The k-Cap Process on Geometric Random Graphs

Mirabel Reid   MREID48@GATECH.EDU
Georgia Tech

Santosh S. Vempala   VEMPALA@GATECH.EDU
Georgia Tech

Abstract

The k-cap (or k-winners-take-all) process on a graph works as follows: in each iteration, a subset of k vertices of the graph are identified as winners; the next round winners are the vertices that have the highest total degree from the current winners, with ties broken randomly. This natural process is a simple model of firing activity and inhibition in the brain and has been found to have desirable robustness properties as an activation function. We study its convergence on directed geometric random graphs in any constant dimension, revealing rather surprising behavior, with the support of the current active set converging to lie in a small ball and the active set itself remaining essentially random within that.

Keywords: Neuroscience, Geometric Random Graphs, Recurrent Networks

1. Introduction

Despite the resounding empirical success of recurrent networks in many applications, their underlying dynamics are still poorly understood. We analyze a simple recurrent process, analogous to the classical Hopfield Network (Hopfield, 1982), which represents the propagation of signals through a graph under inhibition. It has been proposed as a simple model for neural firing behavior, and it forms the backbone of an emerging model for computation in the brain, known as the assembly model.

The k-cap process repeatedly applies k-Winners-Take-All to the degrees of a random graph: at each time step $t > 0$, the *firing set* $A_t \subset V$ consists of the k vertices with the highest degree in A_{t-1} (with ties broken randomly). Given this process, some natural questions arise: How does the firing set A_t evolve with t? When the process does converge, how quickly does it do so?

We make two key contributions to the study of this model. First, we study the k-cap process on Geometric Random Graphs, and we show that the high local density of this model leads to rapid convergence. Second, we illuminate a new notion of convergence which more closely resembles the behavior of firing neurons in the brain.

Motivation from the Brain. The k-cap process forms the backbone of the assembly model for computation, an emerging line of inquiry at the intersection between computer science and theoretical neuroscience. An *assembly* of neurons is a subset of densely interconnected nodes within a brain area which tend to fire together in response to the same input to the brain area (Papadimitriou et al., 2020; Buzsáki, 2019). Assemblies are created through projection, where an outside stimulus fires (repeatedly), activating a subset of neurons. Two

ideas, rooted in experimental findings in neuroscience, lead to the convergence of assembly projection in a random brain graph. The first is inhibition: at each step, the k neurons with the highest total synaptic input are chosen to fire, while the rest are suppressed. The second is plasticity: if a neuron fires immediately following one of its pre-synaptic neighbors, the weight of the edge between them is increased. This causes neurons that 'fire together' to 'wire together', and strengthens internal connections each time an assembly is activated.

Rigorous analysis of the assembly model has thus far been based on a directed Erdös-Rényi random graph, where each pair of neurons has an equal probability of being connected via a synapse. There are two important ways in which this model departs from observed reality. First, the locations of neurons in the brain and the physical distance between them have a significant impact on the probability of connection. Long axons come with a cost in both material and

Figure 1: Two types of convergence. (a) convergence in the high-plasticity regime, and (b) the low-plasticity regime analyzed in this paper.

energy, so neurons tend to prefer to create connections that are close in physical space. The principle of conservation of axonal wiring costs was proposed by Ramon y Cajal in the early 20th century (Ramón y Cajal, 1911), and the relationship between distance and connection probability has been confirmed empirically (Bullmore and Sporns, 2009; Cuntz et al., 2010). Moreover, models that take locality into account are better able to explain statistical deviations of the connectome from the standard random graph model, as observed in experiments (Song et al., 2005). Second, in the standard random graph model, assemblies are shown to correspond to the firing of k neurons, with most of them in a fixed set of size $(1 + o(1))k$ with at most $o(k)$ outside this set. On the other hand, what has been observed is that assemblies represent increased firing activity of a relatively small but significantly larger than k subset of neurons for a period of time (Durstewitz et al., 2000; Buzsáki, 2019). The difference between these two types is illustrated in Figure 1.

We address the first departure by studying the k-cap process on Geometric Random Graphs, a graph model where the connection probability varies with distance. We additionally find that the k-cap process on this graph model exhibits the second type of convergence to a larger-than-k subset. An exciting aspect of our investigation is a rigorous explanation of this phenomenon.

Geometric Random Graphs The behavior of the k-cap process depends heavily on the graph structure. On the complete graph, every vertex has degree k from the firing set A_t; so, assuming random tie-breaking, all vertices fire with probability k/n at each time step. On the other hand, if G is a sparse graph with a planted k-clique H, we expect $A_t = H$ to be a fixed point (for k sufficiently large). While directed Erdős–Rényi graphs are simple to study, we do not expect to see meaningful convergence on this model. We take inspiration
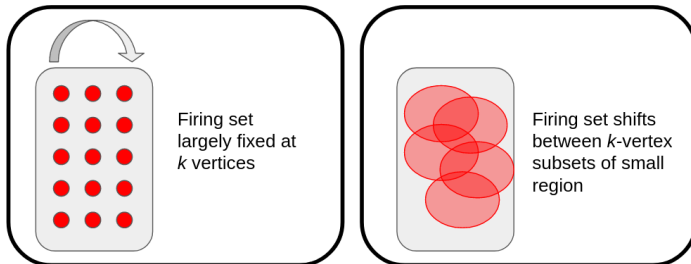
from the graph-theoretic structure of the brain and study this process on the Geometric Random Graph model.

In geometric random graphs, each vertex is assigned a position in a hidden variable space (for example, the cube $[0,1]^d$). The probability that an edge $\mathbb{1}_{(x,y)}$ exists in the graph is a function of the hidden variables of the endpoints. By using an edge probability function which decreases with distance in the hidden variable space, this creates subgraphs which are dense and concentrated within a small diameter subset of the space. The hidden variables can correspond to spatial distance, or they can represent similarity in a wider set of features. For example, the geometric random graph model has been used for social networks, where the hidden variable represents a closeness in "social space" rather than physical distance (Boguná et al., 2004). This model has also been studied in the context of transportation networks, communication networks, and networks of neurons (Bullmore and Sporns, 2009; Barthélemy, 2011).

Properties of geometric random graphs have been thoroughly explored; see Penrose (2003) for comprehensive exposition. In the most common variant of the model, all vertices are placed in a d-dimensional space according to some distribution. If the distance between two vertices is less than r (where r is a parameter of the model), they are connected by an edge; otherwise, they are not. We study a *directed, soft* geometric random graph where the edge probability decays exponentially with squared distance, i.e., the Gaussian kernel. This alternative model introduces asymmetry as well as long-range connections, both of which are important for real-life networks.

1.1. Definition and Intuition

As a warm-up, we consider the infinite limit, i.e., the *continuous* interval $[0,1]$ in one dimension. Then we turn to the discrete setting of graphs, with vertices chosen uniformly from the d-dimensional unit cube. While $d = 2, 3$ apply to the coordinates of a neuron in space, higher dimensions are also relevant and interesting, as vertex location could indicate some set of relevant features (e.g., type of neuron).

A Continuous Process. A natural abstraction of the k-cap process on geometric random graphs is to consider what happens when the number of vertices, n, goes to infinity. On a finite graph, the input to a discrete vertex v is the sum of its edges from $A_t \subset V = \{1/n, 2/n, \ldots, 1\}$. In the infinite limit, we assume that $A_t \subset [0,1]$ is a set of measure α, leading to a corresponding α-cap process. The input to a given point $v \in [0,1]$ is the integral of the edge probability function over A_t and the α fraction of points with the highest input will form A_{t+1}. The formal definition of this process is stated below:

Definition 1 (α-cap Process in 1-D) *Let A_0 be a finite union of intervals on [0,1]. For an integrable function $g : [-1, 1] \to \mathbb{R}$, let $F_t(x) = \int_0^1 A_t(y)g(y - x)\,dy$. Then, A_{t+1} is defined recursively:*

$$A_{t+1}(x) = \begin{cases} 1 & \text{if } F_t(x) \geq C_t \\ 0 & \text{otherwise} \end{cases}$$

where C_t is the solution to $\int_0^1 A_{t+1}(x)dx = \alpha = |A_0|$.

This continuous abstraction leads to a clean convergence phenomenon. We find that A_t converges to a single interval of length α. The speed of convergence depends on the derivative of the edge probability function.

Theorem 2 *Let A_0 be a finite union of intervals in [0,1]. Let $g : [-1, 1] \to \mathbb{R}_+$ be a differentiable, even, nonnegative and integrable function with $g'(x) < 0$ for all $x > 0$. For any such g, the α-cap process converges to a single interval of width α. Moreover, the number of steps to convergence is*

$$O\left(\frac{\max_{[0,1]}|g'(x)|}{\min_{[\frac{\alpha}{8},1]}|g'(x)|}\right).$$

Note that the above conditions capture any distance function that decays smoothly with the distance between its endpoints, e.g., the Gaussian kernel. This process is deterministic given the initial choice of A_0. We are able to bound the convergence using a simple potential function: *The distance between the medians of the leftmost and rightmost intervals of A_t decreases.*

In this way, the intervals are "squeezed" together until they collapse into one. In this continuous version, any sub-interval in $[0, 1]$ of length α is a fixed point; if $A_t = [a, b]$, then $A_{t+1} = A_t$. Moreover, a single interval is the only possible fixed point.

This process is illustrated for an example starting set in Figure 2. While A_0 has five initial intervals, F_0 has only two local maxima, causing the total number of intervals to reduce to 2 in A_1. In subsequent steps, the intervals become increasingly unbalanced. They shift closer together until, ultimately, the larger interval gains all the mass.

Formal Definition of the Discrete Process.
Now we turn to the main setting of this paper, the k-cap process on a finite directed graph G. The following symbols will be used for the rest of the paper. Let n be the number of vertices in the graph and k be the number of vertices activated at each step. A_t represents the set of k vertices activated at step t for $t = 0, 1, 2, \ldots$. Let $\mathbb{1}_{(x,y)}$ be the indicator variable for the directed edge between two vertices x and y.
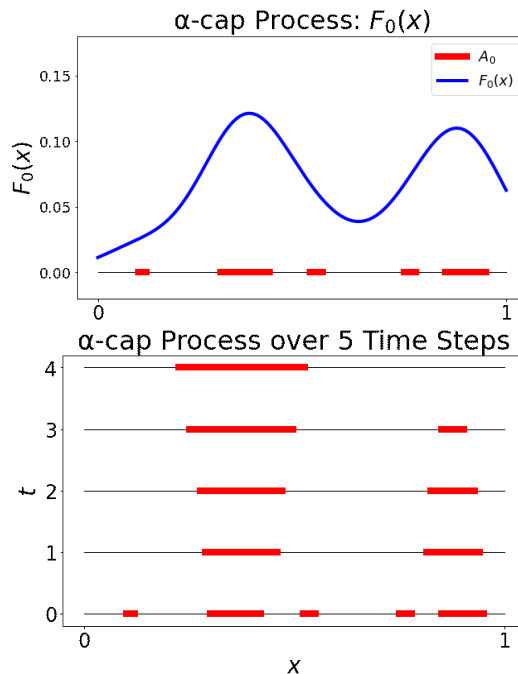

Figure 2: (Top) The influence function, $F_0(x)$, for an example starting set A_0. The edge weight function (g) is the Gaussian kernel with $\sigma = 0.1$. (Bottom) The evolution of the same A_0 over time. This example converges at $t = 4$.

4

Definition 3 (k-cap Process) *Assume $A_t \subset \{1, 2, \ldots, n\}$, and $|A_t| = k$. Let $F_t :$ $\{1, 2, \ldots, n\} \to \{0, 1, 2, \ldots, k\}$ be the synaptic input function at time t, defined as follows:*

$$F_t(x) = \sum_{y \in A_t} \mathbb{1}_{(y,x)}$$

Let C_t be the smallest integer such that $|\{x \mid F_t(x) > C_t\}| \leq k$, and let $A_{t+1} = \{x \mid F_t(x) > C_t\} \cup A_{t+1}^*$, where A_{t+1}^* is a set of points sampled at random from $\{x \mid F_t(x) = C_t\}$ such that $|A_{t+1}| = k$.

In the k-cap process, A_{t+1} is chosen as the k vertices with the highest degree from A_t. If there are ties, the remaining vertices are chosen uniformly at random from the set of vertices with the next highest degree. A_0 can be instantiated in any way, but we assume that it is chosen uniformly at random from the set of vertices.

We analyze the convergence of the k-cap process on a d-dimensional Gaussian geometric random graph; the probability of an edge between two vertices with hidden variables x and y is a Gaussian kernel, parameterized by σ. For simplicity, we use x to represent both the vertex and its hidden variable in $[0, 1]^d$. Throughout this paper, we will use the terms *point* and *vertex* interchangeably.

Definition 4 (d-dim Gaussian Geometric Random Graph) *Let $G = G_\sigma = (V, E)$. Let $V = \{v_1, v_2, \ldots, v_n\}$ where each vertex is a point chosen uniformly at random in $[0, 1]^d$. Each directed edge (x, y) is present in the graph with probability*

$$\mathbb{P}(\mathbb{1}_{(x,y)}) = g(x, y) = e^{-\frac{\|x-y\|_2^2}{2\sigma^2}}$$

All edges are independent conditioned on the locations of the vertices. Unless otherwise stated, assume $\|x - y\| = \|x - y\|_2$ is the Euclidean distance.

Convergence of the Discrete Process. 0The discrete process on graphs turns out to exhibit much more complex behavior than the continuous variant. With the randomness induced by the choice of edges, the convergence behavior also becomes probabilistic rather than ending in a fixed set or distribution. We will prove that for an interesting range of σ, the cap A_t converges with high probability to lie within a small ball (an interval when $d = 1$). A_t will not be a fixed set of points for all $t \geq t^*$, but will instead oscillate randomly between vertices, with the firing probability dropping toward the edges of the ball. This type of convergence is illustrated in Figure 1. It can also be seen in simulation, as shown in Figure 3. This behavior is not only mathematically interesting, but it also matches the observed behavior of assemblies (Durstewitz et al., 2000; Buzsáki, 2019).

Evolution of the k-cap. The evolution of the structure of A_t also reveals interesting properties of the k-cap process on random graphs. In the first step, A_0 is uniformly distributed. However, due to the the *Poisson clumping* phenomenon (Aldous, 2013), there will be a several regions of the ball with a higher concentration than average. As we will show, for σ sufficiently small, A_1 will be concentrated within $k^{1/4+o(1)}$ balls which are small compared to $[0, 1]^d$ (This result is described formally in Theorem 5). As t increases, all but one of these balls will diminish and disappear. With high probability, each ball will shrink by a fixed fraction at each step. After this, we will show that one ball will "win" over the others. This is illustrated in 1 and 2-D in Figures 3 and 4.

5

Figure 4: The k-cap at time steps 0, 3, 6, and 9, with parameters $n = 20000$ and $k = 200$. Dots represent the location of firing neurons in 2-D space at time t.

Parameter Range. In this paper, we focus on $\sigma = \Theta(1/k^{1/d})$. The justification for this parameter range lies in the concentration behavior of uniform random variables. The soft geometric random graph model can be thought of as an approximation of an interval graph with radius $\Theta(\sigma)$. Let $U = \{U_1, U_2, \ldots U_k\}$ be a set of k random variables, each chosen uniformly at random in $[0,1]^d$. For a given radius r, we can compute the maximum number of points which are likely to fall into a ball of radius r (this result is described in Lemma 18). Note the expected number of points in a ball I is $k\mathrm{Vol}(I)$, and $\mathrm{Vol}(I) = \Theta(r^d)$; as r increases, the maximum degree approaches

Figure 3: The k-cap at 4 time steps, with parameters $n = 90000$ and $k = 40$.

the expected value. This phenomenon means that the concentration behavior starts to disappear as σ increases past $1/k^{1/d}$. On the other hand, as r decreases, the maximum degree approaches 1. Therefore, we focus on an intermediate range of σ where the concentration of A_0 leads to interesting behavior.

Additionally, we assume that n is a sufficiently large polynomial function of k; $n = k^\beta$, for a constant $\beta \geq 2 + d$. When n is large and $t = \mathrm{polylog}(n)$, the points in A_t are likely firing for the first time . This lets us make the simplifying assumption that, conditioned on the positions of the vertices in $[0,1]^d$, $F_t(x)$ is a sum of independent indicators.

1.2. Main Results

For clarity, we state the parameter range analyzed in this paper as an assumption and reference it when applicable.

Assumption 1 *The graph G is parameterized by n, σ, which have the following relationships to k: $n = k^\beta$, for some constant $\beta \geq 2 + d$, and $\sigma = \Theta(k^{-1/d})$.*

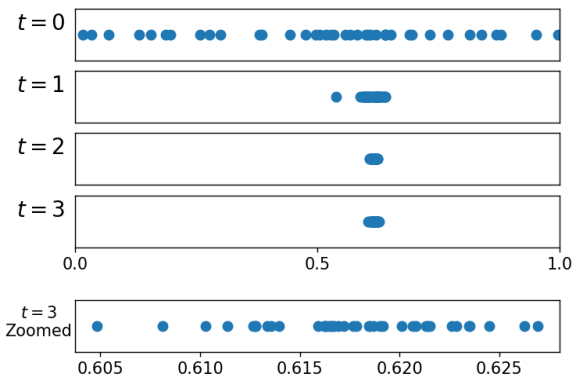We also assume that A_0 is chosen uniformly at random from the set of vertices V.

6

We have the following theorem describing points which have a significant probability of exceeding the threshold at Step 0. By the assumption that $\sigma = \Theta(k^{-1/d})$, the set A_1 is contained in a small region relative to $[0, 1]^d$.

Theorem 5 *Suppose G fulfills Assumption 1, and A_0 is chosen uniformly at random from V. With high probability, A_1 can be covered by $k^{\frac{1}{4}+o(1)}$ balls, each of radius $O(\sigma\sqrt{\ln \ln k})$ and pairwise separated by a distance of at least $2\sigma\sqrt{\ln n}$.*

Let $p_t(x)$ be the probability that x fires at step t, conditioned on the graph edges. Figure 5 (a) shows an empirical demonstration of Theorem 5. Even at the first step, small variation in the expected value lead to a significant imbalance in firing probability.

The next step of the analysis will show that A_t will gradually become concentrated in a single ball of radius $O(\sigma\sqrt{\ln k/k})$. The key idea is that each individual ball shrinks with high probability at each step, almost in place.

Theorem 6 *Suppose G fulfills Assumption 1, and A_0 is chosen uniformly at random from V. There exists a $t^* \leq \ln^c k$, for a constant c, such that A_{t^*} can be covered by a single ball of radius $\Theta\left(\sigma\sqrt{\ln k/k}\right)$.*

This structure is, to within a log factor, the smallest subgraph we can expect A_t to converge to. Within a region of radius $O(\sigma k^{-1/2})$, the edge probability is greater than $e^{-1/k}$; hence, within this ball, any vertex has a constant probability of having degree k from A_t. Note that this directly implies the qualitative convergence behavior illustrated in Figure 1.

Finally, we show that A_t remains concentrated in the infinite time horizon. Conditioned on the structure of the graph, almost all of A_t is contained within a ball of radius $O(\sigma k^{-1/3+\epsilon})$ for all $t \geq t^*$. Note that the diameter bound in Theorem 7 is necessarily weaker, since A_t shifts randomly between k-vertex subsets of the small ball. Given exponential time, there will likely be combinations which happen to activate outlier points. We argue that these outliers cannot compound to split the firing set, and A_t remains concentrated.

Theorem 7 *Suppose G fulfills Assumption 1, and A_0 is chosen uniformly at random from V. For all $t \geq t^*$, with high probability, there exists a ball I_t with radius $r = \sigma k^{-1/3+\epsilon}$, for a constant $\epsilon > 0$, such that $|A_t \cap I_t| > k - k^{2/3}$.*

The proof of this theorem directly implies the following structural property of geometric random graphs, which holds for all $S \subset V$ of size k that are mostly contained in a small ball in $[0, 1]^d$. With high probability over all such sets, the set of k vertices with the highest degree from S are also mostly contained in a small ball.

Corollary 8 *Let $G = (V, E)$ be a geometric random graph such that for every vertex $x \in V$, its location h_x is chosen uniformly at random from $[0, 1]^d$ and for every pair $x, y \in V$, $\mathbb{P}((x, y) \in E) = e^{-\|h_x - h_y\|k^{2/d}}$, where $k = O(|V|^{\frac{1}{2+d}})$. Let $r = k^{-1/d-1/3+\epsilon}$ for any $\epsilon > 0$. Then, with high probability (over the edges of G), for every set $S \subset V$ of size k, if at least $k - k^{2/3}$ points of S are contained in a ball of radius r, then there exists a ball of radius r which contains at least $k - k^{2/3}$ points of S', the set of k points with the highest degree from S.*
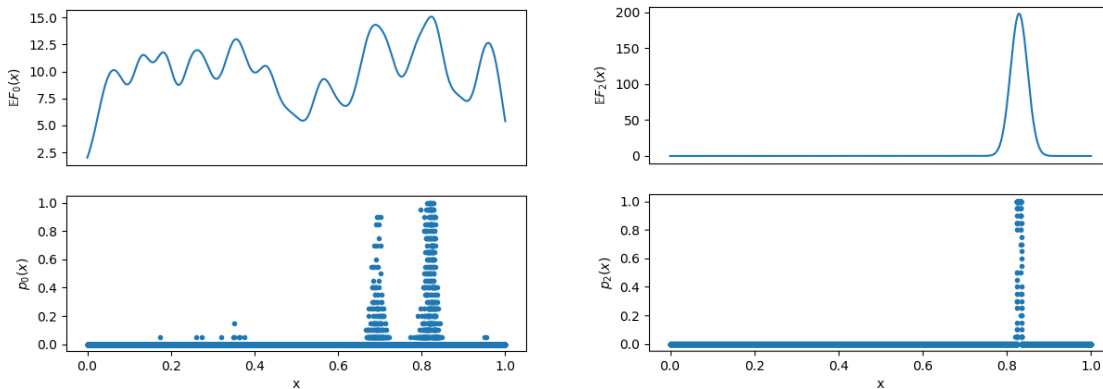
7

Figure 5: Expected input and $p_t(x)$ at $t = 0$ (Left) and $t = 2$ (Right). Parameters used are $n = 20000$ and $k = 200$. The probability $p_t(x)$ was estimated by fixing the firing set, and then repeatedly redrawing the graph edges. The probability of x is defined as the percentage of times x was chosen by the top-k function.

2. Analysis of the discrete k-cap process

In Appendix A, we will introduce a few general results on probability which play key roles in the proof. In Section 2.1, we describe the structure of A_1 given the random initial firing set A_0. Lastly, in Section 2.2, we investigate the evolution of A_t as t increases, and we prove that the set of firing points at t converge to lie in a single small ball.

2.1. Characterization of A_1

Since the initial firing set A_0 is chosen uniformly at random from V, A_1 is concentrated near dense sections of A_0. We argue that the probability that $x \in A_1$ can be characterized by conditioning on the number of points of A_0 within $\tilde{O}(\sigma)$ of x. By analyzing the distribution of dense subsets of a set of uniform random variables, we show that A_1 is contained within a union of $k^{1/4+o(1)}$ small balls.

First, in Lemma 9, we give a lower bound on the first threshold, C_0, by examining the maximum number of uniform random points within a ball of radius r. There are, with high probability, at least k vertices of V which connect to the every point in the intersection of the ball with A_0.

Next, Lemma 10 shows that if $|\{y \in A_0 : \|x - y\| = O(\sigma\sqrt{\ln \ln k})\}|$ is not large, x has a very small probability of achieving an input of C_0. This implies that A_1 must be solely contained within high-density regions of A_0.

Finally, we combine these two lemmas to prove Theorem 5 (stated above). Since the number of high-density regions of A_0 can be bounded using a combinatorial argument, A_1 must be contained within $k^{1/4+o(1)}$ small balls. We note that the constant $1/4$ may not be optimal; however, a tighter bound would only effect the convergence time up to log factors. The proofs of Theorem 5 and Lemma 10 are technical and will be deferred to the appendix.

Lemma 9 *With probability $1 - o(1)$, $C_0 \geq \frac{\ln k}{\ln \ln k}\left(1 + \frac{1}{4}\frac{\ln^{(3)} k}{\ln \ln k}\right)$ (where $\ln^{(3)} k = \ln \ln \ln k$).*

Proof Consider a graph constructed on A_0 as follows. For any $a_1, a_2 \in A_0$, add an edge if $\|a_1 - a_2\| < \frac{1}{2}\sigma\sqrt{\ln\ln k}$. This is a geometric graph, and tight bounds on the maximum degrees for such graphs are known (Penrose (2003), Theorem 6.10). The precise result we use is stated in Appendix A, Lemma 18. Let Δ_k be the maximum degree of the graph. With high probability,

$$\Delta_k \geq \frac{\ln k}{\ln\ln k}(1 + \eta) = M_k \qquad \text{where } \eta = \frac{\ln^{(3)} k}{4\ln\ln k}$$

$\Delta_k + 1$ is the maximum intersection of A_0 with a ball of radius $r = \frac{1}{2}\sigma\sqrt{\ln\ln k}$. Call this ball I. The above result tells us that $|I \cap A_0| \geq M_k$ almost surely.

The maximum distance within the ball is $\sigma\sqrt{\ln\ln k}$. For any $x \in I$, the probability that x connects to M_k points is at least $g(x, x + v)^{M_k}$, where v is a vector of size $\sigma\sqrt{\ln\ln k}$. Substituting:

$$\mathbb{P}(F_0(x) \geq M_k) \geq \exp\left(-\frac{(\sigma\sqrt{\ln\ln k})^2}{2\sigma^2}\right)^{M_k} = \exp\left(-\frac{\ln\ln k}{2}\right)^{(1+\eta)\frac{\ln k}{\ln\ln k}} = k^{-\frac{(1+\eta)}{2}}$$

With a union bound argument, we can lower bound the number of vertices of V in I (stated in Appendix A, Lemma 16).

$$|I \cap V| = \Omega\left(\frac{\text{Vol}(I) \cdot n}{\log n}\right) = \Omega\left(\sigma^d(\ln\ln k)^{d/2} \cdot \frac{n}{\log n}\right)$$

By the assumption that $\sigma = \Theta(k^{-1/d})$ and $n = k^\beta$, there are $k^{\beta-1-o(1)}$ vertices of G in I.

The expected number of points with degree M_k from $|A_0 \cap I|$ is at least $k^{\beta-1-o(1)}k^{-\frac{(1+\eta)}{2}}$, which is much greater than k. Therefore, there are at least k points with input M_k with high probability. This implies that the threshold C_0 is bounded from below by M_k. ∎

Lemma 10 *For any $x \in [0,1]^d$, define $B_r(x) = \{y \in [0,1]^d : \|x - y\| < r\}$. Let $r = \sigma\sqrt{24\beta\ln\ln k}$. Suppose that the overlap between $B_r(x)$ and A_0 is at most $\frac{3\ln k}{4\ln\ln k}$. Conditioned on this event, the probability that $x \in A_1$ is at most $\frac{1}{n^3}$.*

2.2. Convergence of A_t

In this section, we will prove an induction step, Lemma 15, which will lead to the proof of Theorem 6. In Theorem 5, we have proved that A_1 can be covered by $k^{1/4+o(1)}$ balls of radius $O(\sigma\sqrt{\ln\ln k})$, and separated by at least $2\sigma\sqrt{\ln n}$. There are three key properties of this system which make the analysis tractable. First, the separation condition allows us to analyze each interval as a separate system. If $x \in I_a$ and $y \in I_b$, $g(x, y) < \exp\left(-\frac{4\sigma^2\ln n}{2\sigma^2}\right) = n^{-2}$. Therefore, with high probability, the subgraphs defined by I_a and I_b are independent; this means that all $x \in A_t$ will not receive input from outside its interval. Second, since the graph is directed, the edge $\mathbb{1}_{(x,y)}$ is independent of $\mathbb{1}_{(y,x)}$ conditioned on the positions of the vertices $x, y \in [0,1]^d$. Thirdly, we prove in Lemma 12 that for any $t = \text{polylog}(k)$, all points which fire at t are 'new' (i.e., they have not fired at a previous step) with high probability.

This lets us make the simplifying assumption that $F_t(x)$ is a sum of independent indicators. Using these three key simplifications, we prove that, with high probability, each separated interval shrinks to a size of $O(\sigma\sqrt{\ln k/k})$.

We will suppose that the hypothesis of Theorem 5 holds for a step $t \geq 1$; A_t can be covered by $O(k^{1/4+o(1)})$ sufficiently separated balls. Then, we will prove that the separation and coverage continue to hold by induction.

Define $A_t \subset I_1 \cup I_2 \cup \cdots \cup I_i$, where each I_j is a ball of radius $O(\sigma\sqrt{\ln \ln k})$, and all pairs I_a, I_b are separated by a gap of at least $2\sigma\sqrt{\ln n}$. Also define $E[x] = \mathbb{E}F_t(x) = \sum_{z \in A_t} g(x, z)$, and $V[x]^2 = \operatorname{Var} F_t(x) = \sum_{z \in A_t} g(x, z)(1 - g(x, z))$. Note that $E[x]$ and $V[x]$ depend implicitly on t. When the ball I is unambiguous, we will denote $\hat{k} = |A_t \cap I|$ to be the number of firing points lying in I.

The following lemmas will be used to bound C_t at each time step. Using this, we can get precise bounds on $\mathbb{P}(F_t(x) > C_t)$. First, we bound the gradient of $E[x]$ in space.

Lemma 11 *For any vector direction v and point $x \in [0, 1]^d$, $|\nabla_v E[x]| < \frac{k}{\sigma}\sqrt{d/e}$.*

For this proof to be viable, we will need to show that $F_t(x)$ is the sum of independent indicators. By definition of the graph structure, each edge $\mathbb{1}_{(y,x)}$ is drawn independently. However, we will also need to show that, for each $y \in A_t$, its edges $\mathbb{1}_{(y,x)}$ have not been used in previous computations. This follows from the next lemma.

Lemma 12 *Suppose $t = O((\ln k)^c)$ for a constant c. Then, with probability at least $1 - \frac{1}{k^{1/2-o(1)}}$, $A_0 \cap A_1 \cap \cdots \cap A_t = \emptyset$.*

Lemma 12 implies that, conditioned on the set A_t, $F_t(x) = \sum_{z \in A_t} \mathbb{1}_{(z,x)}$ is a sum of independent indicators (with no dependence on previous time steps). Therefore, C_t can be bounded using standard concentration bounds as follows:

Lemma 13 *At any step $t = O((\log k)^c)$, assuming the conditions of Lemma 15, with high probability, $C_t \geq \max_x E[x]$*

Finally, we can use the above lemma to relate the probability that a point fires at time $t + 1$ to its expected value at time t.

Lemma 14 *Let $y \in I_j$, and $\hat{k} = |I_j \cap A_t|$. If there exists an $x \in I_j$ such that $\mathbb{E}F_t(x) > \mathbb{E}F_t(y) + \sqrt{6\beta(\hat{k} - \mathbb{E}F_t(y))\ln k}$, then $\mathbb{P}(F_t(y) > C_t) < \frac{1}{n^3}$*

Finally, we will prove Theorem 6. We will use an induction argument; the base case was shown in Theorem 5, and the following Lemma will form the basis of the induction hypothesis.

Lemma 15 *Suppose that A_t can be covered by i balls, $A_t \subset I_1 \cup \cdots \cup I_i$, which obey the following conditions:*

- *Count: $i = O(k^{1/4+o(1)})$,*

- *Radius: for each j, I_j is a ball of radius r_j, bounded by $r(I_j) = \Omega(\sigma\sqrt{\ln k/|I_j \cap A_t|})$ and $r(I_j) < C\sigma\sqrt{\ln \ln k}$ for some constant C.*

- *Separation: the distance between any two balls is at least $2(1 - o(1))\sigma\sqrt{\ln n}$.*

At the next step, with high probability, $A_{t+1} \subset I'_1 \cup I'_2 \cup \cdots \cup I'_i$, where $d(I_j, I'_j) = \max_{x \in I'_j} \min_{y \in I_j} \|x - y\| < 5(r(I_j) - r(I'_j))$, and $r(I'_j) \leq (1 - \frac{1}{(\ln k)^c})r(I_j)$ for an absolute constant c.

This Lemma will show that the radius of each ball shrinks at each step; that is, A_t is contained within a union of balls of radius r_t, where r_t is a decreasing function of t. The main idea of the proof is to show that, regardless of the actual positions of points in $A_t \cap I$, vertices toward the center of I have a small advantage over vertices toward the edge. Thus, either (1) the position of points in $A_t \cap I$ is particularly unbalanced, and A_{t+1} shifts toward one side, or (2), the radius of I shrinks in all directions.
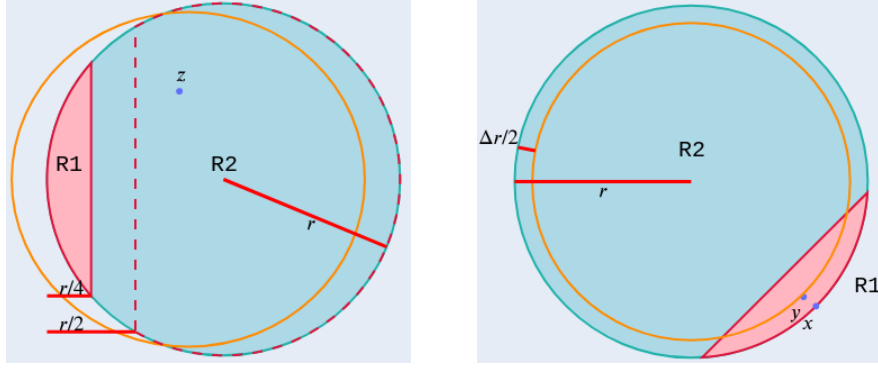

Figure 6: The division of the ball I into two subregions. In case 1, there exists a division of I into two sub regions R_1 and R_2 such that $R_2 \cap A_t < \hat{k}/(\ln k)^\alpha$. We bound the gradient of $E[z]$ for all z in the region enclosed by the dotted line. In case 2, no such division exists. We prove that for y between the outer and inner circles, $\mathbb{P}(y < 1/n^3)$. In both cases, we prove that A_{t+1} falls in the orange circle with high probability.

Proof [Abbreviated Proof of Lemma 15] Fix one ball $I = B_r(p)$. Let $r = r(I)$ be the radius of I, and let p be its center. Define $F_t(x; S) = \sum_{y \in A_t \cap S} \mathbb{1}_{(y,x)}$, the input to x from firing points in S. We will prove that with high probability, $\{x \in V : F_t(x; I) \geq C_t\}$ can be covered by $I' = B_{r'}(p')$, where $r' = (1 - 1/(\log k)^c)r$ and $\max_{z \in I'} dist(I, z) < 5(r(I) - r(I'))$. Call this statement (*).

If statement (*) holds for each $I = I_i$, the lemma is proven. This holds by the separation assumption; for x near I, $F_t(x; I)$ is a good proxy for $F_t(x)$. In particular, if $dist(I, x) < \sigma\sqrt{\ln n}$, then $F_t(x; I) = F_t(x) - o(1)$ with high probability.

To prove this statement, we will consider two cases. In case 1, we suppose that the distribution of $A_t \cap I$ is imbalanced. In particular, there exists a half space dividing I into two spherical caps, with heights $r/4$ and $7r/4$, such that the larger segment contains only $|A_t \cap I|/(\ln k)^\alpha$ points of A_{t+1} (for an $\alpha \geq 1$). See Figure 6. We will show that given this imbalance, statement (*) holds.

11

In case 2, no such division exists. We will show that for any point z near the boundary of I is disadvantaged compared to a point near the center. Thus, $I' = B_{r'}(p)$ for an $r' = (1 - 1/(\log k)^c)r$, which implies statement (*). See Figure 6.

For both cases, the argument will use a bound on the gradient of $E[z]$. With this, we will construct a point w such that $E[w] - E[z]$ is large, and use Lemma 14 to argue that $\mathbb{P}(z \in A_{t+1}) < 1/n^3$. The argument for this is technical, so it will be deferred to the appendix. ∎

This proves that at some $t = (\ln k)^c$, the firing points are contained within $k^{1/4+o(1)}$ small, separated balls. Theorem 6 follows using a coupling argument which shows that one small ball must 'win' all the points.

Proof [Proof of Theorem 6] By Theorem 5, the conditions of Lemma 15 hold at step 1. Additionally, by the condition that each ball does not shift by more than $5(r(I) - r(I'))$ at each step, the separation condition holds inductively for any $t = \text{polylog}(k)$. The maximum distance moved by a single ball by time t is $5(r(I_j^{(0)}) - r(I_j^{(t)})) = O(\sigma\sqrt{\ln \ln k})$, which maintains the separation of $2(1 - o(1))\sigma\sqrt{\ln n}$. Therefore, applying Lemma 15 inductively, the radius of I is reduced to $O(\sigma\hat{k}^{-1/2}\sqrt{\ln k})$ at $t = \text{polylog}(k)$, where $\hat{k} = |A_t \cap I|$. Recall that there are $k^{1/4+o(1)}$ separated balls; a similar method will allow us to eliminate balls that are $\sigma\hat{k}^{-1/2}\sqrt{\ln k}$ in size.

By the pigeonhole principle, at least one ball receives $k^{3/4-o(1)}$ points. Given that the size of the balls are small enough that most vertices within them are neighbors, $C_t \geq k^{3/4-o(1)}$. Therefore, for each j, if I_j is not eliminated, there exists an $x \in I_j$ such that $\mathbb{E}F_t(x) \geq k^{3/4-o(1)}$ (as always, the expectation is over the assignment of edges from A_t to x, conditioned on the above instantiation of A_t).

If one ball contains a much larger proportion of firing points than any other, then it 'wins' all the mass at the next step with high probability. Suppose instead that there are multiple balls at t with similar counts of firing points. In this case, conditioned on A_{t-1}, the size of $I_1 \cap A_t$ is a random variable over the edges from A_{t-1} to I_1. By Lemma 12, we can assume this is independent from previous time steps, and it can vary by $\sqrt{\mathbb{E}|I_1 \cap A_t|}$ with constant probability. Consider two alternative scenarios, which can each occur with constant probability.

$$\begin{cases} (1) & \max_{x \in I_1} \mathbb{E}F_t(x) = X - \Theta(\sqrt{X}) \\ (2) & \max_{x \in I_1} \mathbb{E}F_t(x) = X + \Theta(\sqrt{X}) \end{cases}$$

Let $y = \text{argmax}_{y \in I_2} \mathbb{E}F_t(y)$. So, it is clear that in either scenario (1) or (2), the inputs to x and y differ by the number of points added to I_1.

$$\left| \max_{x \in I_1} \mathbb{E}F_t(x) - \max_{y \in I_2} \mathbb{E}F_t(y) \right| = k^{3/8-o(1)}$$

The distributions of sums of non-identical indicators are well studied, so we can use known tail bounds to bound the ratio between $\mathbb{P}(F_t(z) > C_1)$ and $\mathbb{P}(F_t(z) > C_2)$ for two thresholds C_1, C_2. Precisely, $\frac{\mathbb{P}(F_t(z)>C_1)}{\mathbb{P}(F_t(z)>C_2)} > \exp\left(\frac{(C_2-C_1)^2}{C_2}\right)$. The result we use is stated and proven in the Appendix A, Lemma 21. Since the two thresholds differ by $\Theta(\sqrt{X})$, between the two scenarios, either $p_{t+1}(z)$ increases by a constant factor for all $z \in I_1$, or $p_{t+1}(w)$

decreases by a constant factor for all $w \in I_2$. So, this implies that $|\mathbb{E}F_{t+1}(x) - \mathbb{E}F_{t+1}(y)|$ varies by $\Theta(1) \min_{i \in \{1,2\}}(|I_i \cap A_t|)$ between the two scenarios.

This is a significant variation; as in Lemma 13, for any $x \in [n]$, $C_t \geq \mathbb{E}F_{t+1}(x)$. So, in the case where x has more expected input, $\mathbb{E}F_{t+1}(y) < \mathbb{E}F_{t+1}(x) - \Theta(\mathbb{E}F_{t+1}(y))$. By the Chernoff bound (Lemma 19), the probability that $F_{t+1}(y)$ will exceed C_{t+1} is exponentially small.

$$Pr\left(F_{t+1}(y) > C_{t+1} = (1 + \Theta(1))\mathbb{E}F_{t+1}(y)\right) < \exp\left(-\Theta(1)\mathbb{E}F_{t+1}(y)\right) < \exp\left(-k^{3/4-o(1)}\right)$$

A similar argument applies if $\mathbb{E}F_{t+1}(y) > \mathbb{E}F_{t+1}(x)$. Therefore, since there is a constant probability that the two balls will deviate from each other, either I_1 or I_2 will be eliminated in a constant number of steps.

The same argument applies to any pair of balls (I_i, I_j). Therefore, the number of balls reduces by a constant factor within a constant number of steps. This leads to convergence to a single ball within $O(\ln k)$ steps.

At this point, $|I \cap A_t| = k$, so applying Lemma 15 again, we can conclude that A_t converges to a single ball of size $O(\sigma k^{-1/2}\sqrt{\ln k})$ in $O((\log k)^c)$ steps. ∎

Note that our results also imply that A_t does not converge to a fixed set; rather, it shifts randomly, remaining inside a ball at each step (see Figure 1, with A_t converging in the sense of (b)).

Claim 15.1 *For $t = polylog(k)$, A_t shifts between $k-$vertex subsets of a small ball.*

This follows directly from Lemma 12, which states that $A_0 \cap A_1 \cap \cdots \cap A_t = \emptyset$ with probability at least $1 - k^{-1/2+o(1)}$ Finally, we prove that A_t remains concentrated for all $t \geq t^*$, where t^* is the convergence time described in Theorem 6.

Proof [Abbreviated Proof of Theorem 7] Let $A \subset V$ with $|A| = k$, and let I be any ball surrounding $k - k^{2/3}$ points of A. Assume that $r = r(I) = \sigma k^{-1/3+\epsilon}$ and $I = B_r(p)$.

We consider 2 cases, with minor differences to the proof of Lemma 15 (whose cases are illustrated in Figure 6):

Case 1: There exists a half space dividing I into two spherical caps, R_1 and R_2, with heights $r/4$ and $7r/4$, such that $|R_2 \cap A| \leq k/(\ln k)^2$. In this case, the ball is "imbalanced" in the sense that one portion of of the ball contains the vast majority of the points.

Case 2: No such imbalanced partition of I exists. For all $x \in \partial I$, denoting $R_2 = \{z \in I : x - z \cdot (\frac{x-p}{\|x-p\|} > r/4)\}$, $|R_2 \cap A| \geq k/(\ln k)^2$. We show that for any z within $\Delta r = r/(\ln k)^3$. of the boundary of I, $\mathbb{P}(F(z;A) > C) < 1/n^3$.

We carefully construct two points, z and z' whose expected inputs from A, conditioned on the graph edges, are guaranteed to differ by a constant multiplicative fraction. E.g, let $Z = k - F(z;A)$; $\mathbb{E}Z \approx k\frac{2r^2}{\sigma^2}$ by the assumption that A is well spread within a ball of radius r. By the Chernoff Bound (Lemma 19), we can argue: $\mathbb{P}(Z < \mathbb{E}Z - \Theta(1)\mathbb{E}Z) \leq e^{-\Theta(1)\mathbb{E}Z} \leq e^{-\Theta(1)k^{1/3+2\epsilon}}$.

The probability that there exist $k^{2/3}$ points which violate the condition is at most:

$$\binom{n}{k^{2/3}}\left(e^{-\Theta(1)k^{1/3+2\epsilon}}\right)^{k^{2/3}} < e^{k^{2/3}\log n}e^{-\Theta(1)k^{1+2\epsilon}}$$

13

Since there are at most $\binom{n}{k} = O(e^{k \log n})$ possible k-subsets of V, this is true by the union bound for all subsets A with high probability.

∎

3. Conclusion and further questions

Plasticity. Our proof shows that plasticity is not necessary for the convergence of the k-cap mechanism. Previous analysis of this process on random graphs studied a variant of the problem where edges were given a weight, initially set to 1. If two neighboring vertices fired consecutively, the weight of their edge was boosted by a factor of $1+\beta$. In Erdős–Rényi random graphs, this weight proved to be vital for convergence; it allowed a set of vertices to become associated over time, causing them to fire together (Papadimitriou et al., 2020; Dabagia et al., 2021). We have shown that, given sufficiently local graph structure, it is possible for the process without plasticity to converge to a subset which is small compared to n. This possibly indicates two distinct mechanisms which drive the convergence of assembly projection in the brain (see Figure 1).

Further motivation from Neuroscience. As discussed above, the geometric model embeds the nodes of the graph as points in space, and it strongly prefers to connect nodes which are close to each other. Many real-world graphs have a spatial component and a cost associated with long-range connections, so the geometric graph model has theoretical guarantees which match empirical properties of graphs in many domains. One such property is the clustering coefficient, which measures the prevalence of cliques between the immediate neighborhood of the vertex (Boguná and Pastor-Satorras, 2003). In the graph model we have discussed thus far, the clustering coefficient is quite high; in fact, within a small neighborhood of any vertex the probability that the vertices form a clique is exponentially likely. In particular, high clustering between neurons has been observed in the brain (Song et al., 2005).

While the simplicity of the model makes the analysis tractable, there may be interesting algorithmic insights which can be gleaned by mimicking other empirically observed structures. One relevant property is the *power-law degree distribution* (Bullmore and Sporns, 2009). There are a small set of 'hub' neurons with very high degree (in the geometric random graph, the degree distribution is fairly uniform). One concrete question is whether, in a graph with a power-law degree distribution, the k-cap mechanism is likely to converge to a set of vertices with high degree.

Acknowledgments

References

David Aldous. *Probability approximations via the Poisson clumping heuristic*, volume 77. Springer Science & Business Media, 2013.

Marc Barthélemy. Spatial networks. *Physics reports*, 499(1-3):1–101, 2011.

Marián Boguná and Romualdo Pastor-Satorras. Class of correlated random networks with hidden variables. *Physical Review E*, 68(3):036112, 2003.

Marián Boguná, Romualdo Pastor-Satorras, Albert Díaz-Guilera, and Alex Arenas. Models of social networks based on social distance attachment. *Physical review E*, 70(5):056122, 2004.

Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198, 2009.

György Buzsáki. *The Brain from Inside Out*. Oxford University Press, 2019.

Hermann Cuntz, Friedrich Forstner, Alexander Borst, and Michael Häusser. One rule to grow them all: a general theory of neuronal branching and its practical application. *PLoS computational biology*, 6(8):e1000877, 2010.

Max Dabagia, Christos H Papadimitriou, and Santosh S Vempala. Assemblies of neurons can learn to classify well-separated distributions. *arXiv preprint arXiv:2110.03171*, 2021.

John N Darroch. On the distribution of the number of successes in independent trials. *The Annals of Mathematical Statistics*, 35(3):1317–1321, 1964.

Lutz Dümbgen and Jon A Wellner. The density ratio of poisson binomial versus poisson distributions. *Statistics & probability letters*, 165:108862, 2020.

Daniel Durstewitz, Jeremy K Seamans, and Terrence J Sejnowski. Neurocomputational models of working memory. *Nature neuroscience*, 3(11):1184–1191, 2000.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

Christos H Papadimitriou, Santosh S Vempala, Daniel Mitropolsky, Michael Collins, and Wolfgang Maass. Brain computation by assemblies of neurons. *Proceedings of the National Academy of Sciences*, 117(25):14464–14472, 2020.

Mathew Penrose. *Random geometric graphs*, volume 5. OUP Oxford, 2003.

Martin Raab and Angelika Steger. "balls into bins"—a simple and tight analysis. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 159–170. Springer, 1998.

S Ramón y Cajal. Histology of the nervous system of man and vertebrates (vols. 1, 2).(n. swanson & lw swanson, trans. 1995), 1911.

S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology*, 3(3):e68, 2005.

A Yu Volkova. A refinement of the central limit theorem for sums of independent random indicators. *Theory of Probability & Its Applications*, 40(4):791–794, 1996.

Appendix A. Probability Preliminaries

The following lemma relates to the distribution of uniform random points in $[0, 1]$. It will be referred to frequently throughout the proof.

Lemma 16 *All balls of radius $\sqrt{d/2}\left[\frac{6\log n}{n}\right]^{1/d}$ contain at least one vertex of G with probability at least $1 - (3n^2 \log n)^{-1}$.*

Proof Consider dividing $[0, 1]^d$ into $n/(3\log n)$ boxes with side length $[(3\log n)/n]^{1/d}$.

For any box, the probability that it receives no points of G is $(1 - (3\log n)/n)^n \leq e^{-3\log n} = n^{-3}$. There are $n/(3\log n)$ boxes, so by the union bound, the probability that there exists a box with no points of G is at most $(3n^2 \log n)^{-1}$.

A ball of radius $\sqrt{d/2}\left[\frac{6\log n}{n}\right]^{1/d}$ contains a box of side length $\frac{6\log n}{n}$. Any such box contains at least one box of the partition of $[0, 1]^d$. Thus, all balls of this radius contains a vertex of G. ∎

Lemma 17 (Balls into Bins (Raab and Steger, 1998)) *Suppose m balls are assigned uniformly at random to n bins, where $\frac{n}{polylog(n)} \leq m << n\log n$. Then, with probability $1 - o(1)$, the maximum load is at least:*

$$\frac{\ln n}{\ln \gamma}\left[1 + 0.9\frac{\ln^{(2)}\gamma}{\ln \gamma}\right]$$

where $\gamma = \frac{n\log n}{m}$.

Lemma 18 (Maximum Degree of Geometric Graph (Penrose, 2003), Theorem 6.10) *Let $X = \{x_1, \ldots, x_n\}$ be a set of n points chosen uniformly at random on $[0, 1]^d$. Define a graph $G(X; r)$ such that there exists an edge between x_i and x_j if $\|x_i - x_j\| \leq r$.*

Define a sequence of radii $(r_n)_n$. Let Δ_n be the maximum degree of G. Define $k_n = \frac{\log n}{\log(\log n/(nr_n^d))}$.

If $nr_n^d/\log n \to 0$ and $\log(1/(nr_n^d))/\log(n) \to 0$ as $n \to \infty$. Then:

$$\lim_{n\to\infty}\frac{\Delta_n}{k_n} = 1 \text{ in probability}$$

and

$$\liminf_{n\to\infty}\frac{\Delta_n}{k_n} \geq 1 \text{ Almost surely}$$

Next, the following three lemmas contain different tail bounds for the sums of independent indicators.

Lemma 19 (Chernoff Bound) *Let X be a sum of independent random indicators with mean μ. Then, for any $\delta \geq 0$:*

$$\mathbb{P}(X > (1+\delta)\mu) \leq \exp\left(-\frac{\mu\delta^2}{2+\delta}\right)$$

16

$$\mathbb{P}(X < (1 - \delta)\mu) \le \exp\left(-\frac{\mu\delta^2}{2}\right)$$

Lemma 20 *For any binomial random variable X with parameters k, p, we can bound the probability that it exceeds M for any $M > kp$:*

$$\mathbb{P}(X > M) \le \exp\left(-kD\left(\frac{M}{k} \parallel p\right)\right)$$

where $D\left(a \parallel p\right) = a\log\frac{a}{p} + (1 - a)\log\frac{1-a}{1-p}$.

Lemma 21 *Let $X = \sum_{i=1}^{k} I_i$ be the sum of k independent indicators with probabilities $\mathbb{P}(I_i) = p_i \in (0, 1)$. Let $\mu = \mathbb{E}X$, and let t_1, t_2 be integer values such that $t_1 \ge \lceil\mu\rceil$ and $t_2 > t_1$. Then,*

$$\frac{\mathbb{P}(X \ge t_1)}{\mathbb{P}(X \ge t_2)} > \exp\left(\frac{(t_2 - \lceil\mu\rceil)^2 - (t_1 - \lfloor\mu\rfloor)^2}{2t_2}\right) > \exp\left(\frac{(t_2 - t_1)^2}{t_2}\right)$$

Proof Duembgen et al (Dümbgen and Wellner, 2020) gives a bound on the ratio of two consecutive probabilities. For any c with $\mathbb{P}(X = c - 1) > 0$,

$$\frac{\mathbb{P}(X = c + 1)}{\mathbb{P}(X = c)} < \frac{c}{c + 1}\frac{\mathbb{P}(X = c)}{\mathbb{P}(X = c - 1)} \tag{1}$$

The mode of X is either at $\lfloor\mu\rfloor$, $\lceil\mu\rceil$, or is equally attained at both (Darroch, 1964). The probablity increases monotonically from $X = 0$ up to the mode(s) and then decreases montonically up to $X = k$.

So, we have for any integer $m > \mu$:

$$\frac{\mathbb{P}(X = m + 1)}{\mathbb{P}(X = m)} < 1 \tag{2}$$

Using Equation 1 and 2, for any integer $s \ge \lceil\mu\rceil$

$$\frac{\mathbb{P}(X = s + 1)}{\mathbb{P}(X = s)} < \frac{s}{s + 1}\frac{\mathbb{P}(X = s)}{\mathbb{P}(X = s - 1)} \qquad < \ldots < \frac{\lceil\mu\rceil + 1}{s + 1}\frac{\mathbb{P}(X = \lceil\mu\rceil + 1)}{\mathbb{P}(X = \lceil\mu\rceil)} < \frac{\lceil\mu\rceil + 1}{s + 1}$$

We can rewrite the ratio of $\mathbb{P}(X = t_2)$ and $\mathbb{P}(X = t_1)$ as the product of ratios with a difference of 1:

$$\frac{\mathbb{P}(X = t_2)}{\mathbb{P}(X = t_1)} = \frac{\mathbb{P}(X = t_2)}{\mathbb{P}(X = t_2 - 1)}\frac{\mathbb{P}(X = t_2 - 1)}{\mathbb{P}(X = t_2 - 2)}\cdots\frac{\mathbb{P}(X = t_1 + 1)}{\mathbb{P}(X = t_1)}$$

Substituting the bound above:

$$\frac{\mathbb{P}(X = t_2)}{\mathbb{P}(X = t_1)} < \frac{(\lceil\mu\rceil + 1)^{(t_2 - t_1)}}{t_2(t_2 - 1)(t_2 - 2)\ldots(t_1 + 1)} = \prod_{s=t_1+1}^{t_2}\left(1 - \frac{s - \lceil\mu\rceil - 1}{s}\right)$$

Using the approximation $1 - x \leq e^{-x}$, this is at most:

$$\leq \exp\left(-\sum_{s=t_1+1}^{t_2} \frac{s - \lceil \mu \rceil - 1}{s}\right) \leq \exp\left(-\frac{\sum_{s=t_1}^{t_2} s - \lceil \mu \rceil - 1}{t_2}\right)$$

Expanding the sum in the numerator:

$$= \exp\left(-\frac{\sum_{s=0}^{t_2 - \lceil \mu \rceil - 1} s - \sum_{s=0}^{t_1 - \lceil \mu \rceil - 1} s}{t_2}\right)$$

$$= \exp\left(-\frac{(t_2 - \lceil \mu \rceil - 1)(t_2 - \lceil \mu \rceil) - (t_1 - \lceil \mu \rceil - 1)(t_1 - \lceil \mu \rceil)}{2t_2}\right)$$

$$\leq \exp\left(-\frac{(t_2 - \lceil \mu \rceil)^2 - (t_1 - \lfloor \mu \rfloor)^2}{2t_2}\right)$$

This ratio decreases as t_1 increases and $t_2 - t_1$ remains constant. This means that, for any $i > t_2$, we have $\mathbb{P}(X = i) < \exp\left(-\frac{(t_2 - \lceil \mu \rceil)^2 - (t_1 - \lfloor \mu \rfloor)^2}{2t_2}\right) \mathbb{P}(X = i - (t_2 - t_1))$

$$\mathbb{P}(X \geq t_2) = \sum_{i=\mu+t_2\sqrt{\mu}}^{k} \mathbb{P}(X = i)$$

$$< \sum_{i=\mu+t_2\sqrt{\mu}}^{k} \exp\left(-\frac{(t_2 - \lceil \mu \rceil)^2 - (t_1 - \lfloor \mu \rfloor)^2}{2t_2}\right) \mathbb{P}(X = i - (t_2 - t_1))$$

$$\leq \exp\left(-\frac{(t_2 - \lceil \mu \rceil)^2 - (t_1 - \lfloor \mu \rfloor)^2}{2t_2}\right) \mathbb{P}(X \geq t_1)$$

Expanding $(t_2 - \lceil \mu \rceil)^2 - (t_1 - \lfloor \mu \rfloor)^2$, we get $t_2^2 - t_1^2 - 2\lceil \mu \rceil(t_2 - t_1) = (t_2 - t_1)(t_2 + t_1 - 2\lceil \mu \rceil)$. Since $t_2 > t_1 \geq \lceil \mu \rceil$ by assumption, this exceeds $2(t_2 - t_1)^2$. ∎

Appendix B. Deferred Proofs

B.1. Characterization of A_1

Proof [Proof of Lemma 10]

Let $r = \alpha\sigma\sqrt{\ln \ln k}$. Suppose $|B_r(x) \cap A_0| \leq \frac{3\ln k}{4\ln\ln k}$.

If $x \in A_1$, then $F_0(x) \geq C_0$ by definition. The bound on C_0 in Lemma 9 implies $F_0(x) \geq \frac{\ln k}{\ln\ln k}$. Hence, by assumption, $x \in A_1$ only if it achieves an input of $M = \frac{\ln k}{4\ln\ln k}$ from outside $B_r(x)$.

Conditioned on $|B_r(x) \cap A_0| = \lambda k$, the remaining $(1-\lambda)k$ points are distributed uniformly on $[0,1]^d \setminus B_r(x)$.

By definition, $\mathbb{P}(\mathbb{1}_{(x,Y)} \mid Y = y) = g(x,y) = \exp\left(-\|x-y\|^2/2\sigma^2\right)$. Let f_r be the conditional distribution function of $\|y - x\|$, which has support on $(r, \sqrt{d}]$ ($\sqrt{d}$ being the longest diagonal of the hypercube $[0,1]^d$). Define $\partial B_r(x) = \{y \in [0,1]^d : \|x-y\| = r\}$ to be the spherical shell of radius r around x.

$$f_r(\rho) = \frac{\text{Vol}(\partial B_\rho(x) \cap [0,1]^d)}{1 - \text{Vol}(B_r(x))}$$

The boundaries of the hypercube make f_r somewhat difficult to calculate. Therefore, we will ignore the boundaries and set $f_r(\rho) < \frac{\text{Vol}(\partial B_\rho(x))}{1-\text{Vol}(B_r(x))}$. Note that $\text{Vol}(\partial B_\rho(x)) = \frac{2\pi^{d/2}}{\Gamma(d/2)}\rho^{d-1}$.

$$\mathbb{E}[\mathbb{1}_{(x,y)} \mid y \notin B_r(x)] = \int_r^{\sqrt{d}} e^{-\rho^2/(2\sigma^2)} f_r(\rho)\, d\rho$$

$$\leq \frac{1}{1 - \text{Vol}(B_r(x))} \int_r^{\sqrt{d}} e^{-\rho^2/(2\sigma^2)} \text{Vol}(\partial B_\rho(x))\, d\rho$$

$$\leq 2\frac{2\pi^{d/2}}{\Gamma(d/2)} \int_r^{\infty} \rho^{d-1} e^{-\rho^2/(2\sigma^2)}\, d\rho$$

$$= \frac{4\pi^{d/2}\sigma^d}{\Gamma(d/2)} \int_{r/\sigma}^{\infty} z^{d-1} e^{-z^2/2}\, dz$$

This integral can be estimated by observing that for r sufficiently large, $z^{d-1}e^{-z^2/2}$ is decreasing on $[r/\sigma, \infty]$. Therefore,

$$\int_{r/\sigma}^{\infty} x^{d-1} e^{-x^2/2}\, dx \leq \left(\frac{\sigma}{r}\right)^{d-2} \int_{r/\sigma}^{\infty} x e^{-x^2/2}\, dz = \left(\frac{r}{\sigma}\right)^{d-2} e^{-r^2/(2\sigma^2)} \tag{3}$$

Returning to the original equation, for any $d \geq 1$:

$$\mathbb{E}[\mathbb{1}_{(x,y)} \mid y \notin B_r(x)] \leq \frac{4\pi^{d/2}\sigma^d}{\Gamma(d/2)} \left(\frac{r}{\sigma}\right)^{d-2} e^{-r^2/(2\sigma^2)} = \Theta(1)\sigma^d \left(\frac{r}{\sigma}\right)^{d-2} e^{-r^2/2\sigma^2}$$

Substituting $r = \sigma\alpha\sqrt{\ln\ln k}$, this equals $\Theta(1)\sigma^d(\ln k)^{-\alpha^2/2}(\sqrt{\ln\ln k})^{d-2}$. Again recalling $\sigma^d = \Theta(1/k)$,

$$\mathbb{E}[\mathbb{1}_{(x,y)} \mid y \notin B_r(x)] \leq p = O(1/k)(\ln k)^{-\alpha^2/2}(\alpha\sqrt{\ln\ln k})^{d-2}$$

. We can bound the distribution of $F_0(x)$ by a binomial with probability p. In particular, $\mathbb{P}(F_0(x) > C_0)$ is bounded above by $\mathbb{P}\left(B > \frac{\ln k}{4\ln\ln k}\right)$, where $B \sim Bin(k,p)$. This quantity can be tightly bounded using a well-known entropy bound for tail probabilities, Lemma 20.

$$\mathbb{P}\left(B > \frac{\ln k}{4\ln\ln k}\right) \leq \exp\left(-kD\left(\frac{\ln k}{4k\ln\ln k} \,\|\, p\right)\right)$$

Bounding the divergence term:

$$D\left(\frac{\ln k}{4k \ln \ln k} \,\|\, p\right) = \frac{\ln k}{4k \ln \ln k} \ln \frac{\ln k}{4kp \ln \ln k} + (1 - \frac{\ln k}{4k \ln \ln k}) \ln \frac{1 - \frac{\ln k}{4k \ln \ln k}}{1 - p}$$

$$\geq \frac{\ln k}{4k \ln \ln k} \ln \frac{\ln k}{4kp \ln \ln k} + p - \frac{\ln k}{4k \ln \ln k} \quad \text{using } \ln x \geq 1 - 1/x \; \forall x > 0$$

$$= \frac{\ln k}{4k \ln \ln k} \ln \frac{\ln k^{1+\alpha^2/2}}{O(1)(\ln \ln k)^{1+(d-2)/2}} + \Theta(\frac{1}{k})(\ln k)^{-\alpha^2/2}(\alpha\sqrt{\ln \ln k})^{d-2} - \frac{\ln k}{4k \ln \ln k}$$

$$\geq \frac{\ln k}{k}\left[\frac{1 + \alpha^2/2}{4} - \frac{1 + (d-2)/2}{4}\frac{\ln^{(3)} k}{\ln \ln k} - \frac{O(1)}{\ln \ln k}\right]$$

Plugging this into the original bound, $\mathbb{P}\left(B > \frac{\ln k}{4 \ln \ln k}\right) \leq k^{-\frac{1+\alpha^2/2}{4}+o(1)}$

Since $n = k^\beta$ by definition, we can choose $\alpha = \sqrt{24\beta}$. Then, $\mathbb{P}(F_0(x) > C_0) \leq \mathbb{P}\left(B > \frac{\ln k}{4 \ln \ln k}\right) \leq 1/n^3$ ∎

Proof [Proof of Theorem 5]

We apply Lemma 10 and take the union bound over all x in the graph to conclude the following: with probability, $1 - 1/n^2$, $x \in A_1$ only if the ball $B_r(x)$, where $r = \sigma\sqrt{24\beta \ln \ln k}$, contains more than $\frac{3 \ln k}{4 \ln \ln k}$ points. Since $1/n^2$ is summable, this is true almost surely.

Since the expected number of points of A_0 in $B_r(x)$ is $k * \text{Vol}(B_r(x)) = O((\ln \ln k)^{d/2})$, the number of such high density regions will be relatively small.

To argue this, we consider $d + 1$ overlapping partitions of $[0, 1]^d$ into boxes. Let $L = 2\sigma\sqrt{24\beta \ln \ln k}$. First, tile $[0, 1]^d$ with boxes of width L. Then, shift each interval by half its width in each dimension, leading to d alternate partitions of $[0, 1]^d$.

For any x, the ball $B_{\sigma\sqrt{24\beta \ln \ln k}}(x)$ must be fully contained in a box I_i in at least one partition for some index i. Consider the probability that a given box I_i contains $\frac{3 \ln k}{4 \ln \ln k}$ points of A_0.

For each point in A_0, the probability that it lands in I_i is $\text{Vol}(I_i) = (2\sigma\sqrt{24\beta \ln \ln k})^d = \Theta(1)(\ln \ln k)^{d/2}/k$. Therefore, the number of points in I_i is $|A_0 \cap I_i| \sim Bin(k, \text{Vol}(I_i))$.

Using a well-known entropy bound for the tail probabilities of the binomial distribution (Lemma 20), the probability that $|I_i \cap A_0|$ exceeds $3 \ln k/4 \ln \ln k$ is at most:

$$\mathbb{P}\left(|I_i \cap A_0| > \frac{3 \ln k}{4 \ln \ln k}\right) \leq \exp\left(-kD\left(\frac{3 \ln k}{4k \ln \ln k} \,\|\, \text{Vol}(I_i)\right)\right) \tag{4}$$

Bounding the divergence term (Using the inequality $\ln x \geq 1 - 1/x$):

$$D\left(\frac{3\ln k}{4k\ln\ln k} \;\|\; \mathrm{Vol}(I_i)\right) = \frac{3\ln k}{4k\ln\ln k}\ln\frac{\Theta(1)\ln k}{(\ln\ln k)^{1+d/2}} + \left(1 - \frac{3\ln k}{4k\ln\ln k}\right)\ln\frac{1 - \frac{3\ln k}{4k\ln\ln k}}{1 - \Theta(1)(\ln\ln k)^{d/2}/k}$$

$$\geq \frac{3\ln k}{4k} - \frac{O(1)\ln k\ln^{(3)}k}{k\ln\ln k} + \left(1 - \frac{3\ln k}{4k\ln\ln k}\right)\left(1 - \frac{1 - \Theta(1)(\ln\ln k)^{d/2}/k}{1 - \frac{3\ln k}{4k\ln\ln k}}\right)$$

$$\geq \frac{3\ln k}{4k} - \frac{O(1)\ln k\ln^{(3)}k}{k\ln\ln k} - \frac{3\ln k}{4k\ln\ln k} + \Theta(1/k)(\ln\ln k)^{d/2}$$

$$= \frac{1}{k}\left[\frac{3}{4}\ln k - \frac{O(1)\ln k\ln^{(3)}k}{\ln\ln k}\right]$$

Substituting back into equation 4:

$$\mathbb{P}\left(|I_i \cap A_0| > \ln k/\ln\ln k\right) \leq \exp\left(-\frac{3}{4}\ln k + \frac{O(1)\ln k\ln^{(3)}k}{\ln\ln k}\right) = k^{-\frac{3}{4}+\frac{O(1)\ln^{(3)}k}{\ln\ln k}}$$

There are $\frac{d+1}{\mathrm{Vol}(I_i)} = \Theta\left(\frac{k}{(\ln\ln k)^{d/2}}\right)$ such intervals. The size of $|B_i \cap A_0|$ for each partition can be thought of as the loads in a 'balls into bins' problem; thus, the number of points in non-overlapping boxes are negatively correlated.

With high probability, the number of such intervals with enough points is $k^{1/4+o(1)}$.

The loads of the bins I_i are invariant to permutation; therefore, the probability that two intervals within a distance of $2\sigma\sqrt{\ln n}$ have a large enough load is $o(1)$.

The same can be said for each shifted partition. Therefore, the bins which achieve high input are of size $\Theta(L) = \Theta(\sigma\sqrt{\ln\ln k})$ and separated by a distance of $2\sigma\sqrt{\ln n}$. ∎

B.2. Convergence of A_t

The following lemmas will be used to bound C_t at each time step. Using this, we can get precise bounds on $\mathbb{P}(F_t(x) > C_t)$.

Lemma 22 *For any vector direction v and point $x \in [0,1]^d$, $|\nabla_v E[x]| < \frac{k}{\sigma}\sqrt{d/e}$.*

Proof [Proof of Lemma 11]

For any $i \in \{0, 1, \ldots, d-1\}$:

$$\frac{\partial}{\partial x_i}E[x] = \sum_{z \in A_t}\frac{\partial}{\partial x_i}g(x,z) = \sum_{z \in A_t}-\frac{x_i - z_i}{\sigma^2}\exp\left(\frac{-\|x - z\|^2}{2\sigma^2}\right)$$

Let $\hat{x} = \frac{x-z}{\sigma}$. The maximum of $\left|\hat{x}\exp\left(\frac{-\hat{x}^2}{2}\right)\right|$ occurs at $e^{-1/2}$. Thus, we have

$$\left|\frac{\partial}{\partial x_i}E[x]\right| \leq \sum_{z \in A_t}\frac{|x_i - z_i|}{\sigma^2}\exp\left(\frac{-\|x - z\|^2}{2\sigma^2}\right) < \sum_{z \in A_t}\frac{1}{\sigma}e^{-1/2} = \frac{k}{\sigma}e^{-1/2}$$

For any unit vector v:

$$|\nabla_v E[x]| < \frac{k}{\sigma} e^{-1/2} v \cdot \mathbf{1} = \sqrt{d} e^{-1/2} \frac{k}{\sigma}$$

∎

For this proof to be viable, we will need to show that $F_t(x)$ is the sum of independent indicators. By definition of the graph structure, each edge $\mathbb{1}_{(y,x)}$ is drawn independently. However, we will also need to show that, for each $y \in A_t$, its edges $\mathbb{1}_{(y,x)}$ have not been used in previous computations. This follows from the next lemma.

Lemma 23 *Suppose $t = O((\ln k)^c)$ for a constant c. Then, with probability at least $1 - \frac{1}{k^{1/2-o(1)}}$*

$$A_0 \cap A_1 \cap \cdots \cap A_t = \emptyset$$

Proof Suppose at time s, $\{A_0, A_1 \ldots A_s\}$ are pairwise disjoint. Therefore, at time s, the edges $\{\mathbb{1}_{(y,x)} : y \in A_s, x \in [n]\}$ have not been examined by the k-cap function, and they are conditionally independent.

Now, we will compute the probability that $|A_{s+1} \cap A_i| > 0$ for some $i \leq s$. By Lemma 16, there are at least $\Theta(1)\sigma^d k^{-d/2} n / \log n = \Theta(k^{\beta-d/2-1}/\log n)$ points within $\sigma k^{-1/2}$ of x. By Lemma 11, for all $z \in B_{\sigma k^{-1/2}}(x)$, the expected input to z is not too far from the expected input to x; $\mathbb{E}F_s(z) > \mathbb{E}F_s(x) - (ek)^{1/2}$. For such a z, $\mathbb{P}(F_s(z) > C_{s+1})$ differs from $\mathbb{P}(F_s(x) > C_{s+1})$ by at most a constant factor. Thus, the probability that any given x is chosen is $p_s(x) < k^{-\beta+d/2+1+o(1)} < k^{-3/2+o(1)}$ by the definition of β. There are sk points in $A_0 \cup A_1 \cup \cdots \cup A_s$, so the probability that any given $y \in A_0 \cup A_1 \cup \cdots \cup A_s$ is in A_{s+1} is at most $sk * k^{-3/2+o(1)} = O(k^{-1/2+o(1)})$. Therefore, the probability that $(A_0 \cup A_1 \cup \cdots \cup A_s) \cap A_{s+1} = \emptyset$ is at least $1 - \frac{1}{k^{1/2-o(1)}}$.

The probability that this holds for all $s < t$ is $(1 - \frac{1}{k^{1/2-o(1)}})^t \approx 1 - \frac{t}{k^{1/2-o(1)}}$. Since t is polylog(k), this is at least $1 - \frac{1}{k^{1/2-o(1)}}$. ∎

Lemma 12 implies that, conditioned on the set A_t, $F_t(x) = \sum_{z \in A_t} \mathbb{1}_{(z,x)}$ is a sum of independent indicators (with no dependence on previous time steps). Therefore, C_t can be bounded using standard concentration bounds as follows:

Lemma 24 *At any step $t = O((\log k)^c)$, assuming the conditions of Lemma 15, with high probability, $C_t \geq \max_x E[x]$*

Proof [Proof of Lemma 24]

By Lemma 16, for any point x, there are $\Omega(n \cdot (\sigma k^{-1} \log n)^d)$ points of V in a radius of $\sigma k^{-1} \log n$ of x. By the assumption that $n \geq k^{2+d}$, this is $\Omega(k \log n)$.

For any $y \in B_{\sigma k^{-1} \log n}(x)$, Lemma 11 implies:

$$E[y] > E[x] - \sqrt{d/e} \log n$$

Therefore, if $C_t = E[x]$, then there are $\tilde{\Omega}(k)$ points where $E[y] > C_t - O(\log n)$. Here we will use Lemma 12, which tells us that each $F_t(y)$ is independent conditioned on A_t. Hence, Chernoff type bounds apply; if $\mathbb{P}(F_t(y) > E[y] + O(\log n)) = \Theta(1)$, then with high probability there are k points that exceed C_t.

Using the loose bound given in (Volkova, 1996), we can bound $\mathbb{P}(F_t(x) > C_t)$ using the CDF of the normal distribution. For any sum of independent indicators S with mean μ and variance σ, the CDF can be approximated as follows:

$$\sup_m \left| \mathbb{P}(S \leq m) - G\left(\frac{m + 1/2 - \mu}{\sigma} \right) \right| \leq \frac{\sigma + 3}{4\sigma^3} < \frac{1}{\sigma^2}$$

Where $G(x) = \Phi(x) - \frac{\gamma}{6\sigma^3}(x^2 - 1)\frac{e^{-x^2/2}}{\sqrt{2\pi}}$, and $\gamma = \mathbb{E}[(S - \mu)^3]$ is the skewness. This holds for any $\sigma \geq 10$.

We can assume that the variance of y, $V[y]^2$, exceeds $(\log k)^2$; otherwise, $E[y] = \sum_{y \in A_t} = k(1 - o(1/k))$, so we can assume that $C_t = k$.

Fix $y \in B_{\sigma k^{-1} \log n}(x)$. From the above equation, we find that for any $t > 0$:

$$\mathbb{P}\left(F_t(y) > E[y] + tV[y] - 1/2 \right) > 1 - \left[G(t) + \frac{1}{V[y]^2} \right]$$

Substituting the value of G:

$$\mathbb{P}\left(F_t(y) > E[y] + tV[y] - 1/2 \right) > 1 - \Phi(t) + \frac{\gamma(t^2 - 1)}{6\sqrt{2\pi}V[y]^3} e^{-t^2/2} - \frac{1}{V[y]^2}$$

Here, we will make two approximations. First, the exact value of γ is $\sum_{z \in A_t} g(y, z)(1 - g(y, z))(1 - 2g(y, z))$. Therefore, $\gamma > -V[y]^2$, so $\gamma(t^2 - 1) > -V[y]^2 t^2$. Second, we will substitute the lower tail bound for $1 - \Phi(t)$; $1 - \Phi(t) \geq \frac{1}{\sqrt{2\pi}}\left(t^{-1} - t^{-3} \right) e^{-t^2/2} \geq \frac{1}{t\sqrt{8\pi}} e^{-t^2/2}$ for $t \geq 2$.

This leaves us with:

$$\mathbb{P}\left(F_t(y) > C_t = E[y] + tV[y] - 1/2 \right) > \frac{1}{t\sqrt{8\pi}} e^{-t^2/2} - \frac{t^2}{6\sqrt{2\pi}V[x]} e^{-t^2/2} - \frac{1}{V[y]^2}$$

Setting $t = O(1)$, this occurs with constant positive probability. ∎

Finally, we can use the above lemma to relate the probability that a point fires at time $t+1$ to its expected value at time t.

Lemma 25 *Let $y \in I_j$, and $\hat{k} = |I_j \cap A_t|$. If there exists an $x \in I_j$ such that $\mathbb{E}F_t(x) > \mathbb{E}F_t(y) + \sqrt{6\beta(\hat{k} - \mathbb{E}F_t(y))\ln k}$, then $\mathbb{P}(F_t(y) > C_t) < \frac{1}{n^3}$*

Proof

Let $X = \hat{k} - F_t(y)$. By Lemma 19, $\mathbb{P}(X < (1 - \epsilon)\mathbb{E}X) \leq \exp\left(-\epsilon^2 \mathbb{E}X/2 \right)$.

Thus, setting $\epsilon = \frac{C - F_t(y)}{\mathbb{E}X}$, we have

$$\mathbb{P}(F_t(y) > C) = \mathbb{P}(X < \mathbb{E}X - (C - \mathbb{E}F_t(y)) \leq \mathbb{P}(X < \mathbb{E}X(1 - \epsilon)) \leq \exp\left(-\frac{\epsilon^2 \mathbb{E}X}{2} \right)$$

By Lemma 13, $C_t \geq \mathbb{E}F_t(x)$ for all x. Hence, by the assumption, $C - \mathbb{E}F_t(y) \geq \mathbb{E}F_t(x) - \mathbb{E}F_t(y) > \sqrt{6\beta \mathbb{E}X \ln k}$. Substituting this value for $\epsilon \mathbb{E}X$,

$$\mathbb{P}(F_t(y) > C) \leq \exp\left(-\frac{6\beta \ln k}{2}\right) = k^{-3\beta} = n^{-3}$$

∎

Now, we are ready to prove Lemma 15. This lemma will show that the radius of each ball shrinks at each step; that is A_t is contained within a union of balls of radius r_t, where r_t is a decreasing function of t. The main idea of the proof is to show that, regardless of the actual positions of points in $A_t \cap I$, vertices toward the center of I have a small advantage over vertices toward the edge. Thus, either (1) the position of points in $A_t \cap I$ is particularly unbalanced, and A_{t+1} shifts toward one side, or (2), the radius of I shrinks in all directions.

Proof [Proof of Lemma 15]

Fix one ball $I = B_r(p)$. Let r be the radius of I and p be its center.

To assist with the proof, we will define the following values. Let $\hat{k} = |A_t \cap I|$; we can assume that $\hat{k} > k^{3/4-o(1)}$, since an interval with asymptotically fewer points will be eliminated at the next step. Define $dist(I, z) = \min_{y \in I} \|y - z\|$. Finally, for any set $S \subset [0,1]^d$, let $F_t(z; S) = \sum_{y \in A_t \cap S} \mathbb{1}_{(y,z)}$.

We will prove that with high probability, $\{x \in [0,1]^d : F_t(x; I) \geq C_t\}$ can be covered by $I' = B_{r'}(p')$, where $r' = (1 - 1/(\log k)^c)r$ and $\max_{z \in I'} dist(I, z) < 5(r(I) - r(I'))$. Call this statement (*).

If statement (*) holds for each $I = I_i$, the lemma is proven. This holds by the separation assumption; for x near I, $F_t(x; I)$ is a good proxy for $F_t(x)$. In particular, if $dist(I, x) < \sigma\sqrt{\ln n}$, then $F_t(x; I) = F_t(x) - o(1)$ with high probability.

To prove this statement, we will consider two cases. In case 1, we suppose that the distribution of $A_t \cap I$ is imbalanced. In particular, there exists a half space dividing I into two spherical caps, with heights $r/4$ and $7r/4$, such that the larger segment contains only $\hat{k}/(\ln k)^\alpha$ points of A_{t+1} (for an $\alpha \geq 1$). See Figure 6. We will show that given this imbalance, statement (*) holds.

In case 2, no such division exists. We will show that for any point z near the boundary of I is disadvantaged compared to a point near the center. Thus, $I' = B_{r'}(p)$ for an $r' = (1 - 1/(\log k)^c)r$. See Figure 6.

For both cases, the argument will use a bound on the gradient of $E[z]$. With this, we will construct a point w such that $E[w] - E[z]$ is large, and use Lemma 14 to argue that $\mathbb{P}(z \in A_{t+1}) < 1/n^3$.

Consider two cases:

Case 1: there exists a half space dividing I into two spherical caps, R_1 and R_2, with heights $r/4$ and $7r/4$, such that $|R_2 \cap A_t| \leq \hat{k}/(\ln k)^\alpha$, where $\alpha = 1 + \max(2r^2/(\sigma^2 \ln \ln k), 1)$. In this case, the ball is "imbalanced" in the sense that one portion of of the ball contains the vast majority of the points.

Without loss of generality, let $p = [r, 0, 0, \ldots, 0]$, $R_1 = \{y \in I : y_1 \leq r/4\}$, and $R_2 = \{y \in I : y_1 > r/4\}$ (as illustrated in Figure 6). Let $z = [z_1, z_2, \ldots, z_n]$ where $z_1 \geq 3r/8$ and $dist(I, z) = O(\min\{r, \sigma\hat{k}^{-1/5}\})$.

Then we can bound the derivative with respect to the first coordinate:

$$\frac{\partial}{\partial z_1}\mathbb{E}F_t(z;I) = \sum_{y \in A_t \cap I} \frac{\partial}{\partial z_1}g(y,z) = \sum_{y \in A_t \cap I} -\frac{z_1 - y_1}{\sigma^2}g(y,z) \tag{5}$$

$$= \sum_{y \in R_1 \cap A_t} -\frac{z_1 - y_1}{\sigma^2}g(y,z) + \sum_{y \in R_2 \cap A_t} -\frac{z_1 - y_1}{\sigma^2}g(y,z) \tag{6}$$

The partial derivative $\frac{\partial}{\partial z_1}g(y,z)$ is minimized at $z_1 - y_1 = \sigma$ and maximized at $z_1 - y_1 = -\sigma$. The lower bound on the derivative depends on r as follows:

- If $2r \geq \sigma$:

  $\min_{y \in R_2} \frac{z_1 - y_1}{\sigma^2}g(y,z) > -\frac{1}{\sigma}e^{-1/2}$, and

  $\min_{y \in R_1} \frac{z_1 - y_1}{\sigma^2}g(y,z) > \frac{r}{8\sigma^2}\exp\left(\frac{-(2r + \min\{r,\sigma\hat{k}^{-1/5}\})^2}{2\sigma^2}\right) = \frac{r}{8\sigma^2}\exp\left(\frac{-2r^2 - o(\sigma^2)}{\sigma^2}\right)$

- If $2r < \sigma$,

  $\min_{y \in R_2} \frac{z_1 - y_1}{\sigma^2}g(y,z) > -\frac{2r}{\sigma^2}\exp\left(-2r^2/\sigma^2\right)$, and

  $\min_{y \in R_1} \frac{z_1 - y_1}{\sigma^2}g(y,z) > \frac{r}{8\sigma^2}\exp\left(\frac{-(2r + O(\min\{r,\sigma\hat{k}^{-1/5}\})^2}{2\sigma^2}\right) = \frac{r}{8\sigma^2}\exp\left(\frac{-O(1)r^2}{2\sigma^2}\right)$

Returning to Equation 11,

$$\frac{\partial}{\partial z_1}\mathbb{E}F_t(z;I) \leq -|A_t \cap R_1|\min_{y \in R_1}\frac{z_1 - y_1}{\sigma^2}g(y,z) + |A_t \cap R_2|\max_{y \in R_2}\frac{y_1 - z_1}{\sigma^2}g(y,z) \tag{7}$$

By assumption, $|A_t \cap R_2| \leq \hat{k}/(\ln k)^\alpha$. Replacing this:

- If $2r \geq \sigma$:

$$\frac{\partial}{\partial z_1}\mathbb{E}F_t(z;I) \leq -\hat{k}(1 - o(1))\frac{r}{8\sigma^2}\exp\left(\frac{-2r^2}{\sigma^2} - o(1)\right) + \frac{\hat{k}}{(\ln k)^\alpha}\frac{1}{\sigma}e^{-1/2}$$

  By the definition of α, $(\ln k)^{-\alpha} = e^{-\alpha \ln \ln k} \leq \frac{1}{\ln k}e^{-2r^2/\sigma^2}$. Hence,

$$\frac{\partial}{\partial z_1}E[z] \leq -\frac{\hat{k}}{\sigma}\left[(1 - o(1))\frac{r}{8\sigma}\exp\left(\frac{-2r^2}{\sigma^2}\right) - \frac{1}{\ln k}\exp\left(\frac{-2r^2}{\sigma^2}\right)e^{-1/2}\right]$$

$$\frac{\partial}{\partial z_1}\mathbb{E}F_t(z;I) \leq -\frac{\hat{k}r}{8\sigma^2}\exp\left(\frac{-2r^2}{\sigma^2}\right)(1 - o(1))$$

- If $2r < \sigma$,

$$\frac{\partial}{\partial z_1}\mathbb{E}F_t(z;I) \leq -\frac{r}{8\sigma^2}e^{-O(1)r^2/\sigma^2}\hat{k} + \frac{2r}{\sigma^2}\frac{\hat{k}}{\ln k} = -\Theta(1)\frac{\hat{k}r}{\sigma^2}$$

25

Let $z' = [z'_1, z'_2, \ldots, z'_n]$ where $z'_1 \geq r/2$ and $dist(I, z') < \min\{r/8, \sigma\hat{k}^{-1/5}\}$. Consider the point $w = z' - [\min\{r/8, \sigma\hat{k}^{-1/5}\}, 0, 0, \ldots, 0]$. By definition, the derivative bounds above apply for all points on the line between w and z'. This gives us a lower bound on $E[w] - E[z']$. While $w \notin V$ almost surely, by Lemma 16 there exists a point $w' \in V$ within a radius of $O((\log n/n)^{1/d})$ of w. Applying 11, $E[w'] - E[z'] > E[w] - E[z'] - o(1)$. Then, we will apply Lemma 14 to show that $\mathbb{P}(z' \in A_{t+1}) < 1/n^3$.

The condition of Lemma 14 holds if

$$E[w] - E[z'] \geq \sqrt{6\beta(\hat{k} - E[z])\ln k} \tag{8}$$

- If $2r \geq \sigma$:

$$E[w] \geq E[z'] + \sigma\hat{k}^{-1/5} \cdot \frac{\hat{k}r}{8\sigma^2} \exp\left(\frac{-2r^2}{\sigma^2}\right)(1 - o(1))$$

Since $r = O(\sigma\sqrt{\ln\ln k})$, $\exp\left(2r^2/\sigma^2\right) = \tilde{O}(1)$. Thus, $E[w] - E[z] = \tilde{\Omega}(\hat{k}^{4/5})$. Clearly this exceeds $\sqrt{6\beta\hat{k}\ln k}$, so by Lemma 14, $\mathbb{P}(z' \in A_{t+1}) < 1/n^3$.

- If $20\sigma\hat{k}^{-1/5} < 2r < \sigma$:

For the same reasons as above, we can obtain a similar bound:

$$E[w] \geq E[z'] + \sigma\hat{k}^{-1/5} \cdot \hat{k}\frac{r}{6\sigma^2}e^{-2r^2/\sigma^2}(1 - o(1)) = E[z] + \hat{k}^{4/5-o(1)}\frac{r}{\sigma}$$

Since $r = \Omega(\sigma\hat{k}^{-1/5})$, this exceeds $\sqrt{6\beta\hat{k}\ln k}$, so by Lemma 14, $\mathbb{P}(z' \in A_{t+1}) < 1/n^3$.

- If $2r \leq 20\sigma\hat{k}^{-1/5}$:

$$E[w] \geq E[z'] + \frac{r}{8} \cdot \hat{k}\frac{r}{8\sigma^2}e^{-2r^2/\sigma^2}(1 - o(1)) = E[z] + \Theta(1)\hat{k}\frac{r^2}{\sigma^2}$$

In this case, we can bound $\hat{k} - E[z]$;

$$\hat{k} - E[z] \leq \hat{k}(1 - e^{-2r^2/\sigma^2}) \leq \hat{k}\frac{2r^2}{\sigma^2}$$

Therefore, the condition can be bounded: $\sqrt{6\beta(\hat{k} - E[z])\ln k} \leq \frac{r}{\sigma}\sqrt{12\beta\hat{k}\ln k}$.

There exists a constant C such that for $10\sigma\hat{k}^{-1/5} > r > C\sigma\sqrt{\ln k/\hat{k}}$, $E[w] - E[z] = \Theta(1)\hat{k}\frac{r^2}{\sigma^2} > \frac{r}{\sigma}\sqrt{12\beta\hat{k}\ln k}$. By Lemma 14, $\mathbb{P}(z' \in A_{t+1}) < 1/n^3$.

Finally, we will argue that for any z with $dist(I, z) > r/20$, $\mathbb{P}(z \in A_{t+1}) < 1/n^3$. Let u be the unit vector parallel to $z - p$:

$$\nabla_u E[z] = \sum_{y \in A_t} \nabla_u g(y, z) = \sum_{y \in A_t} \left(u \cdot \frac{y - z}{\sigma^2} \right) g(y, z)$$

$$\geq \frac{dist(I, z)}{\sigma^2} \sum_{y \in A_t} g(y, z)$$

$$= \frac{dist(I, z)}{\sigma^2} E[z]$$

Let w be a point along the line $z - p$, with $dist(I, w) = dist(I, z)/2$. Again, while $w \notin V$ almost surely, by Lemma 16 there exists a point $w' \in V$ within a radius of $O((\log n/n)^{1/d})$ of w. Applying 11, $E[w'] - E[z'] > E[w] - E[z'] - o(1)$. Dividing this again into two cases:

- If $r \geq 2\sigma$:

  There exists a point y in I with $E[y] = \Omega(\hat{k}/(\ln \ln k)^{d/2})$. This is due to the pigeonhole principle; the volume of I is $\Theta(r^d) = O((\ln \ln k)^{d/2}/\hat{k})$. Therefore, there exists a smaller ball of radius σ in I with $\hat{k}/(\ln \ln k)^{d/2}$ points. For y in this smaller ball, $E[y] = \Omega(\hat{k}/(\ln \ln k)^{d/2})$.

  If $E[z] = \tilde{\Omega}(\hat{k})$, then $E[w] - E[z] > dist(I, z)^2/(2\sigma)^2 E[z] = \tilde{\Omega}(\hat{k})$, and by Lemma 14, $\mathbb{P}(z \in A_{t+1}) < 1/n^3$. Otherwise, $E[y] - E[z] = \tilde{\Omega}(\hat{k})$, and again by Lemma 14, $\mathbb{P}(z \in A_{t+1}) < 1/n^3$.

- If $r < 2\sigma$:

  There exists a point y in R_1 with $E[y] \geq e^{-r^2/8\sigma^2} \hat{k} \geq \hat{k}(1 - r^2/8\sigma^2)$.

  If $E[z] = \hat{k}(1 - \gamma r^2/\sigma^2)$, the bound for Lemma 14 is:

  $$\sqrt{6\beta(\hat{k} - E[z]) \ln k} = \frac{r}{\sigma} \sqrt{6\beta\gamma \ln k}$$

  For any $r = \Omega(\sigma\sqrt{\ln k/\hat{k}})$, $E[w] - E[z] \geq \Theta(r^2/\sigma^2 \hat{k}) = \Omega(\ln k)$. There exists a constant C such that for any $r > C\sigma\sqrt{\ln k/\hat{k}}$, this exceeds the bound of Lemma 14, and $\mathbb{P}(z \in A_{t+1}) < 1/n^3$.

In conclusion, we have determined that the set of points $z \in I$ such that $\mathbb{P}(z \in A_{t+1}) > 1/n^3$ are contained within a region $R = \{z : z_1 \leq r/2, dist(I, z) < r/20\}$. The radius of R is $r/20$ plus the width of $\{z \in I : z_1 \leq r/2\}$. Using a geometric argument, this set has width $\sqrt{R^2 - (R/2)^2} = R\sqrt{3}/2$. This region can be enclosed by a ball I' defined as follows (illustrated as an orange circle in Figure 6):

Let $I' = B_{19r/20}(p')$ for $p' = [3r/4, 0, 0, \ldots, 0]$ and $r(I') = \frac{19}{20} r(I)$. It is simple to check that $B_{19r/20}(p')$ contains R;

$$\max_{z \in R} \|p - z\| = r\sqrt{(\sqrt{3}/2 + 1/20)^2 + 1/4^2} < 19r/20$$

27

Additionally, $\max_{z \in B_{19r/20}(p')} dist(z, I) < \max_{z \in B_r(p')} dist(z, I) < r/4$. Therefore, $d(I, I') < r/4 < 5(r(I) - r(I'))$.

Case 2: In this section, we assume that no such imbalanced partition of I exists. For all $x \in \partial I$, denoting $R_2 = \{z \in I : (x - z) \cdot \frac{x-p}{\|x-p\|} > r/4\}$, $|R_2 \cap A_t| \geq \hat{k}/(\ln k)^\alpha$, where $\alpha = 1 + \max\{2r^2/(\sigma^2 \ln \ln k), 1\}$. We will show that for any z within Δr of the boundary of I, $\mathbb{P}(z \in A_{t+1} < 1/n^3)$.

Fix x, and assume without loss of generality that $p = [r, 0, 0, \ldots, 0]$ and $x = [0, 0, \ldots, 0]$

Let $z = [z_1, 0, 0, \ldots, 0]$ where $-\sigma \log n < z_1 < \Delta r = O(r/\ln k)$. Then, $\max_{y \in R_1} \frac{z_1 - y_1}{\sigma^2} g(y, z) = \frac{z_1}{\sigma^2} \exp\left(\frac{-z_1^2}{2\sigma^2}\right) < \frac{\Delta r}{\sigma^2} \exp\left(\frac{-(\Delta r)^2}{2\sigma^2}\right)$ (Note that by construction $\Delta r < \sigma$). Also, separately taking the minima of $y_1 - z_1$ and $g(y, z)$, $\min_{y \in R_2} \frac{y_1 - z_1}{\sigma^2} g(y, z) > \frac{r/4 - z_1}{\sigma^2} \exp\left(\frac{-2r^2}{\sigma^2}\right)$. Returning to Equation 11:

$$\frac{\partial}{\partial z_1} E[z] \geq -|A_t \cap R_1| \frac{\Delta r}{\sigma^2} \exp\left(\frac{-(\Delta r)^2}{2\sigma^2}\right) + |A_t \cap R_2| \frac{r/4 - z_1}{\sigma^2} \exp\left(\frac{-2r^2}{\sigma^2}\right) \qquad (9)$$

By assumption, $|A_t \cap R_2| \geq \hat{k}/(\ln k)^\alpha$. Replacing this:

$$\frac{\partial}{\partial z_1} E[z] \geq -\hat{k} \left[1 - \frac{1}{(\ln k)^\alpha}\right] \frac{\Delta r}{\sigma^2} \exp\left(\frac{-(\Delta r)^2}{2\sigma^2}\right) + \frac{\hat{k}}{(\ln k)^\alpha} \frac{r/4 - \Delta r}{\sigma^2} \exp\left(\frac{-2r^2}{\sigma^2}\right)$$

Define $\Delta r = r/(\ln k)^{2\alpha}$. Again, note that $(\ln k)^{-\alpha} = e^{-\alpha \ln \ln k} \leq \frac{1}{\ln k} e^{-2r^2/\sigma^2}$. So, $e^{-2r^2/\sigma^2} \geq (\ln k)^{1-\alpha}$

$$\frac{\partial}{\partial z_1} E[z] \geq -\hat{k} \left[1 - \frac{1}{(\ln k)^\alpha}\right] \frac{r}{\sigma^2 (\ln k)^{2\alpha}} \exp\left(\frac{-(\Delta r)^2}{2\sigma^2}\right) + \frac{\hat{k}}{(\ln k)^{2\alpha-1}} \frac{r}{5\sigma^2}$$

$$\frac{\partial}{\partial z_1} E[z] \geq \frac{\hat{k}r}{\sigma^2 (\ln k)^{2\alpha}} \left[1 - o(1) + \ln k\right]$$

Suppose $z = [z_1, 0, 0, \ldots, 0]$ where $z_1 < \Delta r/2$. Let $w = z_1 + [\Delta r/2, 0, 0, \ldots, 0]$. Using the lower bound on the derivative,

$$E[w] \geq E[z] + \frac{\Delta r}{2} \cdot \frac{\hat{k}r}{\sigma^2 (\ln k)^{2\alpha}} \left[1 - o(1) + \ln k\right] \geq E[z] + \frac{1}{2} \frac{\hat{k}r^2}{\sigma^2} \ln k$$

For $r = \Omega(\sigma \hat{k}^{-1/4})$, this exceeds $\sqrt{6\beta \hat{k} \ln k}$, so by Lemma 14, $\mathbb{P}(z \in A_{t+1}) < 1/n^3$.

For $r = o(\sigma \hat{k}^{-1/4})$, we can bound $\hat{k} - E[z]$;

$$\hat{k} - E[z] \leq \hat{k}(1 - e^{-2r^2/\sigma^2}) \leq \hat{k} \frac{2r^2}{\sigma^2}$$

Therefore, the condition can be bounded: $\sqrt{6\beta(\hat{k} - E[z]) \ln k} \leq \frac{r}{\sigma} \sqrt{12\beta \hat{k} \ln k}$.

Then, for $\sigma \hat{k}^{-1/4} \ln k > r > \sigma \sqrt{\ln k/\hat{k}}$, $E[w] - E[z] = \frac{1}{2} \frac{\hat{k}r^2}{\sigma^2} \ln k > \frac{r}{\sigma} \sqrt{12\beta \hat{k} \ln k}$. By Lemma 14, $\mathbb{P}(z' \in A_{t+1}) < 1/n^3$.

In summary, there are two cases: in **case 1**, there exists a partition of I such that the vast majority of A_t is located in R_1. In this case, we have shown that $\{z : \mathbb{P}(z \in A_{t+1}) > 1/n^3\} \subset \{z : z_1 \leq r/2\}$. A symmetric argument showed that for any z such that the distance from z to I is at most $r/20$, $\mathbb{P}(z \in A_{t+1}) < 1/n^3$ Therefore, A_{t+1} is contained within a ball $I' = B_{19r/20}(p')$, where $d(I, I') < r/4$

In **case 2**, for all $x \in \partial I$, $|A_t \cap R_2|$ is sufficiently large. In this case, We have shown that for all $z = x + \lambda \frac{p-x}{\|p-x\|}$ where $\lambda < r/\text{polylog}(k)$, $\mathbb{P}(z \in A_{t+1}) < 1/n^3$. This applies for all $x \in \delta I$. Therefore, $I' \subset B_{r(1-1/(\ln k)^c)}(p)$.

By the assumption that each ball is sufficiently separated, we can conclude that, with high probability, $A_{t+1} \subset I_1' \cup I_2' \cup \cdots \cup I_i'$, where the radius of I_j' is smaller than the radius of I_j by at least a factor of $1/\text{polylog}(k)$.

■

Now, we are ready to prove Theorem 6.
Proof [Proof of Theorem 6] By Theorem 5, the conditions of Lemma 15 hold at step 1. Additionally, by the condition that each ball does not shift by more than $5(r(I) - r(I'))$ at each step, the separation condition holds inductively for any $t = \text{polylog}(k)$. The maximum distance moved by a single ball by time t is $5(r(I_j^{(0)}) - r(I_j^{(t)})) = O(\sigma\sqrt{\ln \ln k})$, which maintains the separation of $2(1-o(1))\sigma\sqrt{\ln n}$. Thus, we can apply the Lemma inductively.

By Lemma 15, the radius of I is reduced by a factor of $1 - \frac{1}{\text{polylog}(k)}$ in a single step; thus, to reach $O(\sigma \hat{k}^{-1/2}\sqrt{\ln k})$, the number of steps required is $\text{polylog}(k)$.

This shows that in $\text{polylog}(k)$ time, the radius of each sufficiently separated ball will be reduced to at most $\hat{k}^{-1/2}\sqrt{\ln k}$. Recall that there are $k^{1/4+o(1)}$ separated balls; a similar method will allow us to eliminate balls that are $\sigma\hat{k}^{-1/2}\sqrt{\ln k}$ in size.

If the number of balls is greater than 1, $|I_1 \cap A_t|$ can fall anywhere in the range $M \pm \sqrt{M}$ with constant probability, where $M = \mathbb{E}|I_1 \cap A_t|$. By the pigeonhole principle, at least one ball receives $k^{3/4-o(1)}$ points. Since the size of the balls are at most $\sigma\hat{k}^{-1/2}\sqrt{\ln k} < \sigma k^{-3/8}\sqrt{\ln k}$, $C_t \geq (k^{3/4-o(1)})$. Therefore, for each j, if I_j is not eliminated, there exists an $x \in I_j$ such that $\mathbb{E}F_t(x) \geq k^{3/4-o(1)}$. Consider two alternative scenarios, which can each occur with constant probability.

$$\begin{cases} (1) & \max_{x \in I_1} \mathbb{E}F_t(x) = X - \Theta(\sqrt{X}) \\ (2) & \max_{x \in I_1} \mathbb{E}F_t(x) = X + \Theta(\sqrt{X}) \end{cases}$$

Let $y = \text{argmax}_{y \in I_2} \mathbb{E}F_t(y)$. So, it is clear that in either scenario (1) or (2), the inputs to x and y differ by the number of points added to I_1.

$$\left| \max_{x \in I_1} \mathbb{E}F_t(x) - \max_{y \in I_2} \mathbb{E}F_t(y) \right| = k^{3/8-o(1)}$$

In scenario 2, x receives an extra input of $\Theta\left(\sqrt{X}\right)$. The increased input in this scenario could affect C_t; however, either $\mathbb{E}F_t(x)$ becomes closer to C_t by $k^{3/8-o(1)}$, or $\mathbb{E}F_t(y)$ becomes further from C_t by the same amount.

The distributions of sums of non-identical indicators are well studied, so we can use known tail bounds to bound the ratio between $\mathbb{P}(F_t(z) > C_1)$ and $\mathbb{P}(F_t(z) > C_2)$ for two

thresholds C_1, C_2. Precisely, $\frac{\mathbb{P}(F_t(z)>C_1)}{\mathbb{P}(F_t(z)>C_2)} > \exp\left(\frac{(C_2-C_1)^2}{C_2}\right)$. The result we use is stated and proven in the Appendix A, Lemma 21. Since the two thresholds differ by $\Theta(\sqrt{X})$, between the two scenarios, either $p_{t+1}(z)$ increases by a constant factor for all $z \in I_1$, or $p_{t+1}(w)$ decreases by a constant factor for all $w \in I_2$. So, this implies that $|\mathbb{E}F_{t+1}(x) - \mathbb{E}F_{t+1}(y)|$ varies by $\Theta(1)\min_{i\in\{1,2\}}(|I_i \cap A_t|)$ between the two scenarios.

This is a significant variation; as in Lemma 13, for any $x \in [n]$, $C_t \geq \mathbb{E}F_{t+1}(x)$. So, in the case where x has more expected input, $\mathbb{E}F_{t+1}(y) < \mathbb{E}F_{t+1}(x) - \Theta(\mathbb{E}F_{t+1}(y))$. By the Chernoff bound (Lemma 19), the probability that $F_{t+1}(y)$ will exceed C_{t+1} is exponentially small.

$$Pr\left(F_{t+1}(y) > C_{t+1} = (1 + \Theta(1))\mathbb{E}F_{t+1}(y)\right) < \exp\left(-\Theta(1)\mathbb{E}F_{t+1}(y)\right) < \exp\left(-k^{3/4-o(1)}\right)$$

By the Chernoff bound, the probability that $F_{t+1}(y)$ will exceed C_{t+1} is exponentially small. A similar argument applies if $\mathbb{E}F_{t+1}(y) > \mathbb{E}F_{t+1}(x)$. Therefore, since there is a constant probability that the two balls will deviate from each other, either I_1 or I_2 will be eliminated in a constant number of steps.

The same argument applies to any pair of balls (I_i, I_j). Therefore, the number of balls reduces by a constant factor within a constant number of steps. This leads to convergence to a single ball within $O(\ln k)$ steps.

At this point, $\hat{k} = k$, so applying Lemma 15 again, we can conclude that A_t converges to a single ball of size $O(\sigma k^{-1/2}\sqrt{\ln k})$ in $O((\log k)^c)$ steps. ∎

Finally, we prove that the set A_t, with high probability, remains within a small subset for all $t \geq t^*$.

Proof [Proof of Theorem 7] Let $A \subset V$ with $|A| = k$, and let I be a ball surrounding $k - k^{2/3}$ points of A. Assume that $r = r(I) = \sigma k^{-1/3+\epsilon}$ and $I = B_r(p)$.

We consider 2 cases, identically to the proof of Lemma 15:

Case 1: There exists a half space dividing I into two spherical caps, R_1 and R_2, with heights $r/4$ and $7r/4$, such that $|R_2 \cap A| \leq k/(\ln k)^2$. In this case, the ball is "imbalanced" in the sense that one portion of of the ball contains the vast majority of the points.

Without loss of generality, let $p = [r, 0, 0, \ldots, 0]$, $R_1 = \{y \in I : y_1 \leq r/4\}$, and $R_2 = \{y \in I : y_1 > r/4\}$ (as illustrated in Figure 6). We define $E[z] = F(z; A) = \sum_{y \in A} g(y, z)$.

Then we can bound the derivative with respect to the first coordinate:

$$\frac{\partial}{\partial z_1}E[z] = \sum_{y \in A\cap I}\frac{\partial}{\partial z_1}g(y, z) = \sum_{y \in A\cap I}-\frac{z_1 - y_1}{\sigma^2}g(y, z) + \sum_{y \in A\setminus I}-\frac{z_1 - y_1}{\sigma^2}g(y, z) \tag{10}$$

$$= \sum_{y \in R_1\cap A}-\frac{z_1 - y_1}{\sigma^2}g(y, z) + \sum_{y \in R_2\cap A}-\frac{z_1 - y_1}{\sigma^2}g(y, z) + \sum_{y \in A\setminus I}-\frac{z_1 - y_1}{\sigma^2}g(y, z) \tag{11}$$

Let $z = [z_1, z_2, \ldots, z_n]$ where $z_1 \geq 3r/8$ and $dist(I, z) = O(r)$. This implies:

$\min_{y \in R_2}\frac{z_1 - y_1}{\sigma^2}g(y, z) > -\frac{2r}{\sigma^2}\exp\left(-2r^2/\sigma^2\right)$

$\min_{y \in R_1}\frac{z_1 - y_1}{\sigma^2}g(y, z) > \frac{r}{8\sigma^2}\exp\left(\frac{-O(1)r^2}{2\sigma^2}\right)$, and

$\min_{y \in [0,1]^d \setminus I} \frac{z_1 - y_1}{\sigma^2} g(y,z) > -\frac{1}{\sigma} e^{-1/2}$

Returning to Equation 11:

$$\frac{\partial}{\partial z_1} E[z] \leq -|A_t \cap R_1| \min_{y \in R_1} \frac{z_1 - y_1}{\sigma^2} g(y,z) - |A_t \cap R_2| \min_{y \in R_2} \frac{z_1 - y_1}{\sigma^2} g(y,z) - |A_t \setminus I| \min_{y \in [0,1]^d} \frac{z_1 - y_1}{\sigma^2} g(y,z) \tag{12}$$

By assumption, $|A_t \cap R_2| \leq k/(\ln k)^2$ and $|A_t \setminus I| = k^{2/3}$. Replacing this:

$$\frac{\partial}{\partial z_1} E[z] \leq -\frac{r}{8\sigma^2} e^{-O(1)r^2/\sigma^2} k + \frac{2r}{\sigma^2} \frac{k}{\ln k^2} + \frac{\Theta(1)k^{2/3}}{\sigma} = -\Theta(1)\frac{kr}{\sigma^2}$$

Let $z' = [z_1', z_2', \ldots, z_n']$ where $z_1' \geq r/2$ and $dist(I, z') < r/8$. Consider the point $w = z' - r/8$. By definition, the derivative bounds above apply for all points on the line between w and z'. This gives us a lower bound on $E[w] - E[z']$. While $w \notin V$ almost surely, by Lemma 16 there exists a point $w' \in V$ within a radius of $O((\log n/n)^{1/d})$ of w. Applying 11, $E[w'] - E[z'] > E[w] - E[z'] - o(1)$.

We will prove an analogous result to Lemma 13 for sets A contained mostly within a ball of radius r.

Lemma 26 *Let I be a ball of radius $r = \sigma k^{-1/3+\epsilon}$, for some constant $\epsilon > 0$, surrounding $k - k^{2/3}$ points of A. Let C be the threshold when the k-cap function is applied to A. With high probability, for all such sets $A \subset V$ with $|A| = k$, $C \geq \max_x E[x]$.*

Proof [Proof of Lemma 13]

The derivative of $E[x]$ is, for any dimension i:

$$\frac{\partial}{\partial x_i} E[x] = \sum_{z \in A_t} \frac{\partial}{\partial x_i} g(x,z) = \sum_{z \in A_t} -\frac{x_i - z_i}{\sigma^2} \exp\left(\frac{-\|x-z\|^2}{2\sigma^2}\right)$$

Let $\hat{x} = \frac{x-z}{\sigma}$. The maximum of $\left|\hat{x} \exp\left(\frac{-\hat{x}^2}{2}\right)\right|$ occurs at $e^{-1/2}$. For $y \in I$, this is maximized at $\hat{x} = \frac{2r}{\sigma}$. Thus, we have

$$\left|\frac{\partial}{\partial x_i} E[x]\right| \leq \sum_{z \in A_t} \frac{|x_i - z_i|}{\sigma^2} \exp\left(\frac{-\|x-z\|^2}{2\sigma^2}\right) < \frac{e^{-1/2}k^{2/3}}{\sigma} + k\frac{2r}{\sigma^2} = \frac{2k^{2/3+\epsilon}}{\sigma}(1 + o(1))$$

Therefore, the directional derivative, as in Lemma 11, is at most this value, times a factor of $\sqrt{d}$. For any $y \in B_{\sigma k^{-1+\epsilon}}(x)$, the difference between $E[y]$ and $E[x]$ can be bounded:

$$E[y] = E[x] - o(1)$$

By Lemma 16, for any point x, there are $\Omega((n/\log n) \cdot \sigma^d k^{-d+d\epsilon})$ points in a radius of $\sigma k^{-1+\epsilon}$ of x. By the assumption that $n \geq k^{2+d}$, this is $\tilde{\Omega}(k^{1+d\epsilon})$.

Therefore, if $C = E[x]$, then there are $\tilde{\Omega}(k^{1+d\epsilon})$ points where $E[y] > C_t - O(1)$. Since each edge is chosen independently, Chernoff type bounds apply; if $\mathbb{P}(F_t(y) > E[y] + O(1)) = \Theta(1)$, then with high probability there are k points that exceed C_t.

Using the loose bound given in (Volkova, 1996), we can bound $\mathbb{P}(F_t(x) > C_t)$ using the CDF of the normal distribution. For any sum of independent indicators S with mean μ and variance σ, the CDF can be approximated as follows:

$$\sup_m \left| \mathbb{P}(S \leq m) - G\left(\frac{m + 1/2 - \mu}{\sigma}\right) \right| \leq \frac{\sigma + 3}{4\sigma^3} < \frac{1}{\sigma^2}$$

Where $G(x) = \Phi(x) - \frac{\gamma}{6\sigma^3}(x^2 - 1)\frac{e^{-x^2/2}}{\sqrt{2\pi}}$, and $\gamma = \mathbb{E}[(S - \mu)^3]$ is the skewness. This holds for any $\sigma \geq 10$.

We can assume that $V[y] > 10$; otherwise, $E[y] = \sum_{y \in A_t} = k(1 - o(1/k))$, so we can assume that $C = k$.

Fix $x \in [n]$. From the above equation, we find that for any $t > 0$:

$$\mathbb{P}\left(F_t(x) > E[x] + tV[x] - 1/2\right) > 1 - \left[G(t) + \frac{1}{V[x]^2}\right]$$

Substituting the value of G:

$$\mathbb{P}\left(F_t(x) > E[x] + tV[x] - 1/2\right) > 1 - \Phi(t) + \frac{\gamma(t^2 - 1)}{6\sqrt{2\pi}V[x]^3}e^{-t^2/2} - \frac{1}{V[x]^2}$$

Here, we will make two approximations. First, the exact value of γ is $\sum_{z \in A_t} g(x, z)(1 - g(x, z))(1 - 2g(x, z))$. Therefore, $\gamma > -V[x]^2$, so $\gamma(t^2 - 1) > -V[x]^2 t^2$

Second, we will substitute the lower tail bound for $1 - \Phi(t) \geq \frac{1}{\sqrt{2\pi}}\left(t^{-1} - t^{-3}\right)e^{-t^2/2} \geq \frac{1}{t\sqrt{8\pi}}e^{-t^2/2}$ for $t \geq 2$.

This leaves us with:

$$\mathbb{P}\left(F_t(x) > E[x] + tV[x] - 1/2\right) > \frac{1}{t\sqrt{8\pi}}e^{-t^2/2} - \frac{t^2}{6\sqrt{2\pi}V[x]}e^{-t^2/2} - \frac{1}{V[x]^2}$$

Setting $t = \Theta(1)/V[x]$, this occurs with constant positive probability p.

The probability that there are not k points which exceed $C = E[x]$ is at least $\binom{k^{1+d\epsilon}}{k}(1 - p)^{k^{1+d/2}-k} = (1 - p)^{k^{1+d\epsilon}(1 - o(1))}$.

The number of possible subsets A is at most $\binom{n}{k} < n^k = e^{k \log n}$.

By the union bound, this holds for all subsets A with high probability. ∎

Returning to the proof of the original theorem, we recall that there exists a point w such that:

$$E[w] \geq E[z'] + \frac{r}{8} \cdot k\frac{r}{8\sigma^2}e^{-2r^2/\sigma^2}(1 - o(1)) = E[z] + \Theta(1)k\frac{r^2}{\sigma^2} = E[z'] + \Theta(k^{1/3+2\epsilon})$$

Comparing this to $k - E[z']$

$$k - E[z'] \leq k(1 - e^{-2r^2/\sigma^2}) \leq k\frac{2r^2}{\sigma^2} = 2k^{1/3+2\epsilon}$$

We can apply Lemma 19 to $k - E[z']$; let $Z = k - F(z'; A)$. Then, $\mathbb{P}(Z < (1 - \delta)\mathbb{E}Z]) \leq e^{-\delta^2\mathbb{E}Z/2}$. So,

$$\mathbb{P}(Z < \mathbb{E}Z - \Theta(1)\mathbb{E}Z) < e^{-\Theta(1)k^{1/3+2\epsilon}}$$

The probability that there exist $k^{2/3}$ points which violate the condition is at most:

$$\binom{n}{k^{2/3}}(e^{-\Theta(1)k^{1/3+2\epsilon}})^{k^{2/3}} < e^{k^{2/3}\log n}e^{-\Theta(1)k^{1+2\epsilon}\log n}$$

Since there are at most $\binom{n}{k} = O(e^{k\log n})$ possible k-subsets of V, this is true by the union bound for all subsets A with high probability.

Case 2: In this section, we assume that no such imbalanced partition of I exists. For all $x \in \partial I$, denoting $R_2 = \{z \in I : x - z \cdot (\frac{x-p}{\|x-p\|} > r/4), |R_2 \cap A| \geq k/(\ln k)^2$. We will show that for any z within Δr of the boundary of I, $\mathbb{P}(F(z; A) > C) < 1/n^3$.

Fix x, and assume without loss of generality that $p = [r, 0, 0, \ldots, 0]$ and $x = [0, 0, \ldots, 0]$

Let $z = [z_1, 0, 0, \ldots, 0]$ where $-\sigma \log n < z_1 < \Delta r = r/(\ln k)^3$. Then, $\max_{y \in R_1} \frac{z_1 - y_1}{\sigma^2} g(y, z) = \frac{z_1}{\sigma^2}\exp\left(\frac{-z_1^2}{2\sigma^2}\right) < \frac{\Delta r}{\sigma^2}\exp\left(\frac{-(\Delta r)^2}{2\sigma^2}\right) = \frac{\Delta r}{\sigma^2}(1-o(1))$ Also, separately taking the minima of $y_1 - z_1$ and $g(y, z)$, $\min_{y \in R_2} \frac{y_1 - z_1}{\sigma^2} g(y, z) = \frac{r}{4\sigma^2}(1 - o(1))$. Bounding the derivative again:

$$\frac{\partial}{\partial z_1}E[z] \geq -|A \cap R_1|\frac{\Delta r}{\sigma^2}(1 - o(1)) + |A \cap R_2|\frac{r}{4\sigma^2}(1 - o(1)) - |A \setminus I|\frac{O(1)}{\sigma} \tag{13}$$

By the assumption of the case, $|A \cap R_2| \geq k/(\ln k)^2$. Also, by the assumption of the theorem $|A \setminus I| = O(k^{2/3})$. Replacing this:

$$\frac{\partial}{\partial z_1}E[z] \geq -k\frac{\Delta r}{\sigma^2}(1 - o(1)) + \frac{kr}{4\sigma^2(\ln k)^2}(1 - o(1)) - \frac{O(k^{2/3})}{\sigma}$$

Substituting Δr

$$\frac{\partial}{\partial z_1}E[z] \geq -k(1 - o(1))\frac{r}{\sigma^2(\ln k)^3} + \frac{\Theta(k)r}{\sigma^2(\ln k)^2} - \frac{O(k^{2/3})}{\sigma} = \frac{kr}{\sigma^2(\ln k)^3}[1 - o(1) + \Theta(\ln k)]$$

Suppose $z = [z_1, 0, 0, \ldots, 0]$ where $z_1 < \Delta r/2$. Let $w = z_1 + [\Delta r/2, 0, 0, \ldots, 0]$. Using the lower bound on the derivative,

$$E[w] \geq E[z] + \frac{\Delta r}{2} \cdot \frac{kr}{\sigma^2(\ln k)^3}[1 - o(1) + \ln k] \geq E[z] + \frac{1}{4}\frac{kr^2}{\sigma^2}\ln k$$

Using the same bound as above for $k - E[z]$:

$$k - E[z] \leq k(1 - e^{-2r^2/\sigma^2}) \leq k\frac{2r^2}{\sigma^2}$$

We can apply Lemma 19 to $k - E[z]$; let $Z = k - F(z; A)$. Then, $\mathbb{P}(Z < (1 - \delta)\mathbb{E}Z]) \leq e^{-\delta^2\mathbb{E}Z/2}$. So,

$$\mathbb{P}(Z < \mathbb{E}Z - \Theta(1)\mathbb{E}Z) < e^{-\Theta(1)k^{1/3+2\epsilon}}$$

The probability that there exist $k^{2/3}$ points which violate the condition is at most:

$$\binom{n}{k^{2/3}}(e^{-\Theta(1)k^{1/3+2\epsilon}})^{k^{2/3}} < e^{k^{2/3}\log n}e^{-\Theta(1)k^{1+2\epsilon}}$$

Since there are at most $\binom{n}{k} = O(e^{k\log n})$ possible k-subsets of V, this is true by the union bound for all subsets A with high probability. ∎

B.3. Continuous α-cap process

First, we will prove that single intervals of width α are the only possible fixed points.

Theorem 27 (Fixed Points) *For any even, differentiable, nonnegative, and integrable function $g : [-1, 1] \to \Re_+$ with $g'(x) < 0$ for all $x > 0$, the only fixed points $(A_{t+1} = A_t)$ of the α-cap Process are single intervals of width α.*

The next lemma follows from the properties of g.

Lemma 28 *The following holds for all $b > a$:*

$$\int_a^b g(y - a) \, dy = \int_a^b g(y - b) \, dy$$

We proceed to the proof of the fixed point characterization.
Proof [Proof of Theorem 27] First, we show that if $A_t = [a, b]$ is a single interval, then it is a fixed point. Since A_t is 1 on the interval and 0 elsewhere, we can rewrite $F_t(x)$ as follows:

$$F_t(x) = \int_a^b g(y - x) \, dy \tag{14}$$

Define the threshold $C_t = F_t(a) \int_a^b g(y - a) \, dy$. If $x \in [a, b]$,

$$F_t(x) = \int_a^b g(y - x) \, dy \geq \int_a^b g(y - a) \, dy = C_t$$

It's easiest to see this by breaking $F_t(x)$ into two integrals, $\int_a^b g(y - x) \, dy = \int_a^x g(y - x) \, dy + \int_x^b g(y - x) \, dy$. By Lemma 28, $\int_a^x g(y - x) \, dy = \int_a^x g(y - a) \, dy$. Then, since $a < x$, $g(y - x) > g(y - a)$ for all $y \in [x, b]$. This implies $\int_x^b g(y - x) \, dy \geq \int_x^b g(y - a) \, dy$
Therefore, $F_t(x) \geq F_t(a) = C_t$ for all $x \in [a, b]$.
Similarly, if $x < a$ or $x > b$,

$$F_t(x) = \int_a^b g(y - x) \, dy < \int_a^b g(y - a) \, dy = C_t$$

This implies that if C_t is chosen in this way, then $A_{t+1} = [a, b] = A_t$.
Next, let A_t be the union of finite intervals. Let $A_t = \bigcup_{j=1}^n [a_j, b_j]$ where for all $j < n$, $a_j < b_j < a_{j+1} < b_{j+1}$, and $n > 1$. We will show that this is not fixed.
$F_t(x)$ can be expressed as the following:

$$F_t(x) = \sum_{j=1}^n \int_{a_j}^{b_j} g(y - x) \, dy$$

Consider $F_t(a_n)$ and $F_t(b_n)$.

$$F_t(a_n) = \int_{a_n}^{b_n} g(y - a_n) \, dy + \sum_{j=1}^{n-1} \int_{a_j}^{b_j} g(y - a_n) \, dy$$

34

$$F_t(b_n) = \int_{a_n}^{b_n} g(y - b_n)\, dy + \sum_{j=1}^{n-1} \int_{a_j}^{b_j} g(y - b_n)\, dy$$

By Lemma 28, $\int_{a_n}^{b_n} g(y - a_n)\, dy = \int_{a_n}^{b_n} g(y - b_n)\, dy$. Also, since $a_n < b_n$, $\int_{a_j}^{b_j} g(y - a_n)\, dy >$ $\int_{a_j}^{b_j} g(y - b_n)\, dy$ for all $j < n$. Therefore, $F_t(a_n) > F_t(b_n)$.

Notice that by definition, if $A_{t+1} = A_t$, then $C_t \le F_t(x)$ for all $x \in A_{t+1}$. In particular, $C_t \le F_t(b_n) < F_t(a_n)$. However, since $F_t(x)$ is a continuous function, $C_t < F_t(x)$ in some small window $[a_n - \epsilon, a_n]$. By definition, $a_n - \epsilon \in A_{t+1}$, but $a_n - \epsilon \notin A_t$ for small enough ϵ. This implies $A_t \ne A_{t+1}$. ∎

Now, we will prove the main convergence theorem.

Theorem 2 *Let A_0 be a finite union of intervals in [0,1]. Let $g : [-1, 1] \to \mathbb{R}_+$ be a differentiable, even, nonnegative and integrable function with $g'(x) < 0$ for all $x > 0$. For any such g, the α-cap process converges to a single interval of width α. Moreover, the number of steps to convergence is*

$$O\left(\frac{\max_{[0,1]} |g'(x)|}{\min_{[\frac{\alpha}{8}, 1]} |g'(x)|} \right).$$

Proof [Proof of Theorem 2] At a given step $t \ge 0$, A_t is a union of finite intervals on [0,1]. We will show that if the number of intervals is greater than 1, the distance between the midpoints of the first and last intervals decreases at each step, and this decrease is not diminishing.

Let $A_t = \bigcup_{j=1}^{n} [a_j, b_j]$ where the intervals are disjoint and increasing; for all $j < n$, $0 \le a_j < b_j < a_{j+1} < b_{j+1} \le 1$. Define the midpoint of the kth interval $m_k = \frac{a_k + b_k}{2}$. By 27, if $n = 1$, then the process has converged (i.e. $A_{t+1} = A_t$). If $n > 1$, $[a_1, b_1]$ and $[a_n, b_n]$ are the first and last intervals in A_t, respectively. We will show that the distance between the midpoints, $m_n - m_1$, decreases by at least a constant.

Suppose the first local maximum of F_t occurs at a value $m_1 + \delta$. This proof will show that the shift in the midpoint m_1 is bounded from below by a constant depending on δ.

Recall the definition of F_t and write its derivative:

$$F_t(x) = \sum_{k=1}^{n} \int_{a_k}^{b_k} g(y - x)\, dy$$

$$\frac{dF_t}{dx} = \sum_{k=1}^{n} g(a_k - x) - g(b_k - x)$$

Since g is assumed to be decreasing on $[0, 1]$, if $x < m_k$, then $g(a_k - x) > g(b_k - x)$. This implies F_t is increasing on $[0, m_1)$, so the first local maximum must occur at $m_1 + \delta > m_1$. For ease of notation, we will define the following points; let $m_1 + \delta$ be the smallest local maximum of F_t. Let $z = m_1 - \epsilon$ be the minimum value where $F_t(x) \ge C_t$. Finally, let $z' = m_1 + \epsilon + \epsilon'$ be the smallest value such that $z' > z$, $F_t(z') = F_t(z) = C_t$, and $F_t(x) > C_t$ for all $x \in (z, z')$.

Claim 28.1 *If $z \geq m_1$ (that is, $\epsilon < 0$), then $[z, m_1 + \delta] \subset A_{t+1}$. Hence, the midpoint of the first interval of A_{t+1} is greater than $m_1 + \frac{\delta}{2}$.*

If $z < m_1$, the proof is more involved.

For the rest of this section, note that $\epsilon > 0$. Note that it is possible that $z = m_1 - \epsilon < a_1$, such that the left end of the interval decreases. However, the midpoint of the interval will always increase. The influence on F_t from the first interval is the same for $m_1 - \epsilon$ and $m_1 + \epsilon$.

$$
\begin{aligned}
\int_{a_1}^{b_1} g(y - (m_1 - \epsilon)) \, dy &= \int_{-m_1+\epsilon}^{m_1+\epsilon} g(z) \, dz \\
&= -\int_{m_1-\epsilon}^{-m_1-\epsilon} g(-z) \, dz \\
&= \int_{-m_1-\epsilon}^{m_1-\epsilon} g(z) \, dz = \int_{a_1}^{b_1} g(y - (m_1 + \epsilon)) \, dy
\end{aligned}
$$

For any $[a_k, b_k]$ where $m_1 < a_k < b_k$,

$$
\begin{aligned}
\int_{a_k}^{b_k} g(y - (m_1 + \epsilon)) \, dy &= \int_{a_k-m_1-\epsilon}^{b_k-m_1-\epsilon} g(z) \, dz \\
&> \int_{a_k-m_1+\epsilon}^{b_k-m_1+\epsilon} g(z) \, dz \\
&= \int_{a_k}^{b_k} g(y - (m_1 - \epsilon)) \, dy
\end{aligned}
$$

This implies that $F_t(m_1 + \epsilon) > F_t(m_1 - \epsilon)$ for any $\epsilon > 0$. Since F_t is continuous, there is a small value $\epsilon' > 0$ such that $F_t(m_1 + \epsilon + \epsilon') = C_t$, and $F_t(x) > C_t$ in between. At A_{t+1}, the first interval becomes $[m_1 - \epsilon, m_1 + \epsilon + \epsilon']$, which has the midpoint $m_1 + \frac{\epsilon'}{2}$.

Therefore, the midpoint of the first interval increases. A symmetric argument shows that the midpoint of the last interval must decrease. Next, we will show that ϵ' is bounded below by a constant factor of δ.

Claim 28.2 *Let ϵ, ϵ' be the values defined above. Assume without loss of generality that $b_1 - a_1 \leq \frac{\alpha}{2}$ (If $[a_1, b_1]$ is large, a symmetric argument applies to $[a_n, b_n]$).*

$$
\epsilon' \geq \frac{\alpha}{2} \min\{\frac{\alpha}{4}, \epsilon\} \frac{\min_{y \in [\frac{\alpha}{4}, 1]} |g'(y)|}{g(0) - g(1)}
$$

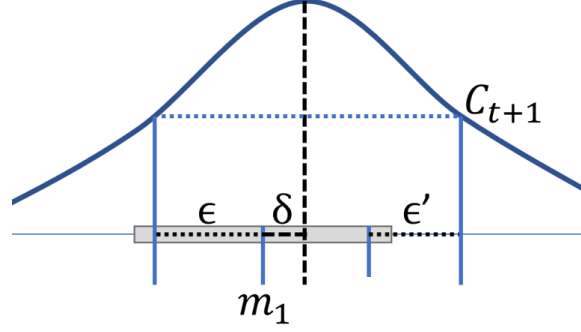Figure 7: An illustration of the terms defined in this proof. The grey box represents the first interval, $[a_1, b_1]$. The curve is F_t.

Proof Let m be the median of A_t; by the assumption, $m \notin [a_1, b_1]$, so $m \geq a_2$.

$$F_t(m_1 + \epsilon) - F_t(m_1 - \epsilon) = \sum_{k=1}^{n} \int_{a_k}^{b_k} g(y - (m_1 + \epsilon))\, dy - \sum_{k=1}^{n} \int_{a_k}^{b_k} g(y - (m_1 - \epsilon))\, dy$$

$$= \sum_{k=2}^{n} \int_{a_k}^{b_k} g(y - (m_1 + \epsilon)) - g(y - (m_1 - \epsilon))\, dy$$

$$= \int_{a_2}^{b_n} [g(y - (m_1 + \epsilon)) - g(y - (m_1 - \epsilon))] A_t(y)\, dy$$

$$\geq \int_{m}^{b_n} [g(y - (m_1 + \epsilon)) - g(y - (m_1 - \epsilon))] A_t(y)\, dy$$

$$\geq \frac{\alpha}{2} [\min_{y \in [m, b_n]} g(y - (m_1 + \epsilon)) - g(y - (m_1 - \epsilon))]$$

By the assumption that $b_1 - a_1 < \frac{\alpha}{2}$, for any $y > m$, $y - m_1 \geq m - m_1 > \frac{\alpha}{4}$. If ϵ is small enough that $m_1 + \epsilon < m$,

$$g(y - (m_1 + \epsilon)) - g(y - (m_1 - \epsilon)) > g(y - m_1) - g(y - (m_1 - \epsilon)) \geq \epsilon \min_{z \in [0, \epsilon]} |g'(y - m_1 + z)|$$

If $\epsilon > m - m_1$, then for any y such that $m < y < m_1 + \epsilon$,

$$g(y - (m_1 + \epsilon)) - g(y - (m_1 - \epsilon)) \geq g(\epsilon) - g(y - (m_1 - \epsilon)) \geq [y - m_1] \min_{z \in [0, y - m_1]} |g'(\epsilon + z)|$$

In both cases, $F_t(m_1 + \epsilon) - F_t(m_1 - \epsilon)$ can be bounded by restricting g'. Let $c_1 = \min_{y \in [\frac{\alpha}{4}, 1]} |g'(y)|$. Since g' is strictly decreasing, $c_1 \leq \min_{y \in [m - m_1, b_n + \epsilon - m_1]} |g'(y)|$. For ϵ

37

small ($\epsilon < m - m_1$):

$$F_t(m_1 + \epsilon) - F_t(m_1 - \epsilon) \geq \frac{\alpha}{2} [\min_{y \in [m, b_n]} g(y - (m_1 + \epsilon)) - g(y - (m_1 - \epsilon))]$$

$$\geq \frac{\alpha \epsilon}{2} \min_{y \in [m, b_n]} \min_{z \in [0, \epsilon]} g'(y - m_1 + z)$$

$$\geq \frac{\alpha \epsilon}{2} \min_{y \in [m, b_n + \epsilon]} g'(y - m_1)$$

$$\geq \frac{\alpha}{2} \epsilon c_1$$

For ϵ large ($\epsilon \geq m - m_1$):

$$F_t(m_1 + \epsilon) - F_t(m_1 - \epsilon) \geq \frac{\alpha}{2} [\min_{y \in [m, b_n]} g(y - (m_1 + \epsilon)) - g(y - (m_1 - \epsilon))]$$

$$\geq \frac{\alpha}{2} \min_{y \in [m, b_n]} [y - m_1] \min_{z \in [0, y - m_1]} |g'(\epsilon + z)|$$

$$\geq \frac{\alpha}{2} \frac{\alpha}{4} \min_{z \in [0, b_n - m_1]} |g'(\epsilon + z)|$$

$$\geq \frac{\alpha}{2} \frac{\alpha}{4} c_1$$

Combining both cases,

$$F_t(m_1 + \epsilon) - F_t(m_1 - \epsilon) \geq \frac{\alpha}{2} \min\{\frac{\alpha}{4}, \epsilon\} c_1 \tag{15}$$

Let $c_2 = g(0) - g(1)$. Since g is continuous and decreasing, $c_2 > 0$ and

$$\left| \frac{dF_t}{dx} \right| = \left| \sum_{k=1}^{n} g(a_k - x) - g(b_k - x) \right| \leq c_2 \tag{16}$$

To recall, we define ϵ and ϵ' such that $F_t(m_1 - \epsilon) = F_t(m_1 + \epsilon + \epsilon') = C_t$. By 15,

$$F_t(m_1 + \epsilon) - F_t(m_1 + \epsilon + \epsilon') \geq \frac{\alpha}{2} \min\{\frac{\alpha}{4}, \epsilon\} c_1$$

and by 16,

$$F_t(m_1 + \epsilon) - F_t(m_1 + \epsilon + \epsilon') \leq c_2 \epsilon'$$

Combining these two equations implies,

$$\epsilon' \geq \frac{\alpha}{2} \min\{\frac{\alpha}{4}, \epsilon\} \frac{c_1}{c_2}$$

$\blacksquare$

The only unbounded value in this equation is ϵ. Recall that $m_1 + \delta$ is defined to be the earliest local maximum of F_t; in the case of small ϵ, $[m_1 - \epsilon, m_1 + \delta]$ is a subset of the first interval of A_{t+1}. If $\epsilon \leq \frac{\delta}{2}$, then the midpoint is at least $m_1 + \frac{\delta}{4}$.

Combining this fact with Claim 28.1, the shift of the midpoint of the first interval is a constant multiple of δ. We will show that δ is also bounded from below by a constant value.

Claim 28.3 *Under the assumption that $b_1 - a_1 < \frac{\alpha}{2}$,*

$$\delta \geq \frac{\min_{y \in [\frac{\alpha}{8}, 1]} |g'(y)|}{\max_{z \in [0,1]} |g'(z)|} \cdot \frac{\alpha}{8}$$

Proof By assumption, at $m_1 + \delta$, $\frac{dF_t}{dx} = 0$.

$$\sum_{j=1}^{n} g(a_j - (m_1 + \delta)) - g(b_j - (m_1 + \delta)) = 0$$

Since g is decreasing and symmetric, the sign of $g(a_j - (m_1 + \delta)) - g(b_j - (m_1 + \delta))$ depends on whether $m_1 + \delta$ is closer to a_j or b_j. Suppose this term is negative for $j = 1, ..., k$ and positive for $j = k+1, ..., n$. Additionally, assume that $\delta < \frac{\alpha}{8}$; since $b_1 - a_1 < \frac{\alpha}{2}$, $m_1 + \delta < a_1 + \frac{\alpha}{4} + \frac{\alpha}{8}$ (Recall that m is defined as the median of A_t, so by definition $m \geq a_1 + \alpha/2$). Using this, we can assume that $a_{k+1} \leq m$.

Again, since g is decreasing, $g(a_j - x) - g(b_j - x) \geq (b_j - a_j) \min_{y \in [a_j, b_j]} |g'(y - x)|$.

$$\sum_{j=k+1}^{n} g(a_j - (m_1 + \delta)) - g(b_j - (m_1 + \delta)) \geq \sum_{j=k+1}^{n} (b_j - a_j) \min_{y \in [a_j, b_j]} |g'(y - (m_1 + \delta))|$$

$$\geq \frac{\alpha}{2} \min_{y \in [m,1]} |g'(y - (m_1 + \delta))|$$

$$\geq \frac{\alpha}{2} \min_{y \in [\frac{\alpha}{8}, 1]} |g'(y)|$$

For $j = 1...k$, $g(a_j - (m_1 + \delta)) - g(b_j - (m_1 + \delta)) < 0$.

$$\sum_{j=1}^{k} g(b_j - (m_1 + \delta)) - g(a_j - (m_1 + \delta)) \leq g(b_k - (m_1 + \delta)) - g(a_1 - (m_1 + \delta))$$

$$\leq (|a_1 - (m_1 + \delta)| - |b_k - (m_1 + \delta)|) \max_{z \in [0,1]} |g'(z)|$$

$$\leq (2(m_1 + \delta) - b_k - a_1)) \max_{z \in [0,1]} |g'(z)|$$

Recall that since $\frac{dF_t}{dx} = 0$, we have:

$$\sum_{j=1}^{k} g(b_j - (m_1 + \delta)) - g(a_j - (m_1 + \delta)) = \sum_{j=k+1}^{n} g(a_j - (m_1 + \delta)) - g(b_j - (m_1 + \delta))$$

Combining the two equations above:

$$(2(m_1 + \delta) - b_k - a_1)) \max_{z \in [0,1]} |g'(z)| \geq \min_{y \in [\frac{\alpha}{8}, 1]} |g'(y)| \frac{\alpha}{2}$$

Therefore, δ is bounded:

$$\delta \geq \frac{\min_{y \in [\frac{\alpha}{8}, 1]} |g'(y)|}{\max_{z \in [0,1]} |g'(z)|} \cdot \frac{\alpha}{4} + \frac{a_1 + b_k}{2} - m_1$$

Since $b_k \geq b_1$, the midpoint of a_1 and b_k is greater than m_1. Combining this fact with the earlier assumption that $\delta < \frac{\alpha}{8}$, we have

$$\delta \geq \frac{\min_{y \in [\frac{\alpha}{8},1]} |g'(y)|}{\max_{z \in [0,1]} |g'(z)|} \cdot \frac{\alpha}{8}$$

∎

This implies that in the case where $b_1 - a_1 \leq \frac{\alpha}{2}$, the change in $m_n - m_1$ is bounded below by a constant value. A symmetric argument applies if $b_n - a_n \leq \frac{\alpha}{2}$. Since $m_n - m_1$ decreases at each step, the process must converge to a single interval in finite steps.

∎

This theorem indicates that the speed of convergence depends on the function g and the size of A_0 (α). This process converges to a fixed point in at most

$$O\left(\frac{\max_{[0,1]} |g'(x)|}{\min_{[\frac{\alpha}{8},1]} |g'(x)|}\right)$$

steps. For example, consider the Gaussian function $g(x) = \exp(\frac{-x^2}{2\sigma^2})$. We have $g'(x) = -\frac{x}{\sigma^2} \exp(\frac{-x^2}{2\sigma^2})$. The maximum is:

$$\max_{[0,1]} |g'(x)| = |g'(\sigma)| = \frac{1}{\sigma} \exp(-1/2)$$

The minimum can occur at either endpoint depending on α and σ:

$$\min_{[\frac{\alpha}{8},1]} |g'(x)| = \min\left(\left|g'\left(\frac{\alpha}{8}\right)\right|, |g'(1)|\right) \approx \min\left(\frac{\alpha}{8\sigma^2}, \frac{1}{\sigma^2} \exp\left(\frac{-1}{2\sigma^2}\right)\right)$$

If $\alpha < 8\left(\frac{-1}{2\sigma^2}\right)$, the lower bound on the number of convergence steps is

$$O\left(\frac{\frac{1}{\sigma} \exp(-1/2)}{\frac{1}{\sigma^2} \exp\left(\frac{-1}{2\sigma^2}\right)}\right) \approx O(\sigma)$$

Otherwise, the bound is

$$O\left(\frac{\frac{1}{\sigma} \exp(-1/2)}{\frac{\alpha}{8\sigma^2}}\right) \approx O(\frac{\sigma}{\alpha})$$

Another example is the inverse square distance $g(x) = \frac{1}{c+x^2}$ for a value $c > 0$. We have $g'(x) = \frac{-2x}{(c+x)^2}$. The maximum occurs at $\max_{[0,1]} |g'(x)| = \frac{-3\sqrt{3}}{8} c^{-3/2}$. The minimum occurs at one of the endpoints:

$$\min_{[\frac{\alpha}{8},1]} |g'(x)| = \min\{|g'(\alpha/8)|, |g'(1)|\} \approx \min\left(\frac{\alpha}{4c^2 + c\alpha}, \frac{2}{(c+1)^2}\right)$$

40

Therefore, the lower bound on the number of convergence steps when c is not small is approximately:

$$O\left(\frac{c^{-3/2}(c^2 + c\alpha)}{\alpha}\right) = O\left(\frac{c^{1/2}}{\alpha} + c^{-1/2}\right)$$

For sufficiently small c, this gives:

$$O\left(c^{-3/2}(c+1)^2\right) = O(c^{-3/2})$$