Electronic Colloquium on Computational Complexity, Report No. 87 (2026)

Tight Bounds for Sketching Intersecting Sets, with Applications

Flavio Chierichetti* Mirko Giacchini† Ravi Kumar‡
Alessandro Panconesi† Erasmo Tani† Andrew Tomkins‡

Abstract

In this work, we study the space complexity of sketching the *intersection profile* of a distribution D on $2^{[n]}$. Specifically, we seek a succinct data structure that, for any query set $S \subseteq [n]$, approximates the quantity $\Pr_{T \sim D}[T \cap S \neq \varnothing]$ to within a small constant additive error. Via a probabilistic packing argument, we show that the worst-case bit complexity of this problem is $\Omega(n^2)$, which we also prove to be tight.

We use this lower bound to settle the complexity of three sketching problems. (i) We show that sketching vertex neighborhood sizes in graphs requires $\Omega(n^2)$ bits, standing in sharp contrast to the $\tilde{O}(n)$ complexity of sketching edge cuts. (ii) We obtain tight lower and upper bounds of $\tilde{\Theta}(n^2)$ for sketching coverage functions with additive and multiplicative errors. (iii) We prove an $\Omega(n^2)$ lower bound for sketching Random Utility Models under the ℓ_∞-norm, improving upon the previous $\Omega(n \log n)$ bound and matching the upper bound to within logarithmic factors.

*Reddit, San Francisco. `flavio.chierichetti@gmail.com`
†Sapienza University of Rome. {`giacchini, ale, tani`}`@di.uniroma1.it`
‡Google, Mountain View. {`ravi.k53, atomkins`}`@gmail.com`

ISSN 1433-8092

1 Introduction

Sketching high-dimensional objects into succinct data structures is a fundamental primitive in algorithm design, enabling efficient processing of massive datasets in learning and optimization. A central theme in this area is to study the space complexity of the data structure: how many bits are necessary and sufficient to (approximately) answer queries about the underlying object?

While most sketching results focus on vectors, matrices, or graphs, many learning problems involve distributions over combinatorial objects, whose sketching complexity is less understood. In this paper, we study the space complexity of sketching distributions over subsets of a universe $[n] = \{1, \ldots, n\}$, focusing on their intersection profiles, which we now define.

Given a distribution D over subsets of $[n]$, its *intersection profile* is the function defined by $F_D(S) = \Pr_{T \sim D}[T \cap S \neq \varnothing]$. Our goal is to construct a succinct data structure, using which, for any query set S, one can compute $F_D(S)$ within a small additive error. We show that this problem serves as a general framework for several seemingly distinct problems in data structures and machine learning, including sketching vertex neighborhood sizes in graphs, approximating coverage functions, and sketching the "winning distribution" of random utility models.

1.1 Main Result

We prove that any sketch that approximates the intersection profile of an arbitrary distribution over subsets of $[n]$ to within a constant additive error needs $\Omega(n^2)$ bits. Here, the error is measured as the maximum discrepancy between $F_D(S)$ and its estimate, over the choice of the set S (the ℓ_∞-norm of the F_D vector). This lower bound is established via a packing argument involving a family of distributions defined by random binary matrices, which we show are pairwise far apart in the ℓ_∞-distance. The technique of proving lower bounds via packing arguments is standard in sketching, but our construction involving the embedding of random matrices to separate intersection profiles is novel to this setting. We complement the lower bound with a simple upper bound, showing that $O(n^2/\varepsilon^2)$ bits suffice, matching the lower bound up to the dependence on ε.

The technical difficulty in showing the lower bound arises from the space of distributions over subsets, which is highly non-uniform. Indeed, for a distribution that is sampled uniformly at random from the simplex, the probability of intersecting any fixed S is concentrated around $1 - 2^{-|S|}$ (see Section B). Consequently, for most instances, a constant number of bits suffices to approximate the profile, since, on input S one can simply return $1 - 2^{-|S|}$ as an estimate for $F_D(S)$ and it will be correct with high probability. However, we show that worst-case instances define a geometry that is much harder to capture. These instances occupy thin, low-measure regions of the parameter space, and their geometric "thinness," along with the non-uniform structure of the space, are the main obstacles in constructing the lower bound. In particular, standard volumetric arguments fail, as random instances do not capture the true geometry of the underlying space.

1.2 Applications

We use our main result to resolve open questions and tighten bounds for three key applications:

(i) *Vertex neighborhoods.* It is known that one can sketch the edge cut function of a graph within multiplicative error ε with $O(n \log n/\varepsilon^2)$ bits, which is optimal [Carlson et al., 2019]. It is natural to ask if a similar result is possible for *vertex* boundaries or equivalently, neighborhood cardinalities. We define the vertex neighborhood sketching problem as the task of generating sketches to approximate the size of the neighborhood for any subset $S \subseteq V$. We prove that, unlike edge cuts, vertex boundaries fundamentally *resist* sparsification: any data structure preserving neighborhood sizes up to an additive $0.01n$ error requires $\Omega(n^2)$ bits, meaning that exactly representing the graph, using, e.g., its adjacency matrix, is optimal. We show our result holds both for the problem of sketching vertex neighborhoods, which include all neighbors of elements of S, and for that of sketching outer vertex neighborhoods (a.k.a., outer vertex boundaries), which only count the neighbors that are *not* part of S itself.

(ii) *Coverage functions.* Coverage functions are an important subclass of submodular functions used extensively in optimization and data mining. A coverage function $f(A) = \frac{1}{m}|\bigcup_{i \in A} S_i|$ measures the (normalized) size of the union of subsets of $[m]$ indexed by A. We show that our lower bound for sketching intersection profiles extends to coverage functions, implying that $\Omega(n^2)$ bits are necessary for constant additive error sketches (and hence for $(1 \pm \varepsilon)$-multiplicative error as well), which is tight. Furthermore, for $(1 \pm \varepsilon)$-multiplicative sketching, we extend an algorithm of Bateni et al. [2017] to provide a sketching using $\tilde{O}(n^2/\varepsilon^2)$ bits, matching our lower bound up to logarithmic factors.

(iii) *Random Utility Models (RUMs).* RUMs are distributions over permutations of a universe of n elements that are used to model preferences in economics and machine learning (see, e.g., [Train, 2009]). For a non-empty subset $S \subseteq [n]$ (also called a "slate") and a permutation π sampled from the RUM, the winner in S is the item ranked highest in S according to π. For each $i \in S \subseteq [n]$, the *winning probability* of i in S is the probability that i wins in S under a permutation drawn from the RUM. Therefore, for each slate S, a RUM induces a *winning distribution* over the items of S.

To sketch the winning distributions of a RUM, we seek a data structure that, for each slate S, returns an estimate of the winning distribution over S. The error in the sketch is taken to be the worst-case distance (over the choice of S) between the true winning distribution and the estimate obtained from the sketch. This distance can either be measured in ℓ_1, as twice the total variation distance, or in ℓ_∞, as the maximum error over the individual probability of any item in S.

Chierichetti et al. [2021] showed that sketching the winning distributions of a RUM under the ℓ_1-distance requires $\Omega(n^2)$ bits, and that $O(n^2 \log n)$ bits suffice. However, for the ℓ_∞-distance—which seeks to estimate the winning probability of specific items in a slate—the best known lower bound was the trivial $\Omega(n \log n)$ bits, which are needed to even sketch a single permutation. We improve this to $\Omega(n^2)$ via a reduction from our sketching intersection profiles problem.

Interestingly, our lower bound also applies to the simpler, related problem of sketching all winning probabilities of a *single* item in the universe, for which this bound is tight.

2 Related Work

Sketching edge cuts. A long line of work has focused on the problem of producing small representations of graphs that multiplicatively approximate the edge-cut function (i.e., the number of edges cut) for all input cuts [Karger, 1993, Benczúr and Karger, 1996, Spielman and Srivastava, 2008, Spielman and Teng, 2011, Allen-Zhu et al., 2015, Lee and Sun, 2018]. Batson et al. [2012] obtained sketches for edge cuts using $O(\frac{n \log n}{\varepsilon^2})$ bits; this is also known to be optimal [Carlson et al., 2019]. It is a natural question to seek similar-sized sketches for the vertex boundary function. We will answer this question in the negative, showing that $\Omega(n^2)$ bits are required for a multiplicative error.

Sketching coverage functions. Coverage functions are a special case of submodular functions, which have attracted great interest in learning theory [Bach, 2013, Feldman and Kothari, 2014, Chakrabarty and Huang, 2015, Yang et al., 2021] and optimization [Balcan and Harvey, 2018]. Moreover, optimization problems on coverage functions entail classical algorithmic problems such as max-cover and set-cover as special cases. From the sketching perspective, Badanidiyuru et al. [2012] showed that coverage functions can be sketched to within a multiplicative $(1 \pm \varepsilon)$ error with bit complexity $\tilde{O}(n^3/\varepsilon^2)$, and Bateni et al. [2017] provided a sketch with $\tilde{O}(nk)$ bits for the special case of k-cover. It is possible to extend the latter to obtain a sketch with $\tilde{O}(n^2)$ bits for general coverage functions. Randomized sketches minimizing the expected squared error have also been studied [Yaroslavtsev and Zhou, 2019]—we do not consider this error in our paper. To the best of our knowledge, no non-trivial lower bound for sketching coverage functions was known in the literature.

Relationship to ℓ_1-distance bounds for Random Utility Models. While our approach shares the high-level packing argument used for the ℓ_1-distance sketching lower bound by Chierichetti et al. [2021] for Random Utility Models—generating $\exp(n^2)$ instances and proving they are pairwise far apart—there is an important technical difference in how the separation is shown. They showed the separation using a fixed class of $\Theta(n)$ *test sets*, guaranteeing that any pair of RUMs will behave differently on at least one of these fixed subsets of the universe. In contrast, for the ℓ_∞-distance, fixing m test sets *a priori* allows the distribution to be approximated with only $O(n \log(n) \log(m))$ bits—which precludes an $\Omega(n^2)$ lower bound unless m is exponentially large! Hence, our proof requires choosing the $\Theta(n)$ test sets adaptively based on the structure of the sampled instances to guarantee the necessary separation probability.

Relationship to VC-dimension bounds. Consider the class of functions $f_S : 2^{[n]} \to \{0, 1\}$ defined as $f_S(T) = [S \cap T \neq \varnothing]$, where $[\cdot]$ indicates the 0-1 indicator. One can easily show that $\mathcal{F} = \{f_S \mid S \subseteq [n]\}$ has VC-dimension, $\text{VCdim}(\mathcal{F}) \geq n$: the set $\{\{1\}, \ldots, \{n\}\}$ is shattered by $\mathcal{F}$ since for any $\{0, 1\}$-labeling of the n points, we can find a function in $\mathcal{F}$ that produces that labeling. On the other hand, $\text{VCdim}(\mathcal{F}) \leq \log_2 |\mathcal{F}|$; hence, $\text{VCdim}(\mathcal{F}) = n$. Now, the class of intersection profiles of power-set distributions, $\mathcal{I}$, is the convex hull of $\mathcal{F}$. Any threshold class of $\mathcal{I}$ must then have VC dimension at least n. The fat-shattering dimension of the real-valued hypothesis class $\mathcal{I}$ at any constant scale also grows like $\Theta(n)$. These bounds can be used to show that $O(n)$ samples (i.e., $O(n^2)$ bits) are enough to learn the behavior of the unknown hypothesis to within a constant error.

Unfortunately, these results do not yield a lower bound on the total number of bits. Indeed, the fact that sketching a single sample requires $\Omega(n)$ bits does not imply that a multiset of k samples requires $\Omega(nk)$ bits. (For example, if $k = 4^n$, one could sketch a multiset of k samples using only $O(n2^n) \leq O(n\sqrt{k})$ bits—for each set $S \subseteq [n]$, just store how many of the k samples intersect S; the bit cost for any of the 2^n many S's is then $O(\log k) = O(n)$.) To prove the $\Omega(n^2)$ lower bound, a different approach is required, which we develop in this paper.

Relationship to communication complexity bounds. The results proved in this paper immediately imply lower bounds on the one-way communication complexity of computing the intersection profile of a power-set distribution. We obtain our results using very simple machinery.

3 Background: Sketching and Intersection Profiles

We begin by reviewing the notion of sketching. In the typical sketching setting, one is interested in producing a succinct representation of an object from a fixed class in a way that allows them to approximately answer certain questions of interests about the object by looking solely at this representation. As is customary, given sets A and B we will denote by A^B the collection of functions from B to A. We will reserve the notation 2^A for the power set of A: the collection of all subsets of A. For a proposition P, we will let $[P]$ be 1 if P is true and 0 otherwise.

Definition 1 (Sketch)**.** Let $\mathcal{F} \subseteq \mathbb{R}^Q$ be a class of real-valued functions on a set Q. A *sketch* for $\mathcal{F}$ is a pair (ϕ, ψ) of functions with $\phi : \mathcal{F} \to \{0,1\}^\kappa$ and $\psi : \{0,1\}^\kappa \times Q \to \mathbb{R}$, where $\kappa \in \mathbb{N}$ is a parameter called the *bit complexity* of the sketch. A sketch (ϕ, ψ) is *ε-multiplicative* if it satisfies:

$$\forall f \in \mathcal{F}, \forall q \in Q: \ |\psi(\phi(f), q) - f(q)| \leq \varepsilon \cdot |f(q)|.$$

A sketch (ϕ, ψ) is *ε-additive* if it satisfies:

$$\forall f \in \mathcal{F}, \forall q \in Q: \ |\psi(\phi(f), q) - f(q)| \leq \varepsilon.$$

Intuitively, given a function f, one wants to produce a short bit string $\phi(f)$ that allows them to approximate the value of $f(q)$ for every possible element $q \in Q$ (sometimes referred to as a *query*).

We will refer to a probability distribution supported on $2^{[n]}$ as a *power-set distribution* on $[n]$, or simply a *power-set distribution*, and we will denote by $\mathcal{D}_n$ the collection of all power-set distributions on $[n]$. The central result of this paper concerns the sketching of intersection profiles of power-set distributions, defined as follows.

Definition 2 (Intersection Profile)**.** For a power-set distribution $D \in \mathcal{D}_n$, its *intersection profile* is the function $F_D \in [0,1]^{2^{[n]}}$ defined as

$$F_D(S) = \Pr_{T \sim D}[T \cap S \neq \varnothing].$$

Note that the intersection profile is closely related to the *zeta transform* ζ_D of the probability mass function (see, e.g., Koivisto and Röyskö [2020]), given by:

$$\zeta_D(S) = \Pr_{T \sim D}[T \subseteq S].$$

In fact, it is easy to see that $F_D(S) = 1 - \zeta_D([n] \setminus S)$, and hence, any additive approximation to $F_D(\cdot)$ will yield an additive approximation to $\zeta_D(\cdot)$. For simplicity, we will work with the intersection profile of a distribution instead of its zeta transform.

We study the problem of additively sketching the class of intersection profiles of power-set distributions: $\mathcal{F}_n = \{F_D(\cdot) \mid D \in \mathcal{D}_n\}$.

For convenience, for two power-set distributions D, D', we define the distance $d_\infty(D, D')$ based on the ℓ_∞-distance between their intersection profiles:

$$d_\infty(D, D') := \|F_D - F_{D'}\|_\infty = \max_{S \subseteq [n]} |F_D(S) - F_{D'}(S)|.$$

4 Sketching Intersection Profiles of Power-Set Distributions

We first present an upper bound for sketching $\mathcal{F}_n$. Specifically, we begin by noting the following result, which follows by a simple application of the Chernoff–Hoeffding bound.

Lemma 3. *For each $\varepsilon > 0$ and for every power-set distribution D, there is a power-set distribution D' that is uniform*[1] *over a support of size $O\left(n/\varepsilon^2\right)$ and such that $d_\infty(D, D') \leq \varepsilon$.*

Proof. Fix $t = \left\lceil \frac{n}{\varepsilon^2} \right\rceil$ and let $T_1, \ldots, T_t$ be i.i.d. samples from D. Define D' as the power-set distribution that samples i uniformly at random from $[t]$ and returns T_i.

Consider any $S \subseteq [n]$. For each $i \in [t]$, $\Pr[T_i \cap S \neq \varnothing] = F_D(S)$. Then, $\mathrm{E}\left[F_{D'}(S)\right] = F_D(S)$, where the expectation is over the random choices of $T_1, \ldots, T_t$. Moreover, by a Chernoff–Hoeffding bound (see, e.g., [Dubhashi and Panconesi, 2009, Theorem 1.1]):

$$\Pr\left[|F_{D'}(S) - \mathrm{E}[F_{D'}(S)]| \geq \varepsilon\right] \leq 2e^{-2\varepsilon^2 t} \leq 2e^{-2n}.$$

By a union bound:

$$\Pr\left[\exists S \subseteq [n] : |F_{D'}(S) - \mathrm{E}[F_{D'}(S)]| \geq \varepsilon\right] \leq 2^n \cdot 2e^{-2n} = 2 \cdot \left(\frac{2}{e^2}\right)^n < \frac{2}{3},$$

since $n \geq 1$. Thus, with probability at least $1/3$, D' satisfies $d_\infty(D, D') \leq \varepsilon$. □

In particular, we obtain an upper bound on the bit complexity of sketches for $\mathcal{F}_n$.

Theorem 4. *There exists an ε-additive sketch for $\mathcal{F}_n = \{F_D(\cdot) \mid D \in \mathcal{D}_n\}$ of bit complexity $\kappa = O(n^2/\varepsilon^2)$.*

Proof. Given D, one can construct the distribution D' guaranteed by Lemma 3 and store the support of D'. Each set in the support can be represented as an n-bit string. □

Our main technical result is that this sketch is optimal in n.

Theorem 5. *Any 0.05-additive sketch for $\mathcal{F}_n = \{F_D(\cdot) \mid D \in \mathcal{D}_n\}$ has bit complexity $\Omega(n^2)$.*

We prove Theorem 5 in the following subsection. Note that our construction works for other constants rather than 0.05, provided they are small enough. However, our construction does not give a lower bound on the bit complexity that scales inversely with this value. Specifically, our proof can be used to show that the bit complexity of an ε-additive sketch, for $\varepsilon < 1/8$, is at least $\frac{1}{4}(\frac{1}{4} - 2\varepsilon)^2 \log_2(e) \cdot n^2$. Therefore, proving a lower bound of $\Omega(n^2/\varepsilon^2)$ for any ε-additive sketch may require some new ideas and for simplicity we restricted here to a fixed constant.

4.1 Proof of the Lower Bound (Theorem 5)

The main idea required to prove Theorem 5 is to construct $2^{\Omega(n^2)}$ power-set distributions that are pairwise far apart. In particular, the key technical step is the following result.

Theorem 6. *For every n, there exists a collection $D_1, \ldots, D_K$ of $K = K(n) = 2^{\Omega(n^2)}$ power-set distributions such that $d_\infty(D_i, D_j) > 0.1$ for each $\{i, j\} \in \binom{[K]}{2}$.*

Before we proceed with its proof, we note that Theorem 5 follows directly from Theorem 6 by a standard argument, which we now give for completeness.

[1] Note that, in the statement of this lemma, we technically allow the support of D' to be a multiset, and D' will actually select an element from its support with probability proportional to the number of times it appears in its support. This causes no issue in the application of this result to the next theorem.

Proof that Theorem 5 follows from Theorem 6. By contradiction, suppose there is a 0.05-additive sketch (ϕ, ψ) for $\mathcal{F}_n$ with bit complexity $\kappa < \log_2 K$, where K is chosen as in the statement of Theorem 6. Now, consider a collection of K set intersection instances $D_1, \ldots, D_K$ with the property guaranteed by Theorem 6. By the pigeonhole principle, there exist D_i and D_j with $i \neq j$ such that $\phi(D_i) = \phi(D_j)$. Also, since $d_\infty(D_i, D_j) > 0.1$ there exists some subset $T \subseteq [n]$ such that $|F_{D_i}(T) - F_{D_j}(T)| > 0.1$.

On the other hand, by the triangle inequality,

$$\begin{aligned}|F_{D_i}(T) - F_{D_j}(T)| &\leq |F_{D_i}(T) - \psi(\phi(D_i), T)| + |\psi(\phi(D_i), T) - F_{D_j}(T)| \\ &= |F_{D_i}(T) - \psi(\phi(D_i), T)| + |\psi(\phi(D_j), T) - F_{D_j}(T)| \\ &\leq 0.05 + 0.05 = 0.1, \text{ a contradiction.}\end{aligned}$$

Hence, every 0.05-additive sketch for $\mathcal{F}_n$ must have bit complexity $\kappa \geq \log_2 K = \Omega(n^2)$, completing the proof. □

We now prove Theorem 6. We do so via a probabilistic packing argument, constructing a family of distributions that are pairwise far apart in the $d_\infty(\cdot, \cdot)$-metric. The construction embeds a random binary matrix into the problem by partitioning the universe items into one of two types: *high-degree* items and *low-degree* items. High-degree items appear in roughly half the sets of our power-set distribution (i.e., they have high degree in the items-sets bipartite graph) and encode the high-entropy random bits of the matrix. Low-degree items appear in exactly one set each (i.e., they have low degree in the items-sets bipartite graph), and help identify the set uniquely. A distribution consisting only of high-degree items could be sketched with few bits because, with high probability, for all sets $S \subseteq [n]$, the set S will have empty intersection with a sampled set with probability $\approx 2^{-|S|}$. On the other hand, a distribution consisting only of low-degree items could be sketched using $O(n \log n)$ bits, since each set could be represented exactly.

To prove our $\Omega(n^2)$ lower bound, we leverage the interplay of the high-degree and low-degree items. The two types of items together make it possible to create $\Theta(n)$ queries such that, for the sketch to answer each of those queries, it has to hold a very good approximation of the neighborhoods of the high-degree items in the items-sets bipartite graph.

Proof of Theorem 6. For simplicity, we assume n is even; let $m = n/2$. Each distribution is associated with a random binary matrix $M \in \{0,1\}^{m \times m}$, where each entry $M_{i,j} \sim \text{Bernoulli}(1/2)$, and is independent of all other entries. The matrix M is used to define two quantities: (i) the distribution D_M and (ii) a family of queries. Formally, for any given $M \in \{0,1\}^{m \times m}$, we let D_M be the uniform distribution on m sets $S_1^{(M)} \ldots, S_m^{(M)} \subseteq [n]$ that, intuitively, correspond to the rows of M paired with an identity matrix. Specifically, for each $i \in [m]$, let

$$S_i^{(M)} = \{j \in [m] \mid M_{i,j} = 1\} \cup \{m + i\}.$$

The query family $T_1^{(M)}, \ldots, T_m^{(M)} \subseteq [n]$ is, intuitively, given by the columns of an identity matrix paired with the complement of M. Specifically, for each $j \in [m]$, let

$$T_j^{(M)} = \{j\} \cup \{m + i \mid i \in [m] \wedge M_{i,j} = 0\}.$$

We show an example in Figure 1. We start by showing some key properties of this construction.

Lemma 7. *Let A and B be two (not necessarily distinct) matrices in $\{0,1\}^{m \times m}$, and let $S_i^{(A)}$ and $T_j^{(B)}$ be defined as above. For any $i, j \in [m]$, $S_i^{(A)} \cap T_j^{(B)} \neq \varnothing$ if and only if $A_{i,j} = 1$ or $B_{i,j} = 0$.*

$$
[M \mid I_m] = \left(\begin{array}{cccc|cccc} M_{1,1} & M_{1,2} & \cdots & M_{1,m} & 1 & 0 & \cdots & 0 \\ M_{2,1} & M_{2,2} & \cdots & M_{2,m} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ M_{m,1} & M_{m,2} & \cdots & M_{m,m} & 0 & 0 & \cdots & 1 \end{array}\right)
$$

$$
[I_m|\mathbf{1}\mathbf{1}^\top - M^\top] = \left(\begin{array}{cccc|cccc} 1 & 0 & \cdots & 0 & 1-M_{1,1} & 1-M_{2,1} & \cdots & 1-M_{m,1} \\ 0 & 1 & \cdots & 0 & 1-M_{1,2} & 1-M_{2,2} & \cdots & 1-M_{2,m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 1-M_{1,m} & 1-M_{2,m} & \cdots & 1-M_{m,m} \end{array}\right)
$$

Figure 1: First, the $m \times n$ matrix $[M \mid I_m]$ with $M_{i,j} \sim \text{Bernoulli}(1/2)$. Each row i of this matrix is the indicator vector of the set $S_i^{(M)}$. The power-set distribution D_M associated with the matrix M selects a subset by sampling a row of this matrix uniformly at random, and interpreting the row as the indicator of the corresponding subset of $[n] = [2m]$. Below that, the matrix $[I_m|\mathbf{1}\mathbf{1}^\top - M^\top]$. Each row j of this matrix, is the indicator of the query set $T_j^{(M)}$

Proof of Lemma 7. Note that the only element of $S_i^{(A)} \cap \{m+1, \ldots, 2m\} = m+i$ and $T_j^{(B)} \cap \{1, \ldots, m\} = j$ and hence $S_i^{(A)} \cap T_j^{(B)} \subseteq \{j, m+i\}$. In particular, $S_i^{(A)} \cap T_j^{(B)}$ is non-empty if and only if either $j \in S_i^{(A)}$ or $m+i \in T_j^{(B)}$. By construction, this happens if and only if $A_{i,j} = 1$ or $B_{i,j} = 0$. □

This leads to two useful consequences.

Corollary 8. *Let D_A be any power-set distribution induced by a binary matrix A as above. Then, for any query $T_j^{(A)}$, we have $F_{D_A}(T_j^{(A)}) = 1$.*

Proof of Corollary 8. Using Theorem 7 with $B = A$, we obtain $S_i^{(A)} \cap T_j^{(A)} \neq \varnothing$ always. Hence,

$$
F_{D_A}(T_j^{(A)}) = \frac{1}{m} \sum_{i=1}^{m} \underbrace{\left[S_i^{(A)} \cap T_j^{(A)} \neq \varnothing\right]}_{\text{binary indicator}} = 1. \qquad \square
$$

Corollary 9. *For any $i, j \in [m]$:*

$$
\Pr_{A,B}[S_i^{(A)} \cap T_j^{(B)} \neq \varnothing] = 3/4,
$$

where the entries of A and B are sampled independently from Bernoulli(1/2).

Proof of Corollary 9. Using Theorem 7 and the independence of A and B,

$$
\Pr_{A,B}[S_i^{(A)} \cap T_j^{(B)} = \varnothing] = \Pr_{A,B}[A_{i,j} = 0 \text{ and } B_{i,j} = 1] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}. \qquad \square
$$

Let D_A and D_B be two independent instances defined by matrices A and B. Let $j \in [m]$ be a fixed column and let Y_j be the random variable representing the value of the intersection profile of

D_A on the query $T_j^{(B)}$ (note that this random variable is measurable in the σ-algebra generated by the jth columns of A and B, and its value is determined once A and B are fixed):

$$Y_j = F_{D_A}(T_j^{(B)}) = \frac{1}{m}\sum_{i=1}^{m}\left[S_i^{(A)} \cap T_j^{(B)} \neq \varnothing\right] = \frac{1}{m}\sum_{i=1}^{m}\left[A_{i,j} = 1 \text{ or } B_{i,j} = 0\right].$$

From this, Y_j is the average of $m = n/2$ i.i.d. Bernoulli random variables of parameter $3/4$ (Corollary 9). By a Hoeffding bound (see, e.g., [Dubhashi and Panconesi, 2009, Theorem 1.1]), we have:

$$\Pr_{A,B}\left[Y_j \geq 0.9\right] = \Pr_{A,B}\left[Y_j - \mathbb{E}[Y_j] \geq \frac{3}{20}\right] \leq \exp\left(-\frac{9n}{400}\right).$$

Note that if $Y_j < 0.9$, then, by Theorem 8, the distance between the intersection profiles on this specific query $T_j^{(B)}$ is:

$$|F_{D_B}(T_j^{(B)}) - F_{D_A}(T_j^{(B)})| = |1 - Y_j| > 0.1.$$

Now, the probability that the distance between D_A and D_B is less than 0.1 on all queries $T_1^{(B)}, \ldots, T_m^{(B)}$ is given by:

$$\Pr_{A,B}[d_\infty(D_A, D_B) \leq 0.1] \leq \prod_{j=1}^{m} \Pr[Y_j \geq 0.9] \leq \prod_{j=1}^{m} \exp\left(-\frac{9n}{400}\right) = \exp\left(-\frac{9n^2}{800}\right),$$

where we are crucially using the fact that, by Theorem 7, each variable Y_j is a function of the jth column of A and B only, and hence all the Y_j's are mutually independent.

Applying a union bound over all pairs in a collection of K distributions $D_1, \ldots, D_K$ corresponding to matrices $M_1, \ldots, M_K$ sampled independently, we obtain:

$$\Pr_{M_1,\ldots,M_K}\left[\exists\{i,j\} \in \binom{[K]}{2} \;\middle|\; d_\infty(D_{M_i}, D_{M_j}) \leq 0.1\right] \leq \binom{K}{2}\exp\left(-\frac{9n^2}{800}\right) < 1,$$

for $K \leq 2^{\left(\frac{9\log_2 e}{1600}n^2\right)}$, completing the proof. □

5 Application: Sketching Vertex Neighborhoods

We now turn to applications of our results. The first application consists in proving lower bounds on sketching vertex neighborhoods. For any unweighted undirected graph $G = (V, E)$ and any $S \subseteq V$, the *cut set* of S is the set of edges with exactly one endpoint in S: $\mathrm{cut}(S) = \{e \mid e \in E \text{ and } |e \cap S| = 1\}$. It is known (Batson et al. [2012], Carlson et al. [2019]) that one can obtain an ε-multiplicative sketch for the function $|\mathrm{cut}(\cdot)|$ with bit complexity $\Theta(n \log n/\varepsilon^2)$ and that this bound is optimal.

Given a graph $G(V, E)$, for a subset $S \subseteq V$, we define $\Gamma_G(S)$ to be the set of *neighbors* of S: $\Gamma_G(S) = \{w \mid w \in V \text{ and } \exists v \in S \text{ such that } \{v, w\} \in E\}$. Observe that the adjacency matrix of the full graph, which requires $\Theta(n^2)$ bits, is a perfect (0-additive) sketch for $|\Gamma_G(\cdot)|$. In contrast to the (edge-based) cut function, we show by using Theorem 5 that this simple representation is asymptotically optimal.

We will actually show that the bit complexity is $\Omega(n^2)$ even if we restrict ourselves to only sketching bipartite graphs $G = (V_1 \cup V_2, E)$, and we only allow queries $S \subseteq V_1$. Specifically, let $\mathcal{B}_{n,m}$ be the set of all bipartite graphs with $|V_1| = n$ and $|V_2| = m$. For a bipartite graph $G = (V_1 \cup V_2, E)$ we restrict the domain of the neighborhood function to vertices of V_1, that is: $\Gamma_G : 2^{V_1} \to 2^{V_2}$.

Theorem 10. *Let $\alpha \le 1/60$ be a constant. There exists $m = \Theta(n)$ such that the bit complexity of any $\alpha(n+m)$-additive sketch for $\{|\Gamma_G(\cdot)| \mid G \in \mathcal{B}_{n,m}\}$ is $\Omega(n^2)$.*

Proof. Let D be any power-set distribution over $[n]$. By Theorem 3, there exists a power-set distribution D' over $[n]$ that is uniform over a support of $N = \Theta(n)$ sets and such that $|F_D(S) - F_{D'}(S)| \le \alpha$ for each $S \subseteq [n]$. Note that $N \ge n$.

Suppose there exists an $\alpha(n+N)$-additive sketch (ϕ, ψ) for $\{|\Gamma_G(\cdot)|\}_{G \in \mathcal{B}_{n,N}}$ with bit complexity κ. Specifically, for any bipartite graph $G = (V_1 \cup V_2, E) \in \mathcal{B}_{n,N}$, let $\Delta_G(\cdot) = \psi(\phi(G), \cdot)$. Then, $||\Gamma_G(S)| - \Delta_G(S)| \le \alpha \cdot (n+N)$ for each $S \subseteq V_1$.

Let the support of D' be $\mathcal{S} = \{S_1, \ldots, S_N\}$, where $S_i \subseteq [n]$ for each i. We build the following bipartite graph $G = (V_1 \cup V_2, E)$, with $V_1 = [n]$ and $V_2 = \mathcal{S}$. For each $x \in V_1$ and $S \in V_2$, we add the edge $\{x, S\}$ to E if and only if $x \in S$. Observe that, for each $S \subseteq [n]$, it holds that:

$$|\Gamma_G(S)| = \sum_{i \in [N]} [S \cap S_i \neq \varnothing] = N \cdot F_{D'}(S).$$

Then, for each $S \subseteq [n]$, we have:

$$\left| \frac{\Delta_G(S)}{N} - \frac{|\Gamma_G(S)|}{N} \right| \le \frac{\alpha \cdot (n+N)}{N} \le 2 \cdot \alpha.$$

Putting it all together, we have that for all $S \subseteq [n]$:

$$\left| \frac{\Delta_G(S)}{N} - F_D(S) \right| \le \left| \frac{\Delta_G(S)}{N} - \frac{|\Gamma_G(S)|}{N} \right| + \left| \frac{|\Gamma_G(S)|}{N} - F_{D'}(S) \right| + |F_{D'}(S) - F_D(S)| \le 3\alpha \le \frac{1}{20}.$$

This immediately implies a $(1/20)$-additive sketch $(\phi, \psi(\cdot,\cdot)/N)$ for $\{F_D\}_{D \in \mathcal{D}_n}$ with bit complexity $\kappa + O(\log n)$. Therefore, by Theorem 5, the bit complexity of the sketch for $\{|\Gamma_G(\cdot)|\}_{G \in \mathcal{B}_{n,N}}$ must be $\kappa = \Omega(n^2)$. □

This result immediately yields a lower bound on the bit complexity of ε-multiplicative sketches of neighborhood cardinalities for arbitrary undirected graphs. In particular, let $\mathcal{G}_n$ be the set of all undirected graphs on n vertices, we have the following:

Corollary 11. *Let $\alpha \le 1/60$. Then, any α-multiplicative sketch for $\{|\Gamma_G(\cdot)| \mid G \in \mathcal{G}_n\}$ has bit complexity $\Omega(n^2)$.*

Moreover, since our construction applies to sketching the neighborhood cardinalities of sets on one side of bipartite graphs, the lower bound applies to the problem of sketching outer neighborhood cardinalities. That is, given an undirected graph G let:

$$\Gamma_G^{\text{out}}(S) := \Gamma_G(S) \setminus S = \{v \in V \setminus S \mid \exists u \in S, \{u, v\} \in E\},$$

then, we have the following result.

Corollary 12. *Let $\alpha \le 1/60$. Then, any α-multiplicative sketch for $\{|\Gamma_G^{out}(\cdot)| \mid G \in \mathcal{G}_n\}$ has bit complexity $\Omega(n^2)$.*

6 Applications: Sketches for Coverage Functions

We now explore the second key application of our result: sketching coverage functions. We recall the key definitions.

Definition 13 (Normalized Coverage Function)**.** Given a universe U of cardinality m and a collection $S_1, \ldots, S_n \subseteq U$ of subsets, the *normalized coverage function* $f : 2^{[n]} \to [0, 1]$ associated with $S_1, \ldots, S_n$ and the universe U is defined as

$$f(A) = \frac{1}{m} \cdot \left| \bigcup_{i \in A} S_i \right|,$$

where $A \subseteq [n]$. Given a collection of positive integer weights $W = \{w_u\}_{u \in U}$ each associated with an element of the universe U, the *weighted coverage function* f_W associated with $S_1, \ldots S_n$ and the universe U is defined as:

$$f_W(A) = \frac{1}{w(U)} \cdot w \left(\bigcup_{i \in A} S_i \right)$$

where, for any subset $B \subseteq U$, we define:

$$w(B) := \sum_{i \in B} w_i.$$

We consider the problem of sketching the family of all normalized coverage functions. Specifically, let $\mathcal{C}_{n,m}$ be the set of all normalized coverage functions with universe size m and a collection of n subsets. We show that the bounds obtained thanks to our Theorem 5 are tight both for additive error sketches and for multiplicative error sketches (in the latter case, up to polylog factors).

6.1 Additive-Error Sketch

By using the vertex neighborhood sketching lower bound we directly obtain a lower bound of $\Omega(n^2)$ for sketching coverage functions:

Theorem 14. *Let $\alpha \leq 1/60$ be a constant. There exists $m = \Theta(n)$ such that any α-additive sketch for the class of normalized coverage functions $\mathcal{C}_{n,m}$ requires bit complexity $\Omega(n^2)$.*

Proof. Consider any bipartite graph $G = (V_1 \cup V_2, E)$, where for simplicity we set $V_1 = [n]$ and $V_2 = [m]$. We consider the normalized coverage function f_G with universe $[m]$ and collection size n. Specifically, for each $i \in [n]$, we add the following S_i to the collection of sets:

$$S_i = \{j \in [m] \mid \{i, j\} \in E\}.$$

Note that, for each $A \subseteq [n]$, we have:

$$f_G(A) = \frac{1}{m} \left| \bigcup_{i \in A} S_i \right| = \frac{|\Gamma_G(A)|}{m}.$$

Suppose there is a α-additive sketch (ϕ, ψ) for $\mathcal{C}_{n,m}$ with bit complexity κ. Specifically, let $\Delta_f(\cdot) = \psi(\phi(f_G), \cdot)$. By definition, for each $A \subseteq [n]$, it holds that $|f_G(A) - \Delta_f(A)| \leq \alpha$. Then, for each

$A \subseteq [n]$, we have:

$$\begin{aligned}
||\Gamma_G(A)| - m \cdot \Delta_f(A)| &= m \cdot \left| \frac{|\Gamma_G(A)|}{m} - \Delta_f(A) \right| \\
&\leq m \cdot \left(\left| \frac{|\Gamma_G(A)|}{m} - f_G(A) \right| + |f_G(A) - \Delta_f(A)| \right) \\
&\leq \alpha \cdot m \\
&\leq \alpha \cdot (n+m).
\end{aligned}$$

Therefore, $(\phi, m \cdot \psi(\cdot,\cdot))$ is an $\alpha(n+m)$-additive sketch with bit complexity $\kappa + O(\log m)$ for the set of bipartite graphs with $|V_1| = n$ and $|V_2| = m$. Therefore, by Theorem 10, for some $m = \Theta(n)$ the bit complexity of (ϕ, ψ) must be $k = \Omega(n^2)$. □

Moreover, a simple application of the Hoeffding bound, shows that our lower bound is tight for additive errors, a result which we now show.

Theorem 15. *For every $\varepsilon > 0$, and every normalized coverage function f, there exists a normalized coverage function f' defined on a universe of cardinality $t = O\left(n/\varepsilon^2\right)$ such that for each $A \subseteq [n]$:*

$$|f(A) - f'(A)| \leq \varepsilon.$$

Thus, there is an ε-additive sketch for $\mathcal{C}_{n,m}$ with bit complexity $O(n^2/\varepsilon^2)$.

Proof. Without loss of generality, we assume $f([n]) = 1$, since any element in U not contained in any subset contributes nothing to the coverage function and can be removed without affecting the value of $f(A)$ for any A.

Fix $t = \left\lceil \frac{(n+2)\ln 2}{2\varepsilon^2} \right\rceil$. If $m \leq t$, the statement trivially follows since the total space to represent the original instance is $O(mn) \leq O(n^2/\varepsilon^2)$. Therefore, we assume $m > t$.

We construct a sub-universe $T \subseteq U$ by selecting t elements from U uniformly at random, without replacement. Now, consider the (random) normalized coverage function f' associated with the collection $S_1 \cap T, \ldots, S_n \cap T$ and the universe T, defined as

$$f'(A) = \frac{1}{t} \cdot \left| \bigcup_{i \in A} S_i \cap T \right|.$$

For each $A \subseteq [n]$, we have $\mathrm{E}\left[f'(A)\right] = f(A)$, where the expectation is over the choice of T. Applying a Hoeffding bound for sampling without replacement (see, e.g., [Serfling, 1974]), we can bound the deviation from the mean as:

$$\Pr[|f'(A) - f(A)| \geq \varepsilon] \leq 2\exp\left(-2\varepsilon^2 t\right) \leq 2^{-(n+1)},$$

by our choice of t.

Applying a union bound over all $A \subseteq [n]$, we obtain

$$\Pr\left[\exists A \subseteq [n] : \left|f'(A) - f(A)\right| \geq \varepsilon\right] \leq 2^n \cdot 2^{-(n+1)} = \frac{1}{2},$$

since $n \geq 1$. Thus, $|f'(A) - f(A)|_\infty < \varepsilon$ for each $A \subseteq [n]$, with probability at least $1/2$.

To store f', notice that we only need to store $S_1 \cap T, \ldots, S_n \cap T$. Since $|T| = t$, this uses space $t \cdot n = O(n^2/\varepsilon^2)$ bits. □

6.2 Multiplicative-Error Sketch

It is easy to see that Theorem 14 also implies that $\Omega(n^2)$ bits are necessary for an ε-multiplicative sketch of the set of all coverage functions $\mathcal{C}_{n,m}$, for a fixed constant $\varepsilon > 0$. Indeed, an ε-multiplicative sketch also implies an ε-additive sketch, since $f(A) \leq 1$ for each $A \subseteq [n]$.

In terms of upper bounds, Badanidiyuru et al. [2012] provide an ε-multiplicative sketch that uses $\tilde{O}(n^3/\varepsilon^2)$ bits.[2] Bateni et al. [2017] provide a multiplicative sketch for a coverage function to approximate the k-cover problem with bit complexity $\tilde{O}(nk)$. Their sketch is based on classical results to approximate the number of distinct elements in a data stream [Bar-Yossef et al., 2002, Cormode et al., 2002], and can be generalized to apply to general coverage functions with bit complexity $\tilde{O}(n^2)$ when $m = n^{O(1)}$. For completeness, in Appendix A, we provide a proof of the following result:

Theorem 16. *Fix any $\varepsilon \in (0,1)$. There exists an ε-multiplicative sketch for the set of all normalized covering functions $\mathcal{C}_{n,m}$ with bit complexity $O(\frac{n^2}{\varepsilon^2}\log m)$.*

When the weights are positive integers, it is possible to extend this upper bound to weighted coverage functions. Specifically, let $\mathcal{C}^W_{n,m}$ be the set of all weighted coverage functions with positive integer weights in $W = \{w_u\}_{u\in U}$.

Corollary 17. *For any $\varepsilon \in (0,1)$ and positive integer weights $W = \{w_u\}_{u\in U}$, the set of weighted coverage functions $\mathcal{C}^W_{n,m}$ can be ε-multiplicatively sketched with bit complexity $O(\frac{n^2}{\varepsilon^2}\cdot\log(w(U)))$.*

Proof. Consider any $f_W \in \mathcal{C}^W_{n,m}$. We show this result by reducing f_W to an unweighted coverage function f.

We build a new universe U' as follows: for each $u \in U$ with positive integer weight $w(u)$, we add elements $u_1, \ldots, u_{w(u)}$ to U'. Similarly, for each $i \in [n]$, we build S'_i from S_i by replacing each $u \in S_i$ with $u_1, \ldots, u_{w(u)}$. Let f be the normalized coverage function induced by universe U' and the collection $S'_1, \ldots, S'_n$. Note that, for each $A \subseteq [n]$, we have:

$$f_W(A) = \frac{1}{w(U)} \cdot w\left(\bigcup_{i\in A} S_i\right) = \frac{1}{|U'|}\cdot \sum_{u\in\cup_{i\in A}S_i} w(u) = \frac{1}{|U'|}\cdot\left|\bigcup_{i\in A} S'_i\right| = f(A).$$

By Theorem 16, $\mathcal{C}_{n,|U'|}$ can be sketched with $O(\frac{n^2}{\varepsilon^2}\log|U'|) = O(\frac{n^2}{\varepsilon^2}\log(w(U)))$ bits which immediately implies a sketch for $\mathcal{C}^W_{n,m}$. □

Note that it is common in the literature to assume that $m = n^{O(1)}$ and that $\max_{u\in U} w(u) = n^{O(1)}$ (e.g., [Badanidiyuru et al., 2012]): in this setting, the previous algorithms provide a sketch with bit complexity $O((n/\varepsilon)^2\log n)$ for $\mathcal{C}^W_{n,m}$.

7 Application: ℓ_∞-Sketching of RUMs

In this section, we describe the implication of our main theorem to the problem of sketching the winning distributions of Random Utility Models (RUMs). A *RUM* on $[n]$ is a probability distribution R over the set of permutations on $[n]$. Given a non-empty $T \subseteq [n]$, we use R_T to denote the distribution of the random variable $\pi(T)$ for $\pi \sim R$, where $\pi(T) = \arg\max_{i\in T}\pi(i)$. We

[2] Badanidiyuru et al. [2012] do not study the bit complexity explicitly, but provide a weighted coverage function with universe size $O(n^2/\varepsilon^2)$ and weights of polynomial size in n that approximates the original coverage function; therefore, a simple upper bound on their bit complexity is $\tilde{O}(n^3/\varepsilon^2)$.

use the terms *slate* to denote T, *winner* to denote $\pi(T)$, and *winning distribution* to denote R_T. For $i \in T$, $R_T(i)$ is the probability that item i is the winner according to a random permutation from R, among the items in T.

Let $\mathcal{R}_n$ be the set of all RUMs over $[n]$. Here, we consider the problem of sketching the winning distributions of all RUMs in $\mathcal{R}_n$ in norm ℓ_p, for $p \in [1, \infty)$. Specifically, for each RUM R over $[n]$, we wish to construct a data structure $\hat{R}$ that can provide an estimate $\hat{R}_T$ of R_T satisfying: $\|R_T - \hat{R}_T\|_p$, for each slate $T \subseteq [n]$. Chierichetti et al. [2021] showed that $\mathcal{R}_n$ can be sketched in ℓ_1-norm with bit complexity $O((n/\varepsilon)^2 \log n)$. They also provided a lower bound of $\Omega(n^2)$ for ℓ_1, but left open the problem for ℓ_∞, for which the best previously known lower bound was $\Omega(n \log n)$. Below, we use Theorem 5 to improve this lower bound to $\Omega(n^2)$. Together with the upper bound on ℓ_1, this shows that sketching $\mathcal{R}_n$ under any ℓ_p-norm, $p \in [1, \infty) \cup \{\infty\}$, requires $\tilde{\Theta}(n^2)$ bits.

In fact, we prove a stronger result: we show that even if we fix a special item $i^* \in [n]$ and we are only interested in sketching $f_R^{i^*}(T) := R_{T \cup \{i^*\}}(i^*)$ for each slate $T \subseteq [n] \setminus \{i^*\}$, then $\Omega(n^2)$ bits are necessary.

Theorem 18. *Let $\alpha \le 1/20$ and $i^* \in [n]$. Any α-additive sketch for the class $\{f_R^{i^*}\}_{R \in \mathcal{R}_n}$ has bit complexity $\Omega(n^2)$.*

Proof. We prove this result by reducing the problem of sketching the intersection profile of a power-set distribution to that of sketching the winning distribution of a RUM. Without loss of generality we will assume that $i^* = n$ and for ease of notation we will write f_R for $f_R^{i^*}$.

Suppose that there is an α-additive sketch (ϕ, ψ) for $\{f_R\}_{R \in \mathcal{R}_n}$ with bit complexity κ. Specifically, for each RUM $R \in \mathcal{R}_n$, let $\Delta_R(\cdot) := \psi(\phi(R), \cdot)$, so that $|\Delta_R(T) - f_R(T)| \le \alpha$ for each $T \subseteq [n-1]$. Let D be any power-set distribution over $[n-1]$. Consider the RUM R defined as follows: sample $S \sim D$, then return a permutation π, where the top-most $|S|$ items of π are the items of S sorted in increasing order; the item in position $|S|+1$ of π is n; the last $n-1-|S|$ items of π are the items of $[n-1] \smallsetminus S$, also sorted in increasing order. Note that n loses against all items in S and wins against all items in $[n-1] \smallsetminus S$. For any $T \subseteq [n-1]$, we have that:

$$f_R(T) = R_{T \cup \{n\}}(n) = \Pr_{S \sim D}[S \cap T = \varnothing] = 1 - F_D(T).$$

Therefore, for each $T \subseteq [n-1]$:

$$|(1 - \Delta_R(T)) - F_D(T)| \le |(1 - f_R(T)) - F_D(T)| + |(1 - \Delta_R(T)) - (1 - f_R(T))| \le \alpha.$$

Therefore, $(\phi, 1 - \psi(\cdot, \cdot))$ is an α-additive sketch with bit complexity κ for $\{F_D\}_{D \in \mathcal{D}_{n-1}}$. Therefore, by Theorem 5, the sketch for $\{f_R\}_{R \in \mathcal{R}_n}$ must have bit complexity $\kappa = \Omega((n-1)^2) = \Omega(n^2)$. □

Observe that with a reduction analogous to the one used in the above proof, it is possible to show that any ε-additive sketch for $\{F_D\}_{D \in \mathcal{D}_{n-1}}$ with bit complexity κ can be transformed into an ε-additive sketch for $\{f_R^{i^*}\}_{R \in \mathcal{R}_n}$ for $i^* \in [n]$ with the same bit complexity κ. Hence, by Theorem 4, $\{f_R^{i^*}\}_{R \in \mathcal{R}_n}$ can be ε-additively sketched with $O(n^2/\varepsilon^2)$ bits.

8 Conclusions and Future Work

In this paper we developed a tool that yields tight lower bounds for the problem of compressing various objects of interest. We applied the tool to study the bit complexity of sketching (i) vertex boundaries in graphs, (ii) coverage functions, and (iii) Random Utility Models.

While we settle the complexity of sketching the intersection profiles of power-set distributions with respect to the universe size n, several questions remain open. A natural question is whether one can determine the optimal dependence on the accuracy parameter ε.

For the problem of sketching RUMs in ℓ_∞-distance, our results leave a gap of $\log n$. Is it possible to improve the lower bound or does there exist a more efficient sketch?

The work of Badanidiyuru et al. [2012] provides a more expensive sketch for coverage functions, but their sketch is *proper*, in that it is itself a coverage function. Do there exist proper sketches of size $\tilde{O}(n^2)$ bits?

Finally, we hope that the techniques developed in this work will prove useful in establishing lower bounds in settings beyond those considered here.

References

Zeyuan Allen-Zhu, Zhenyu Liao, and Lorenzo Orecchia. Spectral sparsification and regret minimization beyond matrix multiplicative updates. In *STOC*, pages 237–245, 2015.

Francis Bach. Learning with submodular functions: A convex optimization perspective. *Foundations and Trends® in Machine Learning*, 6(2-3):145–373, 2013.

Ashwinkumar Badanidiyuru, Shahar Dobzinski, Hu Fu, Robert Kleinberg, Noam Nisan, and Tim Roughgarden. Sketching valuation functions. In *SODA*, pages 1025–1035, 2012.

Maria-Florina Balcan and Nicholas J. A. Harvey. Submodular functions: Learnability, structure, and optimization. *SICOMP*, 47(3):703–754, 2018.

Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, D. Sivakumar, and Luca Trevisan. Counting distinct elements in a data stream. In *RANDOM*, pages 1–10, 2002.

MohammadHossein Bateni, Hossein Esfandiari, and Vahab Mirrokni. Almost optimal streaming algorithms for coverage problems. In *SPAA*, 2017.

Joshua Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-Ramanujan sparsifiers. *SICOMP*, 41(6):1704–1721, 2012.

András A Benczúr and David R Karger. Approximating st minimum cuts in $\tilde{O}(n^2)$ time. In *STOC*, pages 47–55, 1996.

Charles Carlson, Alexandra Kolla, Nikhil Srivastava, and Luca Trevisan. Optimal lower bounds for sketching graph cuts. In *SODA*, pages 2565–2569, 2019.

Deeparnab Chakrabarty and Zhiyi Huang. Recognizing coverage functions. *SICOMP*, 29(3):1585–1599, 2015.

Flavio Chierichetti, Ravi Kumar, and Andrew Tomkins. Light RUMs. In *ICML*, pages 1888–1897, 2021.

Graham Cormode, Mayur Datar, Piotr Indyk, and S. Muthukrishnan. Comparing data streams using hamming norms (how to zero in). In *VLDB*, 2002.

Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

Vitaly Feldman and Pravesh Kothari. Learning coverage functions and private release of marginals. In *COLT*, pages 679–702, 2014.

Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *Ann. Stat.*, pages 209–230, 1973.

David R Karger. Global min-cuts in RNC, and other ramifications of a simple min-cut algorithm. In *SODA*, pages 21–30, 1993.

Mikko Koivisto and Antti Röyskö. Fast multi-subset transform and weighted sums over acyclic digraphs. In *SWAT*, pages 29–1, 2020.

Yin Tat Lee and He Sun. Constructing linear-sized spectral sparsification in almost-linear time. *SICOMP*, 47(6):2315–2336, 2018.

Olivier Marchal and Julyan Arbel. On the sub-Gaussianity of the beta and Dirichlet distributions. *Electronic Communications in Probability*, 22(54):1–14, 2017.

Robert J. Serfling. Probability inequalities for the sum in sampling without replacement. *Ann. Stat.*, 2(1):39 – 48, 1974.

Daniel A Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. In *STOC*, pages 563–568, 2008.

Daniel A Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SICOMP*, 40(4):981–1025, 2011.

Kenneth E Train. *Discrete Choice Methods with Simulation.* Cambridge University Press, 2009.

Fan Yang, Kai He, Linxiao Yang, Hongxia Du, Jingbang Yang, Bo Yang, and Liang Sun. Learning interpretable decision rule sets: A submodular optimization approach. In *NeurIPS*, pages 27890–27902, 2021.

Grigory Yaroslavtsev and Samson Zhou. Approximate F_2-Sketching of Valuation Functions. In *APPROX/RANDOM*, 2019.

A Missing Proofs from Section 6.2

In this section, we show an algorithm that multiplicatively sketches $\mathcal{C}_{n,m}$ with $O((n/\varepsilon)^2 \log m)$ bits. We start by introducing some notation. For a set $S \subseteq \mathbb{N}$ (i.e., without duplicates) and integers $k \geq 1$, and $M \geq 1$, let:

$$\min{}^{(k)}(S) = \begin{cases} S & \text{if} \quad |S| < k, \\ k \text{ minimum distinct values of } S & \text{if} \quad |S| \geq k. \end{cases}$$

$$\hat{F}_0^{(k,M)}(S) = \begin{cases} |S| & \text{if} \quad |S| < k \\ \frac{k \cdot M}{\max(S)} & \text{if} \quad |S| \geq k \end{cases}$$

We will leverage on the well-known k-minimum value (KMV) sketch of Bar-Yossef et al. [2002], which provides the following guarantees:

Theorem 19 (Theorem 1 of Bar-Yossef et al. [2002], paraphrased)**.** *Let $\varepsilon, \delta \in (0,1)$, and let S be a multiset over the universe $[m]$, with $|S| = n$. Let $h : [m] \to [M]$, where $M = m^3$, be a uniformly random function. For $k = O(1/\varepsilon^2)$, let*

$$\hat{S} = \min{}^{(k)}\left(\{h(s) \mid s \in S\}\right).$$

Then, with probability $> 1/2$, it holds that:

$$(1-\varepsilon) \cdot F_0(S) \leq \hat{F}_0^{(k,M)}(\hat{S}) \leq (1+\varepsilon) \cdot F_0(S),$$

where $F_0(S)$ is the number of distinct elements of S. Storing $\hat{S}$ requires $O(\frac{\log m}{\varepsilon^2})$ bits.

Moreover, suppose to run the algorithm $t = O(\log(1/\delta))$ times independently, and let $\hat{S}_1, \ldots, \hat{S}_t$ be the corresponding sets produced. Let v be the median value of $\hat{F}_0^{(k,M)}(\hat{S}_1), \ldots, \hat{F}_0^{(k,M)}(\hat{S}_t)$. Then, with probability $\geq 1 - \delta$, it holds that:

$$(1-\varepsilon) \cdot F_0(S) \leq v \leq (1+\varepsilon) \cdot F_0(S).$$

Note that the original sketch of Bar-Yossef et al. [2002] was developed for streaming algorithms and uses a family of pairwise-independent hash functions rather than sampling h uniformly at random. However, we do not need to store h for our purposes, and we can therefore just pick a uniform at random function for simplicity.

A key property of this sketch is that it is easily composable. In particular, given the sketches for multisets $S_1, \ldots, S_n$ one can compute a sketch for their (multi)union $S = S_1 \cup \cdots \cup S_n$, that will be statistically equivalent to a sketch generated from scratch for S. We remark that Bateni et al. [2017] used exactly the same idea using the sketch of Cormode et al. [2002] in their Appendix D, but limited to sets of size k. For completeness, we provide here a full proof.

Most crucially for our application, storing all sketches for a collection $S_1, \ldots, S_n$ one can compute the sketch for the set:

$$S_A = \bigcup_{a \in A} S_a$$

for *all* subsets $A \subseteq [n]$. The guarantees of the basic sketch yield that *for each* choice of A, with probability at least $1-\delta'$ the sketch for S_A will have small multiplicative error. Setting $\delta' = \Theta(\delta 2^{-n})$ allows us to guarantee that, with probability at least $1-\delta$, *for all* choices of A the sketch will have smaller multiplicative error.

Theorem 16. *Fix any $\varepsilon \in (0,1)$. There exists an ε-multiplicative sketch for the set of all normalized covering functions $\mathcal{C}_{n,m}$ with bit complexity $O(\frac{n^2}{\varepsilon^2}\log m)$.*

Proof. Consider any normalized coverage function $f \in \mathcal{C}_{n,m}$. We show how to construct a data structure that approximates $f(A)$ within a multiplicative error of $(1 \pm \varepsilon)$ for any $A \subseteq [n]$.

Let $\delta = 2^{-(n+1)}$, $M = m^3$, and let $h_1, \ldots, h_t$ be functions from $[m]$ to $[M]$ sampled i.i.d. uniformly at random, where $t = \Theta(\log(1/\delta)) = \Theta(n)$ as required by Theorem 19. Let $k = \Theta(1/\varepsilon^2)$ as required by Theorem 19. If the coverage function f consists of a collection of sets $S_1, \ldots, S_n \subseteq U$, then, we store, for each $i \in [n]$, and $j \in [t]$ the set:

$$\hat{S}_i^{(j)} = \min{}^{(k)}(\{h_j(s) \mid s \in S_i\}).$$

Note that we do not store the functions $h_1, \ldots, h_t$. Therefore, the total bit complexity of our data structure is at most: $n \cdot t \cdot k \cdot \log_2(M) = O((n^2/\varepsilon^2) \cdot \log m)$.

Upon receiving a query $A \subseteq [n]$, we return the estimate:

$$\hat{f}(A) = \text{median}\left\{\hat{F}_0^{(k,M)}\left(\min{}^{(k)}\left(\bigcup_{a \in A} \hat{S}_a^{(j)}\right)\right)\right\}_{j \in [t]}.$$

Observe that, for each $j \in [t]$, we have:

$$\min{}^{(k)}\left(\bigcup_{a \in A} \hat{S}_{aj}\right) = \min{}^{(k)}\left(\bigcup_{a \in A}\{h_j(s) \mid s \in S_a\}\right),$$

therefore, $\hat{f}(A)$ has the same distribution as the estimate computed from scratch by the algorithm of Theorem 19 for the multiset $\cup_{a \in A} S_a$. Hence, with probability at least $1 - \delta$:

$$(1-\varepsilon) \cdot m \cdot f(A) \leq \hat{f}(A) \leq (1+\varepsilon) \cdot m \cdot f(A),$$

where we recall that $f(A) = \frac{1}{m}|\bigcup_{a \in A} S_a|$. In particular, by the union bound, with probability at least $1/2$ we have:

$$\forall A \subseteq [n]: \quad (1-\varepsilon) \cdot m \cdot f(A) \leq \hat{f}(A) \leq (1+\varepsilon) \cdot m \cdot f(A),$$

and the sketch will be correct as needed. □

B Typical Set of Intersection Profiles

In this section, we substantiate the claim that sketching the intersection profile of most power-set distributions requires little memory.

Consider:

$$\mathcal{T}_\varepsilon^{(n)} := \left\{D \in \mathcal{D}_n \,\middle|\, \text{for all } S \in 2^{[n]}, \frac{1}{2^{|S|}} - \varepsilon \leq 1 - F_D(S) \leq \frac{1}{2^{|S|}} + \varepsilon\right\}.$$

We say that a distribution $D \in \mathcal{D}_n$ is *ε-typical* if it belongs to $\mathcal{T}_\varepsilon^{(n)}$ (the *ε-typical set*). We show that as $n \to \infty$ the probability that a uniformly random distribution D is ε-typical tends to 1.

Theorem 20. *For any $\varepsilon > 0$:*

$$\lim_{n\to\infty} \Pr_{D\sim U(D_n)} \left[D \in \mathcal{T}_\varepsilon^{(n)}\right] = 1, \tag{1}$$

where $U(D_n)$ is the uniform distribution over the $(2^n - 1)$-dimensional simplex in $\mathbb{R}^{2^n}$.

In fact, we will prove the following stronger, non-asymptotic version of Theorem 20.

Theorem 21. *For any $\varepsilon > 0$, and any $n \geq 20 \log_2 \frac{1}{\varepsilon}$:*

$$\Pr_{D\sim U(D_n)} \left[D \in \mathcal{T}_\varepsilon^{(n)}\right] \geq 1 - \frac{2}{e^{2^{0.27n}}}.$$

Proof. Recall that the pdf of a Dirichlet distribution with parameters $(a_1, \dots, a_K)$ is given by

$$f(x_1, \dots, x_K) = \frac{\Gamma\left(\sum_{i=1}^K a_i\right)}{\prod_{i=1}^K \Gamma(a_i)} \prod_{i=1}^K x_i^{a_i - 1},$$

where $\Gamma(\cdot)$ is the gamma function and $\sum_{i=1}^K x_i = 1$ and $x_i \geq 0$ for each i.

We note that sampling a distribution D from $U(D_n)$ is equivalent to sampling a vector $X = (X_A)_{A\subseteq[n]}$ from a flat Dirichlet distribution with 2^n parameters $(1, \dots, 1)$, where each X_A indicates the probability of selecting set A from D.

Fix a subset $S \subseteq [n]$. We define $\tilde{X}_S := \sum_{A\subseteq[n]\,:\,A\cap S\neq\varnothing} X_A$ and $\tilde{Y}_S := \sum_{A\subseteq[n]\,:\,A\cap S=\varnothing} X_A$.

We recall the following standard result (see, e.g., Ferguson [1973]):

Proposition 22 (Aggregation Property of Dirichlet Random Variables)**.** *Let $X = (X_1, \dots, X_K) \sim$ Dirichlet$(a_1, \dots, a_K)$. For any $i < j \in [K]$, let $X^{(i\leftrightarrow j)}$ be the vector obtained from X by removing X_j and replacing X_i with $X_i + X_j$, we then have:*

$$X^{(i\leftrightarrow j)} \sim \text{Dirichlet}(a_1, \dots, a_{i-1}, a_i + a_j, a_{i+1}, \dots, a_{j-1}, a_{j+1}, \dots, a_K).$$

By the aggregation property of the Dirichlet distribution, we have:

$$\left(\tilde{X}_S, \tilde{Y}_S\right) \sim \text{Dirichlet}(\alpha, \beta),$$

where $\alpha = |\{A \in 2^{[n]} \mid A \cap S \neq \varnothing\}|$ and $\beta = |\{A \in 2^{[n]} \mid A \cap S = \varnothing\}|$. The marginal of a Dirichlet distribution follows a beta distribution (again, see, e.g., Ferguson [1973]):

$$\tilde{X}_S \sim \text{Beta}\,(\alpha, \beta)\,.$$

The latter is sub-Gaussian with a proxy-variance upper bounded by $\widetilde{\sigma}^2 = \frac{1}{4(\alpha+\beta+1)} \leq \frac{1}{2^{n+2}}$ [Marchal and Arbel, 2017, Theorem 2.1], meaning it satisfies:

$$\Pr[\tilde{X}_S - \mu_S \geq \varepsilon] \leq e^{-\frac{\varepsilon^2}{2\widetilde{\sigma}^2}} \leq e^{-\varepsilon^2 2^{n+1}},$$

where $\mu_S := 1 - \frac{1}{2^{|S|}}$ is the mean of $\tilde{X}_S$, since $\mathrm{E}\,[\tilde{X}_S] = \frac{\alpha}{\alpha+\beta}$. In particular, if $n \geq 20 \log_2 \frac{1}{\varepsilon}$, we have:

$$\Pr[\tilde{X}_S - \mu_S \geq \varepsilon] \leq \exp\left(-2^{0.9n+1}\right) \leq \exp\left(-n\ln(2) - 2^{0.27n}\right) = 2^{-n} e^{-2^{0.27n}}.$$

Applying a union bound, gives:

$$\Pr[\exists S \subseteq [n] : \tilde{X}_S - \mu_S \geq \varepsilon] \leq e^{-2^{0.27n}}.$$

By considering the variables $\{\tilde{Y}_S\}_{S\subseteq[n]}$ and applying the same argument as above, since we have $\tilde{Y}_S = 1 - \tilde{X}_S$, we obtain:

$$\Pr[\exists S \subseteq [n] : \mu_S - \tilde{X}_S \geq \varepsilon] \leq e^{-2^{0.27n}},$$

and hence:

$$\Pr[\exists S \subseteq [n] : |\tilde{X}_S - \mu_S| \geq \varepsilon] \leq 2e^{-2^{0.27n}}.$$

Noting that $\tilde{X}_S = F_D(S)$ completes the proof of Theorem 21 and hence that of Theorem 20. $\square$

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il