omega)$. In general for $k > 1$ the same argument applies and $r^{-k}\omega$ is the set of all $s_1 \sqcup s_2 \sqcup \ldots \sqcup s_k \sqcup \omega$ which satisfy the conditions $\Gamma(s_i, s_{i+1} \sqcup s_{i+2} \sqcup \ldots \sqcup s_k \sqcup \omega)$ for $i = 1, 2, \ldots, k$. Making repeated use of Theorem 3.1, these conditions are expressible in terms of $A(\omega)$, $B(\omega)$, and $R(\omega)$.

For $S_{\mathfrak{m}_2\mathfrak{m}_1}(\omega)$ the argument only needs to be adapted by replacing condition 1 with the condition $R(s) = \mathfrak{m}_2 \setminus \mathfrak{m}_1$. □

3.2 Independent Rows

Assume now that P is a measure where the rows of the matrix are independent (corresponding to hypergraphs generated by a sequence of independent hyperedges). Given $\omega, s \in \Omega$, $P(\omega \sqcup s)$ can be expressed as a product of weights $P_1(\omega)\, P_2(s)$ for ω and s. That is, view Ω as a product space of row sets $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_m$ with P represented as product measure $\mu_1 \times \mu_2 \times \cdots \times \mu_m$. Formally, defining the functions P_1 and P_2 in terms of the μ_i's is a rather arbitrary construction — even with the most natural definition we must somehow decide how any empty rows of $\omega \sqcup s$ get "assigned" to ω and s for purposes of contributing to the weight of $P_1(\omega)$ or $P_2(s)$. It will be desirable that P_1 is defined consistently so that it is a probability measure on certain subsets of Ω.

We consider two ways to proceed: (1) Partition Ω via R, and when $R(\omega) = \mathfrak{m}_1$ define $P_1(\omega) = \Pi_{i \in \mathfrak{m}_1} \mu_i(\omega_i)$. Then P_1 is defined over all Ω and is a probability measure on any partition $\{\omega : R(\omega) = \mathfrak{m}\}$. (2) Presuppose $R(\omega) \subseteq \mathfrak{m}_1$ and define $P_1(\omega) =$

$\Pi_{i\in\mathfrak{m}_1}\mu_i(\omega_i)$. Then P_1 is defined over just a single fixed subset $\{\omega : R(\omega) \subseteq \mathfrak{m}_1\}$ and is a probability measure on this set.

We are going to consider both definitions (1) and (2) simultaneously. A small amount of flexibility in notation permits a single argument to handle both cases. These two cases correspond to the two parts of Theorem 3.4.

Definitions: Let r_0^{-i} denote either r^{-i} or $r^{-i}_{\mathfrak{m}_2\mathfrak{m}_1}$ depending on whether we consider case (1) or case (2). Let S_0 denote the corresponding choice for S or $S_{\mathfrak{m}_2\mathfrak{m}_1}$, and let X_0 denote the corresponding choice for (A, B, R) or (A, B). Define $X_i = X_0(r^i\omega)$, $R_i(\omega) = R(r^i\omega)$, $R_{i,j} = R_i \backslash R_j$ and $X_i' = (A_i, B_i, R_{0,i})$. Let $k \geq 0$ and let E denote the event $\{X_0 = x_0, X_1 = x_1, \ldots, X_{k-1} = x_{k-1}\}$ or $\{X_0' = x_0', X_1' = x_1', \ldots, X_{k-1}' = x_{k-1}'\}$ depending on whether we consider case (1) or case (2).

For $\omega \in \Omega$ we can compute

$$\begin{aligned} P\big(r_0^{-k}\omega \mid E\big) &= \sum_{s\in S_0(\omega)} P(\omega \sqcup s \mid E) = \sum_{s\in S_0(\omega)} P(\omega \sqcup s)\, 1_{\omega\sqcup s\in E}\, (P(E))^{-1} \\ &= \sum_{s\in S_0(\omega)} P_1(\omega)\, P_2(s)\, 1_{\omega\sqcup s\in E}\, (P(E))^{-1}\,. \end{aligned}$$

Factor $P_1(\omega)$ out of the sum, and express the remainder as $z_k(\omega)$ giving

$$P\big(r_0^{-k}\omega \mid E\big) = z_k(\omega)\, P_1(\omega)\,. \tag{3.1}$$

In the case $r_0^{-k} = r^{-k}$, the proportionality factor $z_k(\omega)$ is a function of $A(\omega)$, $B(\omega)$,

and $R(\omega)$ by Theorems 3.1 and 3.2[3], while in the case $r_0^{-k} = r_{\mathfrak{m}_2\mathfrak{m}_1}^{-k}$, $z_k(\omega)$ is a function of $A(\omega)$ and $B(\omega)$. If additionally P respects permutation of rows (or columns), then $z_k(\omega)$ depends on $R(\omega)$ only up to cardinality (respectively, $z_k(\omega)$ depends on $A(\omega)$ and $B(\omega)$ only up to their cardinalities). Independence of columns plays no special role. Requiring permutable rows is equivalent to requiring identically distributed rows since we've already assumed independence.

We know that on an appropriate subset P_1 is a probability measure. Let Q_k denote the measure $P(\_ \mid E)$. The function z_k is the Radon-Nikodym derivative of Q_k with respect to P_1. If we restrict these two probability measures, via conditioning, to a subset where the Radon-Nikodym derivative is constant then the restricted measures must in fact be equal. That is, an equation like (3.1) says P_1 and $Q_k \circ r_0^{-k}$ are conditionally the same measure, conditioned on the random variable z_k. We phrase this observation as the following lemma.

Lemma 3.3 (Pushforward Lemma)**.** Let $k \geq 0$, and let Q_k and z_k be as defined above. Let F be an event on which z_k is constant. Then, when the conditional probabilities are defined,

$$Q_k\big(r_0^{-k}\_ \mid r_0^{-k}F\big) = P_1(\_ \mid F)\,.$$

[3]This requires X_0 to be the particular triple (A, B, R), and not simply any variable with respect to which (A, B, R) is measurable. A generalization in that direction is possible if the variable also satisfies an analogue of theorem 3.1.

Proof. The proof is just a fully explicated version of the observation preceding the statement. Summing (3.1) over $\omega \in F$ to find the constant, we have $z = z_F = Q_k\big(r_0^{-k}F\big)/P_1(F)$ on F. For any event T, use (3.1) again this time summing over $\omega \in T \cap F$ to get $Q_k\big(r_0^{-k}T \cap r_0^{-k}F\big) = z_F P_1(T \cap F)$. Substitute for z_F in this last equation and rearrange to get the theorem. $\square$

Theorem 3.4 (Markov Theorem)**.**

1. Let $X_i = (A_i, B_i, R_i)$. The sequence $X_0, X_1, \ldots$ is a Markov chain with transition kernel $p(x_0, x_1) = P(X_1 = x_1 \mid X_0 = x_0)$.

2. Let $X_i' = (A_i, B_i, R_{0,i})$. The sequence $X_0', X_1', \ldots$ is a Markov chain with transition kernel $p(x_0, x_1) = P(X_1' = x_1 \mid X_0' = x_0)$.

Proof. Let Y_0 denote a random variable such that $\sigma(X_0) \subseteq \sigma(Y_0)$, which will be chosen later. Define $Y_k = Y_0 \circ r^k$. Let F denote the event $\{Y_0 = y\}$, so that $r^{-k}F = \{Y_k = y\}$ and $r^{-k}_{\mathfrak{m}_2\mathfrak{m}_1}F = \{Y_k = y\} \cap \{R_0 \setminus R_k = \mathfrak{m}_2 \setminus \mathfrak{m}_1\}$. Since X_0 is a measurable function of Y_0, z_k is constant on F. Lemma 3.3 states

$$P\big(r_0^{-k}\_ \mid E, r_0^{-k}F\big) = P_1(\_|Y_0 = y)\,. \tag{3.2}$$

To prove 1, let $i \geq 0$ be given. We take $r_0^{-k} = r^{-k}$ and $Y_0 = X_0$ in equation (3.2), and for these choices we recall $E = \{X_0 = x_0, \ldots, X_{k-1} = x_{k-1}\}$ and $r^{-k}F =$

$\{Y_k = y\}$. Evaluate the probability measures in equation 3.2 on the event $\{X_1 = x\}$. Finally, instantiate the resulting equation for both $k = i$ and $k = 0$ to establish

$$P(X_{i+1} = x \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}, X_i = y) = P_1(X_1 = x | X_0 = y)$$

$$P(X_1 = x \mid X_0 = y) = P_1(X_1 = x | X_0 = y)\,.$$

The equality of the left hand sides proves the Markov property.

To prove 2, much is the same until the final step. Let $i \geq 0$ be given and let $x', y' \subseteq [n] \times [n] \times [m]$ be given (x' and y' will play the same roles as x and y in the proof of 1). Denote x' by $(x, \mathfrak{m})$, where $\mathfrak{m} \subseteq [m]$. We take $r_0^{-k} = r_{\mathfrak{m}_2\mathfrak{m}_1}^{-k}$ and $Y_0 = X_0$ in equation (3.2), and for these choices we recall $E = \left\{X'_0 = x'_0, \ldots, X'_{k-1} = x'_{k-1}\right\}$ and $r_{\mathfrak{m}_2\mathfrak{m}_1}^{-k} F = \{X_k = y\} \cap \{R_0 \setminus R_k = \mathfrak{m}_2 \setminus \mathfrak{m}_1\}$. Pick $y, \mathfrak{m}_2$, and $\mathfrak{m}_1$ so that $r_{\mathfrak{m}_2\mathfrak{m}_1}^{-k} F = \{X'_k = y'\}$. Evaluate the probability measures in equation 3.2 on the event $\{X'_1 = x''\}$ where $x'' = (x, \Delta\mathfrak{m})$ and $\Delta\mathfrak{m} = \mathfrak{m} \setminus (\mathfrak{m}_2 \setminus \mathfrak{m}_1)$. Finally, instantiate the resulting equation for both $k = i$ and $k = 0$ to establish

$$P\big(r_{\mathfrak{m}_2\mathfrak{m}_1}^{-i} \{X'_1 = x''\} \mid X'_0 = x'_0, \ldots, X'_{i-1} = x'_{i-1}, X'_i = y'\big)$$

$$= P_1(X'_1 = x'' | X_0 = y) \tag{3.3a}$$

$$P\big(r_{\mathfrak{m}_2\mathfrak{m}_1}^{-0} \{X'_1 = x''\} \mid X'_0 = y'\big) = P_1(X'_1 = x'' | X_0 = y). \tag{3.3b}$$

Note for any $k \geq 0$, one has

$$r_{\mathfrak{m}_2\mathfrak{m}_1}^{-k} \{X'_1 = x''\} = \{X_{k+1} = x, R_{k,k+1} = \Delta\mathfrak{m}, R_{0,k} = \mathfrak{m}_2 \setminus \mathfrak{m}_1\}$$

which by choice of $\Delta\mathfrak{m}$ (namely that $\Delta\mathfrak{m} \cup (\mathfrak{m}_2 \setminus \mathfrak{m}_1) = \mathfrak{m}$) is also the event $\{X'_{k+1} = x'\} \cap \{R_{0,k} = \mathfrak{m}_2 \setminus \mathfrak{m}_1\}$. This later event regarding $R_{0,k}$ is already conditioned on by the left hand sides of equations 3.3 (by the conditioning $X'_k = y'$, where $k = i$ and $k = 0$). Therefore

$$P\big(X'_{i+1} = x' \mid X'_0 = x'_0, \ldots, X'_{i-1} = x'_{i-1}, X'_i = y'\big) = P_1(X'_1 = x'' | X_0 = y)$$

$$P(X'_1 = x' \mid X'_0 = y') = P_1(X'_1 = x'' | X_0 = y)\,.$$

The equality of the left hand sides proves the Markov property. □

In chapter 5, we will be using the Markov chain from case (2). Conditioning on an event like $R_{0,k} = \mathfrak{m}_2 \setminus \mathfrak{m}_1$ conditions on the set of rows (edges) which have been removed after k steps, but it does not stipulate whether any rows in $\mathfrak{m}_1$ were originally empty. For our probability space this has the advantage (over conditioning on R_k) that it preserves the independence of the events $\{j \in e_i\}_{i \in \mathfrak{m}_1}$.

The Markov chain from case (1) is the chain which appears in the literature (referenced in the opening paragraph of this chapter). In the case of the k-core as opposed to the 2-core, the Markov chain would need to be amended by considering the sets of vertices of degrees $1, 2, \ldots, k-1$ and greater equal $k-1$ instead of just 1 and greater equal 1 (naturally one has the freedom to choose, for example, $\geq k$ instead of $\geq k-1$, as these two collection of state variables would have the same

sigma fields). Besides changing the Markov chain itself, the proofs in this chapter would only need a minor alteration by stating the equivalent version of Theorem 3.1 for these state variables (so that for all the degree sets D_i involved, $D_i(\omega_1 \sqcup \omega_2)$ is expressible as combinations of $D_j(\omega_1)$'s and $D_j(\omega_2)$'s).

Proofs of the Markov property from the literature are either given by stating the general principles involved (leaving the details to be checked and without stating the minimal set of assumptions from the model that are actually used) or else are proven using explicit computation of the probabilities involved, i.e. by finding the combinatorial formulae which count the number of hypergraphs with a given property (which, as this chapter demonstrates, is always unnecessary for independent hyperedges).

The hypergraph models referenced from the literature have some form of dependence built in to the hyperedge distribution, arising from conditioning on the size of the hyperedge or conditioning on the degree sequence of the hypergraph itself. They also exhibit homogeneity between vertices, so that the cardinalities of the degree sets in the chain may be used instead of the sets themselves. Finally, they also exhibit an overall symmetry so that $P(\_ \mid X_k = x_k)$ is uniform on the subset of hypergraphs with $\{X_k = x_k\}$ (dubbed "maintenance of uniformity"). The fact that for our model we must work with the sets A and B (and not their cardinalities) as well as nonuniform distributions presents greater complication.

3.3 Removing One Hyperedge at Each Step

In section 5.4 we will want to consider a removal process which iterates a *random* edge removal rule: choose a uniformly random vertex of degree 1, and remove the hyperedge incident to it. This section establishes that the results of the sections 3.1 and 3.2 carry over to this process just as well.

The random removal rule is formalized by supplementing our old probability space with an iid sequence of uniform variables on $[0,1]$. Each random variable represents the randomness we use to decide which vertex of degree 1 to remove at each step. Define then a new probability space $\overline{\Omega} = [0,1]^{\mathbb{N}} \times \Omega$. The new probability measure $P = \mu \times P_\Omega$ on $\overline{\Omega}$ is given as the product $\mu = \mathcal{L} \times \mathcal{L} \cdots$ of uniform (Lebesgue) measure $\mathcal{L}$ on each component $[0,1]$ and a probability measure P_Ω on the component Ω (playing the role of the old probability measure P of the previous section).

We can provide a high level overview of this section in a few sentences. The key property of sections 3.1 and 3.2 is that fiber $r^{-k}\omega$ is an amalgam of ω with a set $S(\omega)$ whose dependence on ω factors through a small set of variables. Since the removal rule is now random, the first difference in this section is that the fiber is additionally an amalgam of the slice of $U(s,\omega) \subseteq [0,1]$ which represents the realizations where we do happen to transition from $s \sqcup \omega$ to ω by r. The second difference is that this

slice is not at all a function of the smaller set of variables — but the *probability* of the slice still is — which is what is needed for z_k from equation 3.1 to be a function of the smaller set of variables. The rest of this section provides a fuller account of the details. The goal though is still to appeal to modification of the of the previous proofs once the new set of definitions is formalized.

Denote elements of $\overline{\Omega}$ as $(\{u_i\}, \omega)$. Define $r : \overline{\Omega} \longrightarrow \overline{\Omega}$ via $(\{u_i\}, \omega) \mapsto (\{u_{i+1}\}, r_{u_1}\omega)$, where for each $u \in [0,1]$, r_u is a fuction $\Omega \longrightarrow \Omega$ so that $r_u\omega$ represents removal of single edge — depending on u and ω — if ω contains at least 1 degree vertex (and $r_u\omega = \omega$ if all vertices have degree at least 2). For example, if ω contains 2 vertices of degree 1, then $r_u\omega$ removes the first if $u \leq 1/2$ and the second if $u > 1/2$. For a fully explicit definition, we shall fix for each $\omega \in \Omega$ an arbitrary partition of $[0,1]$ into $|B(\omega)|$ disjoint subsets of equal measure. When $|B(\omega)| \neq 0$ and u lies in the nth partition then $r_u\omega$ denotes the removal of the nth edge contain a degree 1 vertex.

The next theorem is the analogue of Theorem 3.2 for the random removal rule.

Notation: For $u \in [0,1]$, $v = (v_1, v_2, \dots) \in [0,1]^{\mathbb{N}}$, let $u \diamond v$ denote $(u, v_1, v_2, \dots) \in [0,1]^{\mathbb{N}}$.

Theorem 3.5. Let $k \geq 0$. For $\omega \in \Omega$ define $S(\omega) \subseteq \Omega$ and for each $s \in S(\omega)$ define

$U(s,\omega) \subseteq [0,1]$ as the unique sets so that

$$\bigcup_{v\in[0,1]^{\mathbb{N}}} \{(u \diamond v, s \sqcup \omega) : s \in S(\omega)\,, u \in U(s,\omega)\} = r^{-k}\Big([0,1]^{\mathbb{N}} \times \{\omega\}\Big)\,.$$

Let $\mathfrak{m}_1 \subseteq \mathfrak{m}_2 \subseteq [m]$. For $(v,\omega) \in \overline{\Omega}$ such that $R(\omega) \subseteq \mathfrak{m}_1$, define

$$r^{-k}_{\mathfrak{m}_2\mathfrak{m}_1}(v,\omega) = \{(u \diamond v, s \sqcup \omega) : s \in S_{\mathfrak{m}_2\mathfrak{m}_1}(\omega)\,, u \in U(s,\omega)\}\,.$$

where $S_{\mathfrak{m}_2\mathfrak{m}_1}(\omega) = \{s \in S(\omega) : R(s) = \mathfrak{m}_2 \setminus \mathfrak{m}_1\}$.

Then

1. $S(\omega)$ depends only on $A(\omega)$, $B(\omega)$, and $R(\omega)$.
2. $\mathcal{L}(U(s,\omega))$ depends only on s, $A(\omega)$, and $B(\omega)$.
3. $S_{\mathfrak{m}_2\mathfrak{m}_1}(\omega)$ depends only on $A(\omega)$ and $B(\omega)$.

Proof. The proof is by induction on k and repeated use of Theorem 3.1 just as in the proof of Theorem 3.2. The arguments themselves only need to be adapted slightly. For example, to prove 1 in the case $k = 1$, we change condition 1 to be that $R(s)$ is either a single row disjoint from $R(\omega)$ or the empty set. And to prove 2 in the case $k = 1$, we have $\mathcal{L}(U(s,\omega)) = \frac{B(s\sqcup\omega)\cap A(s)}{B(s\sqcup\omega)} = \frac{B(s)\setminus A(\omega)}{B(s\sqcup\omega)}$, which depends only on s, $A(\omega)$, and $B(\omega)$. □

The rest of this section mirrors the development of section 3.2. Assume that P is a measure with independent rows. We again consider two cases, corresponding to the cases where r_0^{-k} denotes either r^{-k} or $r_{\mathfrak{m}_2\mathfrak{m}_1}^{-k}$ (each of which now refer to the functions defined in this section).

We will copy the notation for random variables of section 3.2 ($A_i, B_i, R_i, R_{i,j}, X_i$, and X_i') for our new probability space by regarding them as functions of Ω (and not $\overline{\Omega}$). Define $\pi_2 : \overline{\Omega} \to \Omega$ by $\pi_2(\{u_i\}, \omega) = \omega$. Then, for example, $X_i \circ \pi_2$ is a random variable on $\overline{\Omega}$. For each $k \geq 0$ we also copy the notation for the event $E \subseteq \Omega$ as $\{X_0 = x_0, X_1 = x_1, \ldots, X_{k-1} = x_{k-1}\}$ or $\left\{X_0' = x_0', X_1' = x_1', \ldots, X_{k-1}' = x_{k-1}'\right\}$ depending on whether we consider case (1) or case (2).

The biggest change from section 3.2 is the argument that precedes the Pushforward Lemma. We now have that for $\omega \in \Omega$,

$$\begin{aligned}
P\left(r_0^{-k}\pi_2^{-1}\omega \mid \pi_2^{-1}E\right) &= \sum_{s \in S_0(\omega)} \mathcal{L}(U(s,\omega))\, P_\Omega(\omega \sqcup s \mid E) \\
&= \sum_{s \in S_0(\omega)} \mathcal{L}(U(s,\omega))\, P_\Omega(\omega \sqcup s)\, 1_{\omega \sqcup s \in E} \left(P_\Omega(E)\right)^{-1} \\
&= \sum_{s \in S_0(\omega)} \mathcal{L}(U(s,\omega))\, P_1(\omega)\, P_2(s)\, 1_{\omega \sqcup s \in E} \left(P_\Omega(E)\right)^{-1} \\
&= z_k(\omega)\, P_1(\omega)\,.
\end{aligned}$$

The conclusion that $z_k(\omega)$ depends only on $A(\omega)$, $B(\omega)$, and $R(\omega)$ in case (1) and only on $A(\omega)$ and $B(\omega)$ in case (2) is now justified using Theorem 3.5. The rest of section 3.2 follows a nearly identical development, and we only state the results.

Lemma 3.6 (Pushforward Lemma)**.** Let $k \geq 0$, let z_k be as defined above, and let F be an event on which z_k is constant. Then, when the conditional probabilities are defined,

$$P\big(r_0^{-k}\pi_2^{-1}\_ \mid \pi_2^{-1}E, r_0^{-k}\pi_2^{-1}F\big) = P_1(\_ \mid F)\,.$$

Theorem 3.7 (Markov Theorem)**.**

1. Let $X_i = (A_i, B_i, R_i)$. The sequence $X_0 \circ \pi_2, X_1 \circ \pi_2, \ldots$ is a Markov chain with transition kernel $p(x_0, x_1) = P(X_1 = x_1 \mid X_0 = x_0)\,.$

2. Let $X_i' = (A_i, B_i, R_{0,i})$. The sequence $X_0' \circ \pi_2, X_1' \circ \pi_2, \ldots$ is a Markov chain with transition kernel $p(x_0, x_1) = P(X_1' = x_1 \mid X_0' = x_0)\,.$

Chapter 4

Preliminary Lemmas

This chapter contains simple probability estimates which are needed later in the course of the main proofs when we compare the Markov chain to the Poisson chain. It is convenient to have these stated here separately for they are statements which can be understood independently of the main probability space and the proofs can be skipped or deferred without loss of understanding.

Theorem 4.1 (Balls in Boxes)**.** Let $X_1, X_2, \ldots, X_b$ be iid with X_i uniformly distributed on $\{1, 2, \ldots, m\}$ and let $Y = m - |\{X_1, X_2, \ldots, X_b\}|$. Then, uniformly in b,

$$P\left(\left|Y - me^{-\frac{b}{m}}\right| \geq s\right) \leq 2e^{-\frac{s^2}{4b}+O\left(\frac{s}{b}\right)},$$

as $m \to \infty$ and $\frac{b}{m}$ ranges over any compact subset of $[0, \infty)$.

As will be the case in all proofs, the implicit sense of convergence (namely, uniformity) meant by any $O(\,\cdot\,)$'s used in the proof is the explicit sense of convergence used in the statement.

Proof. Y is given by

$$Y = \sum_{i=1}^{m} 1_{i \notin \{X_1, X_2, \ldots, X_b\}}$$

and so taking expectations,

$$\begin{aligned} EY &= m\left(1 - \frac{1}{m}\right)^b = me^{b\left(-\frac{1}{m} + O\left(\frac{1}{m^2}\right)\right)} \\ &= me^{-\frac{b}{m} + O\left(\frac{1}{m}\right)} = me^{-\frac{b}{m}} + O(1). \end{aligned}$$

The tail bound comes from McDiramid's inequality, since the outcome of any one of the random variables X_i affects Y by 1. □

Theorem 4.2 (Poisson Approximation)**.** For any $\lambda \geq 0$ and $m \geq 0$ let X be distributed as $\mathrm{Bin}(m, \lambda/m)$. Let $k \geq 0$, then uniformly in λ,

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} + O\left(\frac{1}{m}\right)$$

as $m \to \infty$ and λ/m ranges over any compact subset of $[0, 1)$.

Remark 4.1. The assumption is not that λ tends to a limit. The result includes $\lambda \to \infty$, $\lambda = 0$, and anything in between. It is naturally not uniform in k.

Proof. The proof is direct but careful consideration of the type of convergence is needed during some steps:

$$\begin{aligned}
P(X=k) &= \binom{m}{k}\left(\frac{\lambda}{m}\right)^k\left(1-\frac{\lambda}{m}\right)^{m-k} = \frac{m_{(k)}}{k!}\frac{\lambda^k}{m^k}\left(1-\frac{\lambda}{m}\right)^m\left(1-\frac{\lambda}{m}\right)^{-k} \\
&= \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{m}\right)^m(1+o(1)) \\
&= \frac{\lambda^k}{k!}e^{m\left(-\frac{\lambda}{m}-\left|O\left(\frac{\lambda^2}{m^2}\right)\right|\right)}(1+o(1)) \\
&= \frac{\lambda^k}{k!}e^{-\lambda-\left|O\left(\frac{\lambda^2}{m}\right)\right|}(1+o(1)) \\
&= \frac{\lambda^k}{k!}e^{-\lambda}\left(1+O\left(\frac{\lambda^2}{m}\right)\right) = \frac{\lambda^k}{k!}e^{-\lambda}+O\left(\frac{1}{m}\right).
\end{aligned}$$

In the last step we have used that $\lambda^{k+2}e^{-\lambda}$ is a bounded function of λ, and hence is $O(1)$. $\square$

For more complex Poisson limits let us introduce the probability generating function $\varphi_{\lambda,p}(z)$ where λ, $p \geq 0$ and

$$\varphi_{\lambda,p}(z) = \frac{e^{\lambda p(z-1)} - \lambda p z e^{-\lambda} - (1+\lambda(1-p))e^{-\lambda}}{1-(1+\lambda)e^{-\lambda}}.$$

This is the pgf for $(Z_1|Z_1+Z_2 \geq 2)$ where Z_1 is distributed as poisson(λp) and Z_2 is distributed as poisson$(\lambda(1-p))$. By the thinning lemma, this distribution is also the result of independent thinning of $(Z|Z \geq 2)$ with retention probability p where Z is distributed as poisson(λ).

Notation: Let $\left[z^k\right] f(z)$ denote the coefficient of z^k in the power series expansion

of $f(z)$, and set $\varphi_k(\lambda, p) \coloneqq \left[z^k\right] \varphi_{\lambda,p}(z)$.

Theorem 4.3 (Poisson Approximation 2)**.** For any $0 \leq p \leq 1$, $\lambda \geq 0$ and $m \geq 0$, let X be distributed as $\mathrm{Bin}(mp, \lambda/m)$, Y be distributed as $\mathrm{Bin}(m\,(1-p)\,, \lambda/m)$, and Z distributed as $(X|X+Y \geq 2)$. Let $k \geq 0$. Then, uniformly in λ and p,

$$P(Z = k) = \varphi_k(\lambda, p) + O\left(\frac{1}{m}\right)$$

as $m \to \infty$, λ/m ranges over a compact subset of $[0, 1)$, and p ranges over a compact subset of $(0, 1)$.

Proof. The idea is to express the point probabilities for Z in terms of that of X and Y, and then use the previous theorem twice, once for X (with $\lambda \mapsto \lambda p$, $m \mapsto mp$) and once for Y (with $\lambda \mapsto \lambda\,(1-p)\,$, $m \mapsto m\,(1-p)$). This will give the right estimate under the limits $mp \to \infty$ and $m\,(1-p) \to \infty$, which are equivalent to $m \to \infty$ given the restriction on p. □

Chapter 5

The Size of the 2-core from the Removal Process

This chapter contains the main results. We let $m_i = m - |R_{0,i}|$ denote the number of surviving edges at step i, $b_i = |B_i|$ denote the number of degree 1 vertices, and we let $\deg^{(i)}(j)$ denote the degree of vertex j at step i.

Let $\hat{V}_k^{(i)}$ denote the random counting measure for the number of degree k vertices at step i. That is, the measure assigned to a set of vertices $W \subseteq \{0, 1, \ldots, n-1\}$ is the number of degree k vertices in W at step i. Let the random measures $V_k^{(i)}$

denote the pushforward of $\hat{V}_k^{(i)}$ by the map $j \mapsto j/n$ from $\{0, 1, \ldots, n-1\} \longrightarrow [0, 1]$. Equivalently, $V_k^{(i)}$ is the random measure on $[0, 1]$ defined by

$$V_k^{(i)} = \sum_{j=0}^{n-1} \delta_{\frac{j}{n}} 1_{\deg^{(i)}(j)=k}$$

where δ_x is the Dirac delta measure located at x. We also employ notation such as $V_{\geq 2}^{(i)}$ for $\sum_{k\geq 2} V_k^{(i)}$. The motivation behind the measures $V_k^{(i)}$ is that they are equivalent encodings of the sets of vertices of a given degree. When we discuss the trajectory of the markov chain (m_i, A_i, B_i), we will need to speak of two trajectories being near one another. For sets of vertices, the nearness of two sets will be formulated through the signed measure equal to the difference of the two measures.

Finally we will write, for example, P_{m_i,A_i,B_i} for the conditional probability $P(\_|m_i, A_i, B_i)$.

5.1 Approximations of the Removal Chain's Single Step Transitions

Theorem 5.1 (0th-step)**.** Let $k \geq 0$.

(i) For any interval $U \subseteq [0,1]$ and as $n \to \infty$,

$$E \int_U dV_k^{(0)} = \int_U \frac{\lambda(x)^k}{k!} e^{-\lambda(x)} \, dV_{\geq 0}^{(0)}(x) + O(1)$$
$$= n \int_U d\mu_k + O(1),$$

where $\lambda(x) = \frac{ac_{\text{den}}}{x}$ and μ_k is the non-random measure with Lebsegue density

$$\frac{d\mu_k}{d\mathcal{L}}(x) = \frac{\lambda(x)^k}{k!} e^{-\lambda(x)} \qquad \frac{d\mu_k}{d\mathcal{L}}(0) = 0.$$

(ii) There exists $c \geq 0$ so that

$$P\left(\sup_{0 \leq t \leq 1} \left| \int_t^1 dV_k^{(0)} - E \int_t^1 dV_k^{(0)} \right| \geq s \right) \leq 2e^{-cs^2/n}.$$

We will need to quote a general result about the maximum of a random walks.

Lemma 5.2. Let $S_n = X_1 + X_2 + \cdots + X_n$ where the X_i are independent 0-1 Bernoulli variables. Let $Y_n = S_n - ES_n$, and let $M_n = \sup_{1 \leq i \leq n} Y_i$. Then there exists $c > 0$ so that

$$P(M_n \geq s) \leq 2e^{-cs^2/n}.$$

Proof of Theorem 5.1. To prove (i), let $U = [u_1, u_2] \subseteq [0,1]$ be given. Write $\hat{\lambda}_j$ for

$\frac{am}{2a+j}$, and λ_j for $\lambda(j/n) = \frac{am}{j}$, and let $\varphi(\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}$. We have

$$\begin{aligned} EV_k^{(0)}(U) &= E \sum_{j\,:\,j/n\in U} 1_{\deg^{(0)}(j)=k} \\ &= \sum_{j\,:\,j/n\in U} \left(\varphi\left(\hat{\lambda}_j\right) + O\left(\frac{1}{m}\right)\right) = \sum_{j\,:\,j/n\in U} \varphi\left(\hat{\lambda}_j\right) + O\left(1\right) \\ &= \sum_{j\,:\,j/n\in U} \left(\varphi(\lambda_j) + O\left(\frac{1}{n}\right)\right) + O\left(1\right) \\ &= n\left(\sum_{j\,:\,j/n\in U} \varphi(\lambda_j)\,\frac{1}{n}\right) + O\left(1\right). \end{aligned}$$

To explain the Poisson estimate, see Theorem 4.2. We claim the Riemann sum in parentheses is $\int_{u_1}^{u_2} \frac{\lambda(x)^k}{k!}e^{-\lambda(x)}\,du + O\left(\frac{1}{n}\right)$. Indeed, the sum converges since the integrand is continuous and the error in approximating a definite integral $\int_a^b f(u)\,du$ can be bounded by general considerations to be at most

$$(\#\text{terms})\,|f'(z)|\,(\text{max step size})^2/2$$

for some $z \in [a,b]$. Such an error term is $O\left(\frac{1}{n}\right)$ in our case. This shows overall $EV_k^{(0)}(U) = n\mu_k(U) + O\left(1\right)$.

To prove (ii), we note this is exactly an instance of Lemma 5.2 where the sequence $X_1,\ X_2,\ \ldots$ is the sequence $1_{\deg^{(0)}(n-1)=k},\ 1_{\deg^{(0)}(n-2)=k},\ \ldots$. □

Remark 5.1. Using the identity $V_{\geq 2}^{(0)} = V_{\geq 0}^{(0)} - V_1^{(0)} - V_0^{(0)}$ where $V_{\geq 0}^{(0)} = \sum_{j=0}^{n-1}\delta_{j/n}$, the theorem can be applied to give a tail estimate for $V_{\geq 2}^{(0)}$. This is somewhat subtle

since the theorem, not being uniform in k, apparently cannot be used to estimate an expression like $\sum_{k=2}^{m} V_k^{(0)}$.

Theorem 5.1 provides a sense in which the random initial state (the 0th transition) of the removal chain differs from the expected empirical distribution where we model the graph as having independent Poisson degrees. The next task is to provide a similar statement for how the removal chain at step $i+1$ differs from step i (the ith transition).

The degree of a vertex j post removal of the edges $R_{i,i+1}$ is given by

$$\deg^{(i+1)}(j) = 1_{j \in A_i \setminus B_i} \sum_{e \in [m] \setminus R_{0,i} \setminus R_{i,i+1}} 1_{j \in e}. \tag{5.1}$$

Consider a vertex $j \in A_i \setminus B_i$. Using the Markov property (part 2 of Theorem 3.4) we get that under P_{m_i,A_i,B_i} the indicators $\{1_{j \in e}\}_{e \in [m] \setminus R_{0,i}}$ are identically distributed Bernoullis and dependent only up to their sum being at least 2. This remains true conditional on $R_{i,i+1}$ — under P_{m_i,A_i,B_i}, the vertices in $A_i \setminus B_i$ are independent of those in B_i, and hence independent with $R_{i,i+1}$. Using this fact together with equation (5.1), we have that the distribution of $\deg^{(i+1)}(j)$ under $P_{m_i,A_i,B_i,R_{i,i+1}}$ only depends on $R_{i,i+1}$ up to cardinality, and so $\deg^{(i+1)}(j)$ has the same distribution under $P_{m_i,A_i,B_i,m_{i+1}}$ as it has under $P_{m_i,A_i,B_i,R_{i,i+1}}$. Note this distribution is described by Theorem 4.3.

We formulate two theorems to describe the ith transition, Theorems 5.3 and 5.4.

In both theorems we express how the random transition differs from an expected empirical distribution for Poisson-type degrees, more specifically where the degrees are given by independent, truncated Poisson random variables (representing conditioning that the vertex lies in A_i) that are subsequently thinned (representing the edges $R_{i,i+1}$ which are removed). The difference between the two theorems is that in the first we speak about probabilities conditioned on m_{i+1}, effectively treating the thinning probability as a free parameter. In the second theorem we only condition on step i, so that the amount of thinning is itself random.

In the second theorem, we compare the random transition to an expected empirical distribution where the comparison model treats the number of degree 1 vertices in any fixed hyperedge as Poisson distributed with mean b_i/m_i. This gives $1-e^{-b_i/m_i}$ as the fraction of edges which are removed, or simply e^{-b_i/m_i} as the thinning parameter in the comparison model.

Notation. In the statement of the following theorem and elsewhere, an expression "$g \bullet \xi$" where $g(x)$ is a function and ξ is a measure denotes the measure $A \mapsto \int_A g(x)\ d\xi(x)$.

Theorem 5.3 (1-step)**.** For every $i \geq 0$ define random variables u_i by

$$m_{i+1} = m_i e^{-\frac{b_i}{m_i}} + u_i$$

and for every $i \geq 0$ and $k \geq 0$ define random signed measures $\xi_k^{(i)\star}$ and $\xi_k^{(i)}$ by

$$V_k^{(i+1)} = 1_{k=0} V_{\leq 1}^{(i)} + \varphi_k\big(\lambda^{(i)}, p_{i^\star}\big) \bullet V_{\geq 2}^{(i)} + \xi_k^{(i)\star}$$

and

$$V_k^{(i+1)} = 1_{k=0} V_{\leq 1}^{(i)} + \varphi_k\big(\lambda^{(i)}, p_i\big) \bullet V_{\geq 2}^{(i)} + \xi_k^{(i)}$$

where $\lambda^{(i)}(x) = \frac{\lambda(x) m_i}{m}$, $p_{i^\star} = \frac{m_{i+1}}{m_i}$, and $p_i = e^{-\frac{b_i}{m_i}}$.

(i) As $m_i \to \infty$ and for b_i/m_i bounded above,

$$P_{m_i, A_i, B_i}(|u_i| \geq s) \leq 1_{b_i > 0} 2 e^{-\frac{s^2}{4b_i} + O\left(\frac{s}{b_i}\right)}$$

uniformly in $i \geq 0$.

(ii) Let $k \geq 0$. There exists $c \geq 0$ so that as $m_i \to \infty$, $m_i - m_{i+1} \to \infty$,

$$P_{m_i, A_i, B_i, m_{i+1}}\left(\sup_{0 \leq t \leq 1} \left| \int_t^1 d\xi_k^{(i)\star} \right| \geq s\right) \leq 2 e^{-\frac{cs^2}{n} + O\left(\frac{s}{m_i}\right)}$$

uniformly in $i \geq 0$, $s \geq 0$.

As will be the case in all proofs, the implicit sense of convergence (namely, uniformity) meant by any $O(\,\cdot\,)$'s used in the proof is the explicit sense of convergence used in the statement.

Proof. The proof of part (i) is provided by Theorem 4.1. The prove part (ii), we follow the structure of Theorem 5.1. For convenience, let $P_{i^\star}$ denote $P_{m_i, A_i, B_i, m_{i+1}}$.

Let $k \geq 0$ and $U = [u_1, u_2] \subseteq [0,1]$ be given. Writing $\lambda_j^{(i)}$ for $\lambda^{(i)}(j/n) = m_i/j$, then

$$\begin{aligned} E_{i^\star} V_k^{(i+1)}(U) &= E_{i^\star} \sum_{j/n \in U} 1_{j \in A_i \setminus B_i} 1_{\deg^{(i+1)}(j)=k} \\ &= \sum_{j/n \in U} 1_{j \in A_i \setminus B_i} P_{i^\star}\big(\deg^{(i+1)}(j) = k\big) \\ &= \sum_{j/n \in U} 1_{j \in A_i \setminus B_i} \left(\varphi_k\Big(\lambda_j^{(i)}, p_{i^\star}\Big) + O\left(\frac{1}{m_i}\right)\right) \\ &= \int_U \varphi_k\big(\lambda^{(i)}, p_{i^\star}\big)\ dV_{\geq 2}^{(i)} + O\left(\frac{n}{m_i}\right). \end{aligned}$$

To explain the probability estimate, see Theorem 4.3 (with $m \mapsto m_i$, $\lambda \mapsto \lambda_j^{(i)}$, $p \mapsto p_{i^\star}$, and $\frac{\lambda}{m} \leq \frac{a}{2}$) recalling the discussion preceding the present theorem about the law of $\deg^{(i+1)}(j)$.

Finally, to get the tail bound we use that the same tail bound in Theorem 5.1 holds here. We can account for an extra deviation $O\left(\frac{n}{m_i}\right)$ by replacing $s \mapsto s - O\left(\frac{n}{m_i}\right)$, which gives the bound in the theorem. $\square$

Theorem 5.4 (Combined 1-step)**.** Let $k \geq 0$. There exists $c_1, c_2 > 0$ so that as $m_i \to \infty$, $b_i \to \infty$, b_i/m_i bounded above,

$$P_{m_i, A_i, B_i}\left(\sup{}_{0 \leq t \leq 1} \left| \int_t^1 d\xi_k^{(i)} \right| \geq s \right) \leq 4e^{-\frac{s^2}{c_1 b_i \vee c_2 n} + O\left(\frac{s}{m_i}\right) + O\left(\frac{s}{b_i}\right)}$$

uniformly in $i \geq 0$, $s \geq 0$.

Proof. The theorem has proper conditions for part (i) of Theorem 5.3 to hold. We

also claim that the conditions on m_i and b_i imply that $m_i - m_{i+1} \to \infty$. Importantly then, the $P_{m_i,A_i,B_i,m_{i+1}}$-probability upper bound in part (ii) of Theorem 5.3 holds here as P_{m_i,A_i,B_i}-probability bound. The justification is that the P_{m_i,A_i,B_i}-probability can be expressed by integrating this $P_{m_i,A_i,B_i,m_{i+1}}$-probability against the law of m_{i+1}. And since the upper bound on the integrand is uniform in m_{i+1}, it carries through as an upper bound on the integral.

Let $f(x) = 1_{t \geq x}$. The proof now is to combine the tail bounds from part (i) and part (ii) using that

$$\left| \int f \, d\xi_k^{(i)} \right| > s \Rightarrow \left| \int f \, d\xi_k^{(i)\star} \right| > s/2 \text{ or } \left| \int f \, d\left(\xi_k^{(i)} - \xi_k^{(i)\star} \right) \right| > s/2. \tag{5.2}$$

The function $\varphi_k = \varphi_k(\lambda, p)$ has bounded partial derivative $\frac{\partial \varphi_k}{\partial p}$ over the domain $\lambda \geq 0$ and $0 \leq p \leq 1$. Therefore the deviation $\left| \varphi_k\left(\lambda^{(i)}, p_{i^\star}\right) - \varphi_k\left(\lambda^{(i)}, p_i\right) \right|$ is at most some $c_1 > 0$ times $|p_{i^\star} - p_i| = |u_i|$. We conclude

$$\left| \int f \, d\left(\xi_k^{(i)} - \xi_k^{(i)\star} \right) \right| \leq c_1 \, |u_i| \, .$$

Taking the union bound of the right hand side of (5.2) gives the upper bound in the theorem. $\square$

5.2 The Event that the Removal Chain Mimics a Fixed Trajectory

Our next goal is to describe the deviation of the removal chain's random trajectory from some fixed, deterministic trajectory. Theorems 5.4 and 5.3 provide a deterministic limit law for the single step transitions of the removal chain. The candidate limit law for a cumulative number of transitions i has the form suggested by iterating the single step limit law. Put simply, the limit at time i is obtained by performing the deterministic transitions $0 \to 1 \to 2 \to \ldots \to i$ according to right hand sides of the previous theorems, discarding the error terms u_i and $\xi_k^{(i)}$. We may think of the iterated limit itself being a (deterministically) evolving process, or more appropriately as a discrete time dynamical system.

Let us begin by unambiguously defining what deterministic sequence is being suggested by providing a (non-random) real sequence $m_{\mathsf{deter},i}$, and for each $k \geq 0$ a sequence of (non-random) measures, $V_{\mathsf{deter},k}^{(i)}$ where the intended interpretation is that m_i fluctuates around $m_{\mathsf{deter},i}$, and $V_k^{(i)}$ fluctuates around $V_{\mathsf{deter},k}^{(i)}$.

Notation. Recall that an expression "$g \bullet \xi$" where $g(x)$ is a function and ξ is a measure denotes the measure $A \mapsto \int_A g(x)\, d\xi(x)$.

Given an initial value $m_{\mathsf{deter},0}$ and initial measures $V^{(0)}_{\mathsf{deter},0}$, $V^{(0)}_{\mathsf{deter},1}$, $V^{(0)}_{\mathsf{deter},\geq 2}$, we inductively define for $i \geq 0$ and $k \geq 0$

$$m_{\mathsf{deter},i+1} = m_{\mathsf{deter},i} p_{\mathsf{deter},i} \tag{5.3a}$$

$$V^{(i+1)}_{\mathsf{deter},k} = 1_{k=0} V^{(i)}_{\mathsf{deter},\leq 1} + \varphi_k(\nu_i, p_{\mathsf{deter},i}) \bullet V^{(i)}_{\mathsf{deter},\geq 2} \tag{5.3b}$$

where

$$p_{\mathsf{deter},i} = e^{-\frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}}} \tag{5.4}$$

$$b_{\mathsf{deter},i} = V^{(i)}_{\mathsf{deter},1}([0,1]) \tag{5.5}$$

$$\nu_i(x) = \frac{a c_{\mathrm{den}} m_{\mathsf{deter},i}}{x m_{\mathsf{deter},0}} \tag{5.6}$$

$$V^{(i)}_{\mathsf{deter},\geq 2} = V^{(0)}_{\mathsf{deter},\geq 0} - V^{(i)}_{\mathsf{deter},1} - V^{(i)}_{\mathsf{deter},2}. \tag{5.7}$$

The definition we are taking is that the terms of the sequences $m_{\mathsf{deter},i}$ and $V^{(i)}_{\mathsf{deter},k}$ are functions $m_{\mathsf{deter},i} : \mathcal{S}_0 \longrightarrow \mathbb{R}^+$ and $V^{(i)}_{\mathsf{deter},k} : \mathcal{S}_0 \longrightarrow \mathcal{M}$ of the initial conditions, where $\mathcal{S}_0 = \mathbb{R}^+ \times \mathcal{M}^3$ and $\mathcal{M}$ is the set of finite signed measures on $[0,1]$. We also generalize these definitions by defining augmented functions $m_{\mathsf{deter},i} : \mathcal{S}_0 \times \mathcal{S}^i \longrightarrow \mathbb{R}^+$ and $V^{(i)}_{\mathsf{deter},k} : \mathcal{S}_0 \times \mathcal{S}^i \longrightarrow \mathcal{M}$ where $\mathcal{S} = \mathbb{R} \times \mathcal{M}^2$. The $3i$ additional parameters — denoted $u_{\mathsf{deter},i}$, $\xi^{(i)}_{\mathsf{deter},0}$, and $\xi^{(i)}_{\mathsf{deter},1}$ — represent linear offsets at each step, that is

$$m_{\mathsf{deter},i+1} = m_{\mathsf{deter},i} p_{\mathsf{deter},i} + u_{\mathsf{deter},i} \tag{5.8a}$$

$$V^{(i+1)}_{\mathsf{deter},k} = 1_{k=0} V^{(i)}_{\mathsf{deter},\leq 1} + \varphi_k(\nu_i, p_{\mathsf{deter},i}) \bullet V^{(i)}_{\mathsf{deter},\geq 2} + \xi^{(i)}_{\mathsf{deter},k}. \tag{5.8b}$$

The explicit purpose of the generalized definitions is that if we were to evaluate the functions $m_{\mathsf{deter},i}$, $V^{(i)}_{\mathsf{deter},0}$, and $V^{(i)}_{\mathsf{deter},1}$ with the corresponding random variable in each parameter — meaning take the evaluations $m_{\mathsf{deter},0} \mapsto m_0$, $V^{(0)}_{\mathsf{deter},k} \mapsto V^{(0)}_k$, $u_{\mathsf{deter},j} \mapsto u_j$, $\xi^{(j)}_{\mathsf{deter},k} \mapsto \xi^{(j)}_k$ — then $m_{\mathsf{deter},i} = m_i$, $V^{(i)}_{\mathsf{deter},0} = V^{(i)}_0$, and $V^{(i)}_{\mathsf{deter},1} = V^{(i)}_1$ surely.

A certain perspective is useful for motivating what comes next. Suppose we are given two random variables X and Y related by a function, $Y = f(X)$. Knowing a Lipschitz constant for f allows us to bound the deviation in Y under any event which stipulates a maximum deviation of X. In our context, Y is any one of the three random variables m_i, $V^{(i)}_0$, or $V^{(i)}_1$ and the analogue of X is the collection of variables m_0, $V^{(0)}_k$, u_j, and $\xi^{(j)}_k$. We have laid the groundwork by describing the events which stipulate a maximum deviation of X (Theorems 5.3, 5.4) and specifying the (multivariate) functions relating X and Y (the functions $m_{\mathsf{deter},i}$ and $V^{(i)}_{\mathsf{deter},k}$).

Theorem 5.5. For any $\eta > 0$ there exists $L > 0$ such that for all $i \geq 0$ the functions $m_{\mathsf{deter},i}$, $V^{(i)}_{\mathsf{deter},0}$, and $V^{(i)}_{\mathsf{deter},1}$ of $3i + 4$ variables have Lipschitz constant L^i in each parameter. This is subject to the functions being restricted to the domain $D_{i-1}(\eta)$,

defined as follows:

$$E_i(\eta) = \left\{ \frac{V^{(i)}_{\mathsf{deter},\geq 2}([0,1])}{m_{\mathsf{deter},i}} \leq \eta, \frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} \leq \eta \right\} \tag{5.9}$$

$$D_i(\eta) = \bigcap_{1 \leq j \leq i} E_j(\eta)\,. \tag{5.10}$$

Notes. The notation in (5.9) defines a subset of $\mathcal{S}_0 \times \mathcal{S}^{\mathbb{N}}$ in an abbreviated fashion. For example, by the set $\left\{ \frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} \leq \eta \right\}$ we mean

$$\left\{ (s_0, s_1, s_2, \dots) \in \mathcal{S}_0 \times \mathcal{S}^{\mathbb{N}} : \frac{b_{\mathsf{deter},i}\,(s_0, s_1, s_2, \dots)}{m_{\mathsf{deter},i}\,(s_0, s_1, s_2, \dots)} \leq \eta \right\},$$

where $s_0 = \left(m_{\mathsf{deter},0}, V^{(0)}_{\mathsf{deter},0}, V^{(0)}_{\mathsf{deter},1}, V^{(0)}_{\mathsf{deter},\geq 2} \right)$ and $s_j = \left(u_{\mathsf{deter},j}, \xi^{(j)}_{\mathsf{deter},0}, \xi^{(j)}_{\mathsf{deter},1} \right)$ for $j \geq 1$. Then, $D_i(\eta) = D \times \mathcal{S}^{\infty}$ where D is a nontrivial subset of $\mathcal{S}_0 \times \mathcal{S}^i$. We defer introducing the notion of Lipschitz for measure valued arguments until the proof.

The following Lemmas will be referred to in the proof of Theorem 5.5.

Lemma 5.6. Let $f(x,y) = xe^{-y/x}$. Then $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are bounded for $x, y \geq 0$.

Proof. Let $z = y/x$. The theorem follows from using $\frac{\partial f}{\partial x} = e^{-y/x} - \frac{y}{x} e^{-y/x} = (1-z)\,e^{-z}$ and $\frac{\partial f}{\partial y} = -e^{-y/x} = -e^{-z}$, both of which are easily seen to be bounded functions of $z \geq 0$. □

Lemma 5.7. The functions $\lambda \frac{\partial \varphi_0}{\partial \lambda}$, $\lambda \frac{\partial \varphi_1}{\partial \lambda}$, $p \frac{\partial \varphi_0}{\partial p}$, $p \frac{\partial \varphi_1}{\partial p}$ are bounded for $\lambda \geq 0$ and $0 \leq p \leq 1$.

Proof. First, straightforward calculations give

$$\lambda\frac{\partial\varphi_0}{\partial\lambda} = \frac{\lambda p\left(e^{-\lambda p}-e^{-\lambda}\right)+\lambda^2 p\left(e^{-\lambda}-pe^{-\lambda p}\right)}{1-(1+\lambda)e^{-\lambda}} + \frac{\lambda^2 e^{-\lambda}\left(\lambda p\left(e^{-\lambda p}-e^{-\lambda}\right)\right)}{\left(1-(1+\lambda)e^{-\lambda}\right)^2}$$

$$\lambda\frac{\partial\varphi_0}{\partial\lambda} = \frac{-\lambda p\left(e^{-\lambda p}-e^{-\lambda}\right)+\lambda^2(1-p)e^{-\lambda}}{1-(1+\lambda)e^{-\lambda}} - \frac{\lambda^2 e^{-\lambda}\left(e^{-\lambda p}-(1+\lambda(1-p))e^{-\lambda}\right)}{\left(1-(1+\lambda)e^{-\lambda}\right)^2}$$

$$p\frac{\partial\varphi_1}{\partial p} = \frac{\lambda p\left(e^{-\lambda p}-e^{-\lambda}\right)-\lambda^2 p^2 e^{-\lambda p}}{1-(1+\lambda)e^{-\lambda}}$$

$$p\frac{\partial\varphi_0}{\partial p} = \frac{-\lambda p\left(e^{-\lambda p}-e^{-\lambda}\right)}{1-(1+\lambda)e^{-\lambda}}.$$

Using the following Taylor expansions in λ,

$$\lambda p\left(e^{-\lambda p}-e^{-\lambda}\right) = \lambda^2 p(1-p) - \lambda^3\frac{1}{2}p\left(1-p^2\right) + \lambda^4\frac{1}{6}p\left(1-p^3\right)+\cdots$$

$$\lambda^2 p\left(e^{-\lambda}-pe^{-\lambda p}\right) = \lambda^2 p(1-p) - \lambda^3 p\left(1-p^2\right) + \lambda^4\frac{1}{2}p\left(1-p^3\right)+\cdots$$

$$e^{-\lambda p}-(1+\lambda(1-p))e^{-\lambda} = \lambda^2\frac{1}{2}(p-1)^2 + \lambda^3\frac{1}{6}(p-1)^2(p+2) + \lambda^4\frac{1}{24}(p-1)^2\left(p^2+2p+3\right)+\cdots$$

$$1-(1+\lambda)e^{-\lambda} = \lambda^2\frac{1}{2} - \lambda^3\frac{1}{3} + \lambda^4\frac{1}{8}\cdots,$$

one gets

$$\left(1-(1+\lambda)e^{-\lambda}\right)^2\lambda\frac{\partial\varphi_0}{\partial\lambda} = \lambda^5\frac{1}{12}p(p-1)(3p-1) - \lambda^6\frac{1}{12}p\left(2p^3-3p+1\right)+\cdots$$

$$\left(1-(1+\lambda)e^{-\lambda}\right)^2\lambda\frac{\partial\varphi_0}{\partial\lambda} = -\lambda^5\frac{1}{12}p(1-p)^2(3p-1) + \lambda^6\frac{1}{24}p(1-p)^2(p+2)+\cdots$$

$$\left(1-(1+\lambda)e^{-\lambda}\right)p\frac{\partial\varphi_1}{\partial p} = -\lambda^2 p(2p-1) + \lambda^3\frac{1}{2}p\left(3p^2-1\right)+\cdots$$

$$\left(1-(1+\lambda)e^{-\lambda}\right)p\frac{\partial\varphi_0}{\partial p} = \lambda^2 p(1-p) - \lambda^3\frac{1}{2}p\left(1-p^2\right)+\cdots.$$

This allows us to compare the leading order of λ in the numerators and denominators of $\lambda\frac{\partial\varphi_0}{\partial\lambda}$, $\lambda\frac{\partial\varphi_1}{\partial\lambda}$, $p\frac{\partial\varphi_0}{\partial p}$, and $p\frac{\partial\varphi_1}{\partial p}$. One deduces that these functions are bounded near $\lambda = 0$. Away from $\lambda = 0$ matters are simpler since the denominators in $\lambda\frac{\partial\varphi_0}{\partial\lambda}$, $\lambda\frac{\partial\varphi_1}{\partial\lambda}$, $p\frac{\partial\varphi_0}{\partial p}$, $p\frac{\partial\varphi_1}{\partial p}$ are bounded away from zero. And so the lemma follows from the more basic observation that $\lambda^k e^{-\lambda}$ and $\lambda^k e^{-\lambda p}$ are bounded. □

Lemma 5.8. Let ξ be a finite signed measure on $[0, 1]$.

(i) $\sup_f \int_0^1 f\, d\xi = \sup_t \int_t^1 d\xi$ where f ranges over monotonically increasing functions satisfying $0 \le f \le 1$.

(ii) $\inf_f \int_0^1 f\, d\xi = \inf_t \int_t^1 d\xi$ where f ranges over monotonically increasing functions satisfying $0 \le f \le 1$.

(iii) Let M be an upper bound for the absolute value of both quantities in (i) and (ii). Let $\varphi : [0, 1] \longrightarrow \mathbb{R}$ with $\varphi(0) = a$ be a function of bounded variation C. Then

$$\left| \int_0^1 \varphi\, d\xi \right| \le |a| \left| \int_0^1 d\xi \right| + CM.$$

Proof. We will prove (i) and (iii), as the proof for (ii) is similar to (i).

Proof of (i): It suffices to prove the claim for step functions f. The statement in (i) asserts that a maximizer of $\sup_f \int_0^1 f\, d\xi$ may be taken to be either $f = 0$ or else

a function f with 2 steps. Given a function f with $n > 2$ steps, let $I_1, I_2 \subseteq [0,1]$ denote the smallest and second smallest steps of nonzero height. Define a new step function g with $n-1$ steps by adjusting the value of f on the interval I_1: if $\xi(I_1) > 0$ then define $g(I_1) = f(I_2)$, otherwise define $g(I_1) = 0$. The function g then satisfies $\int_0^1 f\, d\xi \leq \int_0^1 g\, d\xi$ and has fewer steps.

Proof of (iii): The Jordan decomposition for functions of bounded variation permits us to write φ as the difference of two increasing functions. In particular, we take $\varphi = \varphi(0) + C_1 f_1 - C_2 f_2$ where $C_1, C_2 \geq 0$ and f_1 and f_2 are positive, increasing functions satisfying $0 \leq f_i \leq 1$. Then $C_1 + C_2 \leq C$, the total variation of φ. Using this expression for φ,

$$\begin{aligned}\int_0^1 \varphi\, d\xi &= \int_0^1 a\, d\xi + C_1 \int_0^1 f_1\, d\xi - C_2 \int_0^1 f_2\, d\xi \\ &\leq \int_0^1 a\, d\xi + C_1 \sup{}_t \int_t^1 d\xi - C_2 \inf{}_t \int_t^1 d\xi.\end{aligned}$$

This shows

$$\int_0^1 \varphi\, d\xi \leq |a| \left| \int_0^1 d\xi \right| + (C_1 + C_2)\, M,$$

and the complementary lower bound follows from a similar argument applied to $-\varphi$.

□

Discussion. The proof of Theorem 5.5 is based around an argument for multivariable functions which is a generalization of a much more readily phrased argument for single

variable functions: for a one variable function $f(x)$ with Lipschitz constant L then the ith iterate $f^{(i)}(x)$ has Lipschitz constant which grows exponentially as L^i. In the multivariate setting one may have a Lipschitz constant associated to each parameter of the function $f(x_1, x_2, \dots)$. The multivariate generalization of nested composition may be described by a kind of rooted tree: rooted trees where each vertex v is labeled by a n-ary function g_v, with $n = n_v$ depending on v, such that n_v is the number of descendants of v. It is preferable to speak of the functions themselves as the vertices, although strictly speaking the use of labels allows distinct vertices to be labeled by the same function. The tree describes a nested composition through the interpretation that the descendants $h_1, h_2, \dots, h_n$ of a vertex g correspond to composing these n functions into each argument of g. So the root vertex f represents the outermost function and the height of the tree indicates the deepest level of nesting of function arguments inside f. Here we are considering 0-ary functions to be simply values, indicating that the sequence of compositions stops in that argument.

In one variable, the ith iterate $f^{(i)}(x)$ is represented by a chain (a 1-ary tree) with $i + 1$ vertices. Each edge corresponds to a Lipschitz constant L, and the unique path from the root to the leaf of the chain has i edges. The product of the Lipschitz constant for each edge in this path gives L^i and this is the Lipschitz constant for the nested composition. In the multivariate setting, the Lipschitz constant for a

parameter x is given by summing the corresponding product over all paths starting from the root and ending with a leaf labeled by x. For a tree of height i there are at most exponentially many (in i) such paths if the tree has bounded degree. And each path contributes at most an exponential factor L^i if all edges in the tree have a common Lipschitz constant L.

Proof of Theorem 5.5. Consider four rooted trees which are rooted by $m_{\mathsf{deter},i}$, $b_{\mathsf{deter},i}$, $V^{(i)}_{\mathsf{deter},0}$, $V^{(i)}_{\mathsf{deter},1}$ for some fixed i. Each tree consists of internal vertices labeled by $m_{\mathsf{deter},j}$, $b_{\mathsf{deter},j}$, $V^{(j)}_{\mathsf{deter},0}$, $V^{(j)}_{\mathsf{deter},1}$ where $1 \leq j \leq i$ and leaves labeled by $m_{\mathsf{deter},0}$, $b_{\mathsf{deter},0}$, $V^{(0)}_{\mathsf{deter},0}$, $V^{(0)}_{\mathsf{deter},1}$, $V^{(0)}_{\mathsf{deter},\geq 2}$, $u_{\mathsf{deter},i}$, $\xi^{(j)}_{\mathsf{deter},0}$ $\xi^{(j)}_{\mathsf{deter},1}$ where $0 \leq j \leq i-1$. The descendant relations are such that the trees represent the nested composition recursively specified by equations (5.8).

We need to address how the discussion regarding Lipschitz constants and nested composition of multivariate functions is compatible with measure valued functions and arguments. For simplicity, we will specialize the answer to just the particular trees involved in the present theorem.

Consider the measure valued function $T_f : \mathcal{M} \longrightarrow \mathcal{M}$ defined by $\xi \mapsto f \bullet \xi$. Assign the norm $\|\cdot\| : \mathcal{M} \longrightarrow \mathbb{R}$ defined as $\xi \mapsto \left|\sup_t \int_t^1 d\xi\right| \vee \left|\inf_t \int_t^1 d\xi\right|$ to the space $\mathcal{M}$

of signed measures. Should the inequality $\left|\int_0^1 \varphi\, d\xi\right| \leq L\,\|\xi\|$ hold, we will interpret[4] it as saying that T_φ has Lipschitz constant L on the one dimensional domain of signed measures $\{c\xi : c \in \mathbb{R}\}$. Every edge in our rooted trees representing composition of a measure parameter in fact represents either an application of T_φ for some function φ or it represents an application of $\xi \mapsto \int_0^1 \varphi\, d\xi$ (these are the edges which directly lead to $b_{\mathsf{deter},i}$).

The strategy then is to show each edge in our rooted tree has Lipschitz constant L in this extended sense.

Lemma 5.6 provides the Lipschitz condition for $m_{\mathsf{deter},i+1}$ as a function of $m_{\mathsf{deter},i}$ and $b_{\mathsf{deter},i}$.

Define functions $\lambda(m_i, b_i) = \frac{c_{\text{den}} m_i}{xm}$ and $p(m_i, b_i) = e^{-\frac{b_i}{m_i}}$ for the purpose of viewing $\varphi = \varphi_k(\lambda, p)$ as a function of two variables m_i and b_i. Using Lemma 5.7 and the chain rule, we get $\frac{\partial \varphi_k}{\partial m_i} = O\left(\frac{1}{m_i} + \frac{b_i}{m_i^2}\right)$ and $\frac{\partial \varphi_k}{\partial b_i} = O\left(\frac{1}{m_i}\right)$. Using the domain assumption, this provides the Lipschitz condition for $V^{(i+1)}_{\mathsf{deter},0}$ and $V^{(i+1)}_{\mathsf{deter},1}$ as functions of $m_{\mathsf{deter},i}$ and $b_{\mathsf{deter},i}$.

[4]This is definition is not exactly an instance of Lipschitz, since we are using a seminorm $\mu \mapsto \left|\int_0^1 d\mu\right|$ for the range.

Finally, as stated earlier, we must consider the family of transformations $\{T_\varphi\}_{\varphi\in\mathcal{C}}$ where φ varies over some class $\mathcal{C}$ of functions. Our goal is to show this family of transformations has a uniform Lipschitz constant. However the nature of finding a uniform Lipschitz constant requires constraints on the class of functions $\mathcal{C}$ depending on the domain of measures. To illustrate, note the total variation $\int d\,|\xi|$ is representable as $|T_\varphi \xi|$ for some function φ which varies as $+1$ or -1 according to the sign of the measure ξ. For our signed measures $\xi_k^{(i)}$ from Theorem 5.3, the total variation is $O(n)$ while the norm is roughly of order $\sqrt{n}$. So a large enough family, e.g. one which includes all bounded functions φ, will have Lipschitz constant of order $\sqrt{n}$ — an inadequate result.

Consider then the particular chains in our rooted trees where every vertex of the chain is labeled by a measure. Such a chain of height k necessarily represents an expression of the form $\varphi^{(1)} \ldots \varphi^{(k)} \bullet \xi$ for some sequence of functions $\varphi^{(i)}$, and if the chain is not maximal then the parent of the chain is some vertex $b_{\mathsf{deter},j}$ appearing as an expression $\int_{[0,1]} \varphi^{(1)} \ldots \varphi^{(k)}\, d\xi$.

We complete the proof by showing, for every such chain, this integral is at most $L^k \|\xi\|$. In our situation, the class of functions $\mathcal{C}$ we must serve consists of k-fold products of functions $\varphi_0(\nu_i(x), p_{\mathsf{deter},i})$ and $\varphi_1(\nu_i(x), p_{\mathsf{deter},i})$. Essentially, the special property in our favor is that the total variation of the functions in this class is well

controlled, and along the lines of Lemma 5.8 this will produce a Lipschitz bound.

Working with the k-fold products directly instead of the individual transformations T_φ makes the presentation easier, but the high level interpretation of what we are showing can be abstracted as saying the family $\{T_\varphi\}_{\varphi\in\mathcal{C}}$ has uniform Lipschitz constant L over a particular domain of measures, and that $T_\varphi(\xi)$ does not leave this domain.

We note that Lemma 5.8 can be stated in greater generality. First, the total variation of a function φ does not change if the domain is reparametrized by precomposing φ with a smooth invertible function. And so the domain $[0,1]$ assumption may be replaced by a domain $[0,\infty)$, and the endpoint assumption $\varphi(0)=0$ may be replaced by $\lim_{x\to\infty}\varphi(x)=0$.

For functions f, g of bounded variation both satisfying the endpoint condition $a=0$, one has $\mathrm{TV}(fg)\leq \mathrm{TV}(f)\,\mathrm{TV}(g)$. Therefore a family of k-fold products of functions will have total variation uniformly bounded by L^k provided L bounds the total variation of the individual functions.

Each function $\varphi(x) = \varphi_k(\nu_i(x)\,, p_{\mathsf{deter},i})$ is a reparametrization of $\varphi(\lambda) = \varphi_k(\lambda, p_{\mathsf{deter},i})$. The functions $f(\lambda)=\varphi_k(\lambda,p)$ where $0\leq p\leq 1$ is a constant have

uniformly bounded total variation: $\mathrm{TV}(f) \leq \int_0^\infty \left|\frac{\partial f}{\partial \lambda}\right| d\lambda$ and it follows from Lemma 5.7 that $\left|\frac{\partial f}{\partial \lambda}\right|$ has tails of order $O\left(e^{-\lambda}\right)$. $\square$

Evaluate the functions $m_{\mathsf{deter},i}$, $V^{(i)}_{\mathsf{deter},0}$, and $V^{(i)}_{\mathsf{deter},1}$ with the same initial conditions as the removal chain and with the corresponding random variables for each offset parameter: $m_{\mathsf{deter},0} \mapsto m_0$, $V^{(0)}_{\mathsf{deter},k} \mapsto V^{(0)}_k$, $u_{\mathsf{deter},j} \mapsto u_j$, $\xi^{(j)}_{\mathsf{deter},k} \mapsto \xi^{(j)}_k$. We may also consider the set $D_i(\eta)$ a random event in this way, which we shall denote as $D_i^{\mathsf{event}}(\eta)$ to avoid abuse of notation. More explicitly, $D_i^{\mathsf{event}}(\eta)$ is defined as the inverse image of $D_i(\eta)$ by the map $\Omega \longrightarrow \mathcal{S}_0 \times \mathcal{S}^\infty$ which is defined componentwise by the functions (i.e. random variables) $\Omega \longrightarrow R$ and $\Omega \longrightarrow \mathcal{M}$ given by m_0, $V^{(0)}_k$, u_j, and $\xi^{(j)}_k$. Theorem 5.5 says that on the event $D^{\mathsf{event}}_{i-1}(\eta)$ the deviation of the removal chain from the deterministic trajectory at time i is at most $L^i K_{i-1}$, where

$$
\begin{aligned}
k_{\mathrm{I},j} &= |u_j| \\
k_{\mathrm{II},j} &= \sup{}_{t\in[0,1]} \left|\xi^{(j)}_0([t,1])\right| \\
k_{\mathrm{III},j} &= \sup{}_{t\in[0,1]} \left|\xi^{(j)}_1([t,1])\right| \\
k_j &= k_{\mathrm{I},j} \vee k_{\mathrm{II},j} \vee k_{\mathrm{III},j} \\
K_i &= \sup{}_{j\leq i} k_j.
\end{aligned}
$$

Theorem 5.9. $K_i = O_p\left(\sqrt{n}\log n\right)$ as $n \to \infty$, uniformly in $i \geq 0$.

Proof. The exponential tail bounds from Theorems 5.4 and part (i) of 5.3 on the random variables k_j tell us that these variables are individually of size $O\left(\sqrt{n}\log n\right)$ with exceptional probability that is $o\left(\frac{1}{n}\right)$. Since there are only $O\left(n\right)$ many variables involved, a union bound on the exceptional events gives an exceptional probability of $o\left(1\right)$. □

5.3 The Deterministic System's Trajectory

The goal of this section is to analyze the limit of the deterministic process. We specialize our initial conditions to $m_{\mathsf{deter},0} = c_{\text{den}}n$ and $V^{(0)}_{\mathsf{deter},k} = n\mu_k$ where μ_k is the measure from part (iii) of Theorem 5.1, and set $u_{\mathsf{deter},i} = 0$, $\xi^{(i)}_{\mathsf{deter},k} = 0$. For this section, n is always fixed and limits are as $i \to \infty$. Let $\beta_i = \frac{m_{\mathsf{deter},i}}{m}$, which denotes the proportion of hyperedges still remaining after step i.

Lemma 5.10. $V^{(i)}_{\mathsf{deter},\geq 2}$ has Lebesgue density

$$\frac{dV^{(i)}_{\mathsf{deter},\geq 2}}{d\mathcal{L}}(x) = P\left(\,\text{Poiss}(\nu_i(x)) \geq 2\,\right).$$

Proof. Proof by induction. When $i = 0$ the statement comes from the definition of μ_k. Assume the result holds for some $i > 0$. By equation (5.3b), one gets $V^{(i+1)}_{\mathsf{deter},\geq 2} =$

$\varphi_{\geq 2}(\nu_i, \beta_i) \bullet V^{(i)}_{\mathsf{deter},\geq 2}$. Using the induction hypothesis and the definition of φ,

$$\begin{aligned}
\frac{dV^{(i+1)}_{\mathsf{deter},\geq 2}}{d\mathcal{L}}(x) &= \varphi_{\geq 2}(\nu_i(x)\,, p_{\mathsf{deter},i})\, P\,(\,\mathrm{Poiss}(\nu_i(x)) \geq 2\,) \\
&= \frac{1 - (1 + \nu_i(x)\, p_{\mathsf{deter},i})\, e^{\,-\nu_i(x) p_{\mathsf{deter},i}}}{1 - (1 + \nu_i(x))\, e^{\,-\nu_i(x)}} P\,(\,\mathrm{Poiss}(\nu_i(x)) \geq 2\,) \\
&= 1 - (1 + \nu_i(x)\, p_{\mathsf{deter},i})\, e^{\,-\nu_i(x) p_{\mathsf{deter},i}} \\
&= P\,(\,\mathrm{Poiss}(\nu_i(x)\, p_{\mathsf{deter},i}) \geq 2\,)\,. \qquad \square
\end{aligned}$$

Lemma 5.11 (Recurrence for β_i)**.** For all $i \geq 0$,

$$\log \beta_{i+1} = -a \int_{a c_{\mathrm{den}} \beta_i}^{\infty} \frac{e^{\,-t}}{t}\, dt.$$

Proof. We begin by organizing some equations:

$$b_{\mathsf{deter},0} = n \int_0^1 \nu_0(x)\, e^{\,-\nu_0(x)}\, dx \tag{5.11}$$

$$b_{\mathsf{deter},i} = n \int_0^1 \nu_0(x)\, \beta_i \left(e^{\,-\nu_0(x)\beta_i} - e^{\,-\nu_0(x)\beta_{i-1}} \right) dx \quad \text{for } i \geq 1 \tag{5.12}$$

$$\log \beta_{i+1} = -\left(\frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} + \cdots + \frac{b_{\mathsf{deter},0}}{m_{\mathsf{deter},0}} \right). \tag{5.13}$$

Equations (5.11) and (5.12) come the definition (5.5) for $b_{\mathsf{deter},i}$, Lemma 5.10, and use of the identities $\nu_i = \nu_0 \beta_i$ and $\nu_i p_{\mathsf{deter},i} = \nu_0 \beta_{i+1}$. Equation (5.13) comes from rewriting definition (5.3a) for $m_{\mathsf{deter},i}$ as $\log \beta_{i+1} = -\frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} + \log \beta_i$ and iteration. When $b_{\mathsf{deter},i}$ is divided by $m_{\mathsf{deter},i}$, the integrals in (5.11) and (5.12) have an overall

factor of $\frac{n\beta_i}{m_{\mathsf{deter},i}} = \frac{1}{c_{\mathrm{den}}}$, leading to a telescoping sum in (5.13) which simplifies to

$$\log \beta_{i+1} = -\frac{1}{c_{\mathrm{den}}} \int_0^1 \nu_0(x)\, e^{-\nu_0(x)\beta_i}\, dx.$$

We perform the substitution $t = \nu_0(x)\,\beta_i = \frac{ac_{\mathrm{den}}\beta_i}{x}$ in this last integral (giving $dt = -\frac{t\,dx}{x} = -\frac{t^2\,dx}{ac_{\mathrm{den}}\beta_i}$) to complete the proof. □

From the defining equation (5.3a) for $m_{\mathsf{deter},i}$, one has $\beta_{i+1} = \beta_i e^{-b_{\mathsf{deter},i}/m_{\mathsf{deter},i}}$. Evidently then, β_i is a decreasing, non-negative sequence and therefore converges to some $\beta_\infty \geq 0$ as $i \to \infty$. So either $\beta_\infty = 0$ or else β_∞ is a solution to the equation

$$\log \beta = -a \int_{ac_{\mathrm{den}}\beta}^{\infty} \frac{e^{-t}}{t}\, dt. \tag{5.14}$$

If this equation has any solutions, then they are all strictly less than $\beta_0 = 1$. Let β' denote a credible candidate for β_∞: the largest solution if there is any, and zero otherwise. We may rewrite the recurrence of Lemma 5.11 as the fix point iteration of $f(\beta)$ where

$$f(\beta) = \exp\left(-a \int_{ac_{\mathrm{den}}\beta}^{\infty} \frac{e^{-t}}{t}\, dt\right), \tag{5.15}$$

and of course the fixed points of f are the solutions to (5.14). Since $f(1) < 1$, the portion of the graph $y = f(\beta)$ with $\beta' < \beta \leq 1$ lies below the line $y = \beta$, from which we may conclude β_i converges to β' — that is, $\beta_\infty = \beta'$.

The proof of the next theorem requires the following simple lemma.

Lemma 5.12 (Exponential Integral)**.**

$$\int_x^\infty \frac{e^{-t}}{t}\,dt = -\gamma - \log x + h(x)$$

where $h : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{\geq 0}$ is a bijection with $h(0) = 0$.

Proof. Using $\log x = \int_1^x \frac{1}{t}\,dt$ we have $h(x) = \gamma + \int_1^\infty \frac{e^{-t}}{t}\,dt + \int_1^x \frac{1-e^{-t}}{t}\,dt$. From this and integral tables, $h(0) = 0$. Differentiating, $h'(x) = \frac{1-e^{-x}}{x} > 0$, so that h is monotonic. Since $h(x) = \gamma + o(1) + \log x$ as $x \to \infty$, the range of h includes all non-negative reals. □

For purposes that will become clear during the following proof, define $\beta_{\mathsf{min}} = \frac{\log a}{a c_{\text{den}}}$.

Theorem 5.13 (Limiting Proportion of Hyperedges)**.**

1. Case $a < 1$: $\beta_\infty > 0$, with $\beta_\infty \downarrow 0$ as $c_{\text{den}} \downarrow 0$.

2. Case $a = 1$:

   If $c_{\text{den}} \leq e^{-\gamma}$ then $\beta_\infty = 0$.

   If $c_{\text{den}} > e^{-\gamma}$ then $\beta_\infty > 0$ with $\beta_\infty \downarrow 0$ as $c_{\text{den}} \downarrow e^{-\gamma}$.

3. Case $a > 1$:

If $c_{\text{den}} < c_*$ then $\beta_\infty = 0$.

If $c_{\text{den}} = c_*$ then $\beta_\infty = \beta_{\mathsf{min}} = \frac{\log a}{ac_{\text{den}}}$.

If $c_{\text{den}} > c_*$ then $\beta_\infty > \frac{\log a}{ac_*} > \beta_{\mathsf{min}}$ with $\beta_\infty \downarrow \frac{\log a}{ac_*}$ as $c_{\text{den}} \downarrow c_*$.

Proof. Rewrite (5.14) by setting $\theta = ac_{\text{den}}\beta$ to get the equivalent equation

$$\log\theta + a\int_\theta^\infty \frac{e^{-t}}{t}\,dt = \log\left(ac_{\text{den}}\right). \tag{5.16}$$

As c_{den} varies, the right hand side expression, $\log ac_{\text{den}}$, varies over $\mathbb{R}$. The range of the left hand side expression, $H(\theta)$, determines the existence or non-existence of solutions for θ, and hence solutions for β (take note, however, the solutions for θ and β are not linearly related as functions of c_{den}). Observe that as $\theta \to \infty$, $H(\theta) = \log\theta + O\left(1\right)$. Using Lemma 5.12 to expand the integral near 0, we have

$$H(\theta) = (1-a)\log\theta - a\gamma + ah(\theta)\,,$$

revealing that near $\theta = 0$, $H(\theta) = (1-a)\log\theta + O\left(1\right)$.

Considering $a < 1$, this behavior at 0 and ∞ implies that for every $c_{\text{den}} > 0$ there exists a solution.

Considering $a > 1$, this behavior at 0 and ∞ implies $H(\theta)$ has an absolute minimum, and so has no solution if $\log\left(ac_{\text{den}}\right)$ is less than this minimum. Differentiating,

$H'(\theta)$ is zero at $\theta_* \coloneqq \log a$. Substituting $\theta \mapsto \theta_*$ into (5.16) gives the critical value of $c_{\text{den}} = c_*$ (where c_* is as defined in Theorem 2.1), below which $\beta_\infty = 0$. This establishes $\beta_\infty = \beta_{\mathsf{min}}$ when $c_{\text{den}} = c_*$. For $c_{\text{den}} > c_*$, the largest solution θ is to the right of the minimum θ_*, which is equivalent to $\beta_\infty > \beta_{\mathsf{min}}$.

Finally considering $a = 1$, equation (5.16) simplifies to

$$h(\theta) = \gamma + \log\left(c_{\text{den}}\right).$$

By the properties of h from Lemma 5.12, this has a unique positive solution precisely when $c_{\text{den}} > e^{-\gamma}$.

Return now to any $a > 0$. It only remains to prove monotonicity of β_∞ as a function of c_{den}. Differentiating equation (5.14) by c_{den},

$$\frac{1}{\beta}\frac{d\beta}{dc_{\text{den}}} = ae^{-ac_{\text{den}}\beta}\left(\frac{1}{\beta}\frac{d\beta}{dc_{\text{den}}} + \frac{1}{c_{\text{den}}}\right).$$

So the equation provides $\frac{d\beta_\infty}{dc_{\text{den}}} > 0$ whenever $ae^{-ac_{\text{den}}\beta_\infty} < 1$. It is useful for now and later to notice $ae^{-ac_{\text{den}}\beta} < 1$ is equivalent to $\beta > \beta_{\mathsf{min}} = \frac{\log a}{ac_{\text{den}}}$, and $\beta_\infty > \beta_{\mathsf{min}}$ is equivalent to $\beta_\infty > \beta_*$ where

$$\beta_* = \inf\left\{\beta_\infty(c_{\text{den}}) : \beta_\infty(c_{\text{den}}) > 0,\, c_{\text{den}} > 0\right\}.$$

Note though that β_{min} depends on both a and c_{den}, while β_* depends only a. These

equivalences prove $\beta_\infty(c_{\text{den}})$ is strictly increasing except for the established intervals on which it is constantly zero. $\square$

We say a sequence $x_i \to x_\infty$ *converges exponentially* if there exists $C > 0$ so that $|x_i - x_\infty| = O\left(e^{-Ci}\right)$, and *converges super-exponentially* if $-\log |x_i - x_\infty| = \omega(i)$.

Theorem 5.14 (Rate of Convergence for β_i)**.**

1. Case $a < 1$: $\frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} \to 0$ and β_i converges exponentially.

2. Case $a = 1$:

   If $c_{\text{den}} \leq e^{-\gamma}$ then $\frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} \to -\gamma - \log c_{\text{den}}$.

   If $c_{\text{den}} > e^{-\gamma}$ then $\frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} \to 0$.

   If $c_{\text{den}} \neq e^{-\gamma}$ then β_i converges exponentially.

3. Case $a > 1$:

   If $c_{\text{den}} < c_*$ then $\frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} \to \infty$ and β_i converges super-exponentially.

   If $c_{\text{den}} \geq c_*$ then $\frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} \to 0$.

   If $c_{\text{den}} > c_*$ then β_i converges exponentially.

Proof. Taking the derivative of $f(\beta)$ from (5.15),

$$f'(\beta) = \frac{f(\beta)}{\beta} a e^{-ac_{\text{den}}\beta}.$$

We recall from the fix point discussion preceding (5.15) that $f(\beta) \leq \beta$ when $\beta \geq \beta_\infty$. Also note that $ae^{-ac_{\text{den}}\beta} < 1$ when $\beta > \beta_{\mathsf{min}}$. So under the condition $\beta_\infty > \beta_{\mathsf{min}}$, one has $f'(\beta) < 1$ for all $\beta \geq \beta_\infty$. This says f is a contraction mapping and β_i converges exponentially. The condition $\beta_\infty > \beta_{\mathsf{min}}$ covers the cases: $a < 1$; $a = 1$, $c_{\text{den}} > e^{-\gamma}$; and $a > 1$, $c_{\text{den}} > c_*$.

Since $-\log \frac{\beta_{i+1}}{\beta_i} = \frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}}$, either expression converges to 0 when $\beta_\infty > 0$. This accounts for all limits $\frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} \to 0$ in the statement of the theorem.

Now, combining lemmas 5.11 and 5.12 we get

$$\log \frac{\beta_{i+1}}{\beta_i} = a\left(\gamma + \log\left(ac_{\text{den}}\right) - h(ac_{\text{den}}\beta_i)\right) + (a-1)\log\beta_i, \tag{5.17}$$

which when $\beta_\infty = 0$ says that as $i \to \infty$,

$$\log \frac{\beta_{i+1}}{\beta_i} = a\left(\gamma + \log\left(ac_{\text{den}}\right)\right) + (a-1)\log\beta_i + o\left(1\right).$$

Considering the case $a = 1$, $\beta_\infty = 0$ we conclude $-\log \frac{\beta_{i+1}}{\beta_i} = \frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} \to \gamma + \log c_{\text{den}}$. If additionally $c_{\text{den}} < e^{-\gamma}$, then $\gamma + \log c_{\text{den}} < 0$ and β_i converges exponentially. Considering the case $a > 1$, $\beta_\infty = 0$ we conclude $-\log \frac{\beta_{i+1}}{\beta_i} = \frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} \to \infty$ and β_i converges super-exponentially. □

Lemma 5.15.

$$\frac{V^{(i)}_{\mathsf{deter},\geq 2}([0,1])}{m_{\mathsf{deter},i}} = \frac{1-e^{-ac_{\mathrm{den}}\beta_i}}{c_{\mathrm{den}}\beta_i}$$
$$\frac{V^{(i)}_{\mathsf{deter},2}([0,1])}{m_{\mathsf{deter},i}} = \frac{1}{2}ae^{-ac_{\mathrm{den}}\beta_i}.$$

Proof. From the definition,

$$V^{(i)}_{\mathsf{deter},\geq 2}([0,1]) = n\int_0^1 1-(1+\nu_0(x)\,\beta_i)\,e^{-\nu_0(x)\beta_i}\,dx$$
$$V^{(i)}_{\mathsf{deter},2}([0,1]) = n\int_0^1 \frac{1}{2}\nu_0^2(x)\,\beta_i^2 e^{-\nu_0(x)\beta_i}\,dx.$$

Perform the substitution $t = \nu_0(x)\,\beta_i = \frac{ac_{\mathrm{den}}\beta_i}{x}$ (giving $dt = -\frac{t\,dx}{x} = -\frac{t^2\,dx}{ac_{\mathrm{den}}\beta_i}$) to get

$$V^{(i)}_{\mathsf{deter},\geq 2}([0,1]) = n\,ac_{\mathrm{den}}\beta_i\int_{ac_{\mathrm{den}}\beta_i}^{\infty}\frac{1-(1+t)\,e^{-t}}{t^2}\,dt$$
$$V^{(i)}_{\mathsf{deter},2}([0,1]) = n\,ac_{\mathrm{den}}\beta_i\int_{ac_{\mathrm{den}}\beta_i}^{\infty}\frac{1}{2}e^{-t}\,dt = \frac{n}{2}ac_{\mathrm{den}}\beta_i e^{-ac_{\mathrm{den}}\beta_i}.$$

For the first expression, we can employ the identity

$$\frac{1-(1+t)\,e^{-t}}{t^2} = \int_0^1 se^{-st}\,ds$$

and use Fubini's Theorem to derive the integration formula

$$\int_a^b \frac{1-(1+t)\,e^{-t}}{t^2}\,dt = \int_0^1\int_a^b se^{-st}\,dt\,ds$$
$$= \int_0^1 e^{-sa}-e^{-sb}\,ds = \frac{1}{a}\left(1-e^{-a}\right)-\frac{1}{b}\left(1-e^{-b}\right).$$ □

We will say that *the deterministic system lies in* $D_i(\eta)$ if the sequences $m_{\mathsf{deter},j}$, $V^{(j)}_{\mathsf{deter},0}$, and $V^{(j)}_{\mathsf{deter},1}$ satisfy the bounds of (5.9) for $0 \leq j \leq i$ — formally, if $(s_0, 0, 0, \dots)$ is an element of $D_i(\eta) \subseteq \mathcal{S}_0 \times \mathcal{S}^{\mathbb{N}}$, where $s_0 \in \mathcal{S}_0$ is defined by $s_0 = \left(m_{\mathsf{deter},0}, V^{(0)}_{\mathsf{deter},0}, V^{(0)}_{\mathsf{deter},1}, V^{(0)}_{\mathsf{deter},\geq 2}\right)$ and these later quantities being as defined in the beginning of this chapter.

Corollary 5.16.

(i) If a and c_{den} are such that $\beta_\infty = 0$, then for all $\epsilon > 0$ there exists $\eta > 0$ and $i > 0$ such that for all $n \geq 1$, $m_{\mathsf{deter},i} < \epsilon n$ and the deterministic system lies in $D_{i-1}(\eta)$.

(ii) Unless $a > 1$ and $c_{\text{den}} < c_*$, there exists $\eta > 0$ such that for all $n \geq 1$ and $i \geq 0$ the deterministic system is contained in $D_i(\eta)$.

Proof. An important observation is that $m_{\mathsf{deter},i}$ and $V^{(0)}_{\mathsf{deter},k}$ are homogenous functions of their parameters $m_{\mathsf{deter},0}$, $V^{(0)}_{\mathsf{deter},0}$, $V^{(0)}_{\mathsf{deter},1}$, $V^{(0)}_{\mathsf{deter},\geq 2}$ — if these initial conditions at $i = 0$ are scaled by some α, then the entire sequence for all $i > 0$ is scaled by α. Because our initial conditions are scaled by n, the trajectory of the sequence $\frac{m_{\mathsf{deter},i}}{n} = c_{\text{den}}\beta_i$ is independent of n and likewise the exit time of the set $D_i(\eta)$ is a function of i but not n.

For (i), we may pick i so that $c_{\text{den}}\beta_i < \epsilon$ (since $\beta_\infty = 0$) and choose the smallest set $D_{i-1}(\eta)$ which contains the trajectory of the deterministic system up to time $i-1$.

For (ii), the sequence $\frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}}$ converges by Theorem 5.14. The sequence $\frac{1}{m_{\mathsf{deter},i}} V^{(i)}_{\mathsf{deter},\geq 2}([0,1])$ increases to a limit which is at most a by Lemma 5.15. Take η to be a common bound on both sequences. $\square$

We remark that for all $i \geq 0$ such that the deterministic system is contained in $D_i(\eta)$, the bound $\frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} \leq \eta$ implies that $m_{\mathsf{deter},i+1} \geq e^{-\eta} m_{\mathsf{deter},i}$, bounding the exponential rate decay for $m_{\mathsf{deter},i}$. For any such i and any $0 < \sigma < 1$, it follows that there exists $\epsilon > 0$ such that $m_{\mathsf{deter},i} \geq n^\sigma$ whenever $i \leq \epsilon \log n$.

5.4 The 2-Core and the Limiting State of the Removal Chain

Theorem 5.17.

(i) Let $\epsilon > 0$, and let η and i be as in (i) of Corollary 5.16 for this ϵ. Then as

$n \to \infty$,

$$P(D_i^{\mathsf{event}}(\eta)) = 1 - o(1).$$

(ii) Let η_0 be as in (ii) of Corollary 5.16, and let $\eta > \eta_0 \vee c_{\text{den}}^{-1}$. Then there exists $\epsilon > 0$ so that for $i \leq \epsilon \log n$ and as $n \to \infty$,

$$P(D_i^{\mathsf{event}}(\eta)) = 1 - o(1).$$

Proof. The strategy is to estimate the differences $z_1 = \frac{b_{\mathsf{deter},i}}{m_{\mathsf{deter},i}} - \frac{b_i}{m_i}$ and $z_2 = \frac{V^{(i)}_{\mathsf{deter},\geq 2}([0,1])}{m_{\mathsf{deter},i}} - \frac{V_2^{(i)}([0,1])}{m_i}$. Our hypotheses inform us the deterministic system lies in $D_i(\eta_0)$, so the event that these random differences are sufficiently small will contain the event $D_i^{\mathsf{event}}(\eta)$. The differences z_1 and z_2 can each be represented as deviations of the function $f(x,y) = \frac{y}{x}$, which has partial derivatives $\frac{\partial f}{\partial x} = -\frac{y}{x^2}$ and $\frac{\partial f}{\partial y} = \frac{1}{x}$.

On the event $D_{i-1}^{\mathsf{event}}(\eta)$ and excluding an additional event E_i of exceptional probability $o\left(\frac{1}{n}\right)$ according to Theorem 5.9, the quantities m_i, b_i, and $V_{\geq 2}^{(i)}([0,1])$ deviate at most $L^i K_{i-1}$ from the deterministic quantities. Let $\sigma_1, \sigma_2 \in (0,1)$, to be chosen later. Let $\epsilon_1 > 0$ be as in the remarks following Corollary 5.16 so that $m_{\mathsf{deter},i} > n^{\sigma_1}$. Let $\epsilon_2 > 0$ be sufficiently small so that $L^i = O(n^{\sigma_2})$ for $i \leq \epsilon_2 \log n$. Let $h(n) = \frac{\eta \vee 1}{n^{\sigma_1}}$, which is greater than $\frac{\eta \vee 1}{m_{\mathsf{deter},i}}$ for $i \leq \epsilon_1 \log n$.

For the remaining part of the proof we consider $i \leq (\epsilon_1 \wedge \epsilon_2) \log n$. By bound-

ing the first derivatives of f over $\left\{(x,y) \in \mathbb{R}^2 : x \geq 0,\ y \geq 0,\ \frac{y}{x} \leq \eta\right\}$, one obtains the bound $|z_k| \leq O\left(hL^i K_{i-1}\right)$ holding on the event $D_{i-1}^{\mathsf{event}}(\eta) \cap E_i^c$ of the preceding paragraph. As $O\left(hL^i K_{i-1}\right) = O\left(n^{\frac{1}{2}+\sigma_2-\sigma_1}\log n\right)$, let us pick σ_1 and σ_2 so that this exponent is negative. With this choice and with n sufficiently large, $D_{i-1}^{\mathsf{event}}(\eta) \cap E_i^c$ contains the event that the differences z_1 and z_2 are both small, which in turn contains the event $D_i^{\mathsf{event}}(\eta)$.

In other words, we have shown the symmetric difference $D_{i-1}^{\mathsf{event}}(\eta) \oplus D_i^{\mathsf{event}}(\eta)$ has probability $1 - o\left(\frac{1}{n}\right)$. It follows that $D_i^{\mathsf{event}}(\eta) \oplus D_0^{\mathsf{event}}(\eta)$ has probability $1 - o(1)$. For $\eta > c_{\mathrm{den}}^{-1}$, one has $\eta m > n$, and so $D_0^{\mathsf{event}}(\eta)$ has probability 1. □

Proof of Main Theorem 2.1 for special cases: Let R once again denote $\lim_{i\to\infty} m_i$, the number of hyperedges in the 2-core. Let $\epsilon > 0$ be given. We will prove the cases of the Main Theorem which assert $R = o(n)$ by showing $R \leq \epsilon n$ with probability tending to one as $n \to \infty$. Pick η and i as in (i) of Corollary 5.16 so that $m_{\mathsf{deter},i} < \frac{\epsilon}{2} n$. Let L be as in Theorem 5.5 for η, so that if we stop the process at i, then with high probability $m_i = m_{\mathsf{deter},i} + O\left(L^i K_{i-1}\right)$. This becomes $m_i = m_{\mathsf{deter},i} + O\left(\sqrt{n}\log n\right)$, as i is fixed.

Using the simple observation that $R \leq m_i$, we get $R \leq m_i = \frac{\epsilon}{2} n + O\left(\sqrt{n}\log n\right)$, which is less than ϵn for n sufficiently large. □

For the other cases when $\beta_\infty > 0$ we seek to prove $R = \beta_\infty m + o(n)$, for we have already proved in Theorem 5.13 facts about c and β which align with the cases in the Main Theorem. (Note that unlike Theorem 5.13, the Main Theorem does not assert any statement for the case $c = c_*$ when $a > 1$). While we know quite well how the process mimics the deterministic system up to some time $i = \epsilon \log n$, we do not yet know if the number of edges m_i at time i mimics the number of edges in the 2-core. To this end we need to consider what happens between time i and reaching the 2-core so that the simple observation $R \leq m_i$ may be improved to $R = m_i + o(n)$. We will adopt a different approach and look at a slowed down removal process where only one vertex with degree 1 is removed at a time, the vertex being chosen uniformly at random. This may lead to vertices being removed in a very different order than the original process, but the terminal state—the 2-core of the original hypergraph—remains the same.

Let S_i denote the number of vertices with degree 1 after i steps of this new process and let T denote the stopping time when there are no more degree 1 vertices present. To be precise, our definition depends on a choice of an initial hypergraph. (The eventual intention is to run the original process for $\epsilon \log n$ iterations and use the resulting hypergraph as the starting hypergraph for the small step process, but this does not concern our definition of S_i.) Time indexing in the new process therefore

begins anew, in the sense that i counts the number of small steps (as opposed to total number of steps) and likewise T denotes $\inf\{i : S_i = 0\}$ (as opposed to the total time). Since each step of the slowed down process changes the number of hyperedges by 1, the result we are after is that S_i converges to 0 quickly—in time $T = o(n)$. Such a result means that the number of hyperedges R in 2-core differs from m_i by $o(n)$.

In the small step process, let us denote by $v_k^{(i)}$ the number of vertices of degree k after i steps and by r_i the number of hyperedges after i steps (so r_i deterministically decreases by 1).

Suppose vertex v in hyperedge e is removed at step i. The increment $S_{i+1} - S_i$ decreases by the number of degree 1 vertices in e (including v itself) and increases by the number of degree 2 vertices in e. The latter is distributed as the sum of Bernoulli variables (one for each degree 2 vertex) with success probability $\frac{2}{r_i}$.

So the process S_i lends the following conservative approximation (from the point of view of convergence to 0): at every step it decreases by 1 minus a binomial random variable, or more explicitly $S_{i+1} - S_i \leq -1 + \mathrm{Bin}\left(v_2^{(i)}, \frac{2}{r_i}\right)$. At time $i = 0$, we may approximate the expectation of this binomial using the deterministic system. From Lemma 5.15, $\frac{2V^{(i)}_{\mathsf{deter},2}([0,1])}{m_{\mathsf{deter},i}} = ae^{-ac_{\mathrm{den}}\beta_i}$ which is strictly less than 1 provided $\beta_i > \beta_{\mathsf{min}}$.

This holds for all cases with $\beta_\infty > 0$ once the case $c = c_*$ is excluded for $a > 1$. Therefore $E(S_1 - S_0) < 0$, and moreover Lemma 5.15 informs us that it is uniformly (in n) bounded away from 0.

We want to establish a simple condition for when $E[S_{t+1} - S_t \mid \mathcal{F}_t]$ is negative and bounded away from 0. Since the possibility of $t = T$ precludes it being strictly less than 0, we more precisely seek for every $0 < \sigma < 1$ some $\epsilon > 0$ so that $E[S_{t+1} - S_t \mid \mathcal{F}_t] \leq -\epsilon 1_{T>t}$ holds for all $t < n^\sigma$. With each step, $v_2^{(i)}$ may increase (if the hyperedge being removed contained degree 3 vertices), which acts to slow the convergence of S_i. We rely on another conservative approximation: there are at most n degree 3 vertices, so at each step $v_2^{(i)}$ increases by at most $\mathrm{Bin}\left(n, \frac{3}{r_i}\right)$. Essentially speaking, this is $O(1)$ at each step, which over the course of $o(n)$ many steps will not slow the process significantly.

These thoughts in mind, we define a new process $\overline{S}_i$, coupled so that $\overline{S}_i \geq S_i$ and $\overline{S}_{i+1} - \overline{S}_i \geq S_{i+1} - S_i$, which we achieve by defining the new increments as $\overline{S}_{i+1} - \overline{S}_i = -1 + \mathrm{Bin}\left(U_i, \frac{2}{r_i}\right)$ where $U_{i+1} - U_i = \mathrm{Bin}\left(n, \frac{3}{r_i}\right)$. With the bound $E(U_t - U_i) \leq \frac{3tn}{r_t} = o(n)$, it follows that $E\left(\overline{S}_{t+1} - \overline{S}_t\right) \leq E\left(\overline{S}_1 - \overline{S}_0\right) + o\left(\frac{n}{r_t}\right)$. For $\beta_\infty > 0$ one has $r_0 = \Omega(n)$, and so $r_t = \Omega(n) - t = \Omega(n)$. Therefore $o\left(\frac{n}{r_t}\right) = o(1)$.

Lemma 5.18. Given $0 < \sigma < 1$, there exists $y > 0$ and $\delta > 0$ such that

$$E\left[\exp\left(y\left(S_{t+1} - S_t\right)\right) \mid \mathcal{F}_t\right] \leq \exp\left(-\delta 1_{T>t}\right)$$

for all $t \leq n^{\sigma}$.

Proof. Let $f(y)$ denote $E\left[\exp\left(y\left(S_{t+1} - S_t\right)\right) \mid \mathcal{F}_t\right]$. The expectation exists since, for example, $0 \leq S_i \leq n$ for any i. The Taylor expansion for $f(y)$ is given by

$$f(y) = f(0) + yE\left[S_{t+1} - S_t \mid \mathcal{F}_t\right] + \cdots.$$

Since $f(0) = 1$, the function $g(y) = \log f(y)$ has Taylor expansion

$$g(y) = yE\left[S_{t+1} - S_t \mid \mathcal{F}_t\right] + \cdots.$$

Considering then the behavior of $g(y)$ near 0, it attains a negative local minimum $-\delta$ at some $y > 0$. □

Theorem 5.19. Suppose $S_0 = O\left(n^{\sigma_1}\right)$ for some $0 < \sigma_1 < 1$. Then $T = O\left(S_0\right)$.

Proof. The proof is a general martingale argument. Pick $\sigma_1 < \sigma < 1$ and let y and δ be as in the Lemma for this σ. Let $\overline{M}_i = \exp\left(yS_i + \sum_{j=0}^{i-1} \delta 1_{T>j}\right)$, and let $M_i = \overline{M}_{i \wedge n^{\sigma}}$ which is a supermartingale with respect to $\mathcal{F}_i$. One has the Markov bound $P\left(T > t\right) \leq I^{-1} E M_{t \wedge T}$, where $I = \inf_{T>t} M_{t \wedge T}$. Being a supermartingale, $EM_{t \wedge T} \leq EM_0 = \exp\left(yS_0\right)$. And $\inf_{T>t} M_{t \wedge T} = \inf_{T>t} M_t \geq \exp\left(\delta\left(t \wedge n^{\sigma}\right)\right)$. Together,

$P(T > t) \leq \exp(yS_0 - \delta(t \wedge n^\sigma))$. This probability may be made small by choosing $t = CS_0$ (which is $o(n^\sigma)$, and so passes through the min function) for a sufficiently large constant C. □

Proof of Main Theorem 2.1, remaining cases: Pick η and $\epsilon > 0$ as in (ii) of Corollary 5.16. Let L be as in Theorem 5.5 for η, so that if we stop the process at $i = \epsilon \log n$, then with high probability $m_i = m_{\mathsf{deter},i} + O(L^i K_{i-1})$. Now pick instead $i = \epsilon_1 \log n$ where $0 < \epsilon_1 < \epsilon$ is chosen sufficiently small depending on L so that $m_i = m_{\mathsf{deter},i} + o(n)$.

For this i, one has $b_i = O(n^\sigma)$ for some $0 < \sigma < 1$. Using Theorem 5.19, $R = m_i + O(b_i) = m_i + o(n)$. And as $i \to \infty$, we have $m_i = \beta_\infty m + o(n)$. Putting these last two estimates together, $R = \beta_\infty m + o(n)$. □

Bibliography

[1] A.Z. Broder, A.M. Frieze, and E. Upfal, *On the satisfiability and maximum satisfiability of random 3-CNF formulas*, Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 1993, pp. 322–330.

[2] J. P. Buhler, H. W. Lenstra, Jr., and Carl Pomerance, *Factoring integers with the number field sieve*, The development of the number field sieve, Lecture Notes in Math., vol. 1554, Springer, Berlin, 1993, pp. 50–94. MR 1321221

[3] Richard Crandall and Carl Pomerance, *Prime numbers*, Springer-Verlag, New York, 2001, A computational perspective. MR 1821158 (2002a:11007)

[4] Ernie Croot, Andrew Granville, Robin Pemantle, and Prasad Tetali, *Running time predictions for factoring algorithms*, Algorithmic number theory, Lecture

Notes in Comput. Sci., vol. 5011, Springer, Berlin, 2008, pp. 1–36. MR 2467835 (2009k:11202)

[5] ———, *On sharp transitions in making squares*, Ann. of Math. (2) **175** (2012), no. 3, 1507–1550. MR 2912710

[6] Martin Dietzfelbinger, Andreas Goerdt, Michael Mitzenmacher, Andrea Montanari, Rasmus Pagh, and Michael Rink, *Tight thresholds for cuckoo hashing via XORSAT*, CoRR **abs/0912.0287** (2009).

[7] John D. Dixon, *Asymptotically fast factorization of integers*, Math. Comp. **36** (1981), no. 153, 255–260. MR 595059 (82a:10010)

[8] O. Dubois and J. Mandler, *The 3-XORSAT threshold*, Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on, IEEE, 2002, pp. 769–778.

[9] Svante Janson and Malwina J. Luczak, *A simple solution to the k-core problem*, Random Structures Algorithms **30** (2007), no. 1-2, 50–62. MR 2283221 (2007k:05201)

[10] Michael Molloy, *Cores in random hypergraphs and Boolean formulas*, Random Structures Algorithms **27** (2005), no. 1, 124–135. MR 2150018 (2006f:05168)

[11] Boris Pittel and Gregory B. Sorkin, *The satisfiability threshold for k-XORSAT*, (2011), preprint.

[12] Boris Pittel, Joel Spencer, and Nicholas Wormald, *Sudden emergence of a giant k-core in a random graph*, J. Combin. Theory Ser. B **67** (1996), no. 1, 111–151. MR 1385386 (97e:05176)

[13] C. Pomerance, *Analysis and comparison of some integer factoring algorithms*, Computational methods in number theory, Part I, Math. Centre Tracts, vol. 154, Math. Centrum, Amsterdam, 1982, pp. 89–139. MR 700260 (84i:10005)

[14] Carl Pomerance, *The quadratic sieve factoring algorithm*, Advances in cryptology (Paris, 1984), Lecture Notes in Comput. Sci., vol. 209, Springer, Berlin, 1985, pp. 169–182. MR 825590 (87d:11098)

[15] ———, *The role of smooth numbers in number-theoretic algorithms*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994) (Basel), Birkhäuser, 1995, pp. 411–422. MR 1403941 (97m:11156)

[16] ———, *Multiplicative independence for random integers*, Analytic number theory, Vol. 2 (Allerton Park, IL, 1995), Progr. Math., vol. 139, Birkhäuser Boston, Boston, MA, 1996, pp. 703–711. MR 1409387 (97k:11174)

[17] Robert D. Silverman, *The multiple polynomial quadratic sieve*, Math. Comp. **48** (1987), no. 177, 329–339. MR 866119 (88c:11079)

[18] Gérald Tenenbaum, *Introduction to analytic and probabilistic number theory*, vol. 46, Cambridge university press, 1995.