Submodular Clustering beyond $1 - 1/e$

Kiarash Banihashem[*]
University of Maryland
College Park, MD, USA
kiarash@umd.edu

MohammadHossein Bateni[*]
Google Research
New York City, NY, USA
bateni@google.com

Hossein Esfandiari[*]
Google Research
London, UK
esfandiari@google.com

Samira Goudarzi[*]
University of Maryland
College Park, MD, USA
samirag@umd.edu

MohammadTaghi Hajiaghayi[*]
University of Maryland
College Park, MD, USA
hajiagha@umd.edu

Abstract

Submodular clustering provides a flexible framework for modeling coverage and representation quality in metric spaces, capturing a broad range of objectives arising in machine learning, including distance-based coverage and data summarization. Given a set of points P in a metric space and a decreasing service function ϕ, the goal is to select k centers U to maximize total service $\sum_{p \in P} \phi(d(U, p))$, where $d(U, p)$ is the distance from p to its nearest center. While this objective is submodular and admits a standard $1 - 1/e$ approximation via greedy algorithms, it has remained unclear whether and when this barrier can be surpassed. We resolve this question by characterizing exactly which functions ϕ permit approximation guarantees beyond $1 - 1/e$. We identify a natural parameter γ_ϕ capturing how rapidly ϕ decreases, and show that, assuming P $\neq$ NP, a better-than $1 - 1/e$ approximation ratio is achievable in polynomial time if and only if $\gamma_\phi > 0$.

1 Introduction

Clustering has emerged as a central algorithmic primitive in modern machine learning systems, serving not only as an exploratory tool but as a structural component within learning, optimization, and inference pipelines. Recent work highlights the role of clustering in improving scalability and data efficiency, for example through clustering-based sensitivity sampling for training large models [ACH+24], error diagnosis and interpretability via clustering of influence embeddings [WAT+23], and systems-level acceleration such as clustered attention mechanisms and compression of large embedding tables [AAH+24, TA23]. Clustering also plays a key role in federated and distributed learning, where partitioning clients into coherent groups is essential for handling heterogeneity and improving convergence [KKdV24, LTK+22]. From a theoretical perspective, these applications have motivated extensive work on clustering with provable guarantees under constraints such as fairness, privacy, and dynamic updates [BIO+19, CEM+22, BCLP23, SLS25].

A complementary perspective views clustering through the lens of *coverage maximization*: rather than minimizing the cost of assigning points to centers, the goal is to select k centers maximizing the number or fraction of points effectively covered. This formulation arises naturally in influence maximization [KKT03], sensor placement [LKG+07, KSG08], and data summarization [LB11, TIWB14]. Both clustering and coverage maximization have been extensively studied from approximation and hardness perspectives [Fei98, CGTS99, JV01, CN06, DGKR03, JLN+05, DGRS05, STT07, OR10, LS13, CM14, ABC+16, CAEMN22]. Coverage maximization objectives are naturally submodular,

[*]Authors are ordered alphabetically.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

and submodular optimization has been studied thoroughly across a range of settings and computational models [NWF78, Fei98, FMV11, CCPV11, FNSZ23].

A natural class of problems sitting at the center of clustering, coverage, and submodularity is *submodular clustering*, which is parametrized by a decreasing *service function* $\phi : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ and is defined as follows. Given a set of points P in a metric space with distance function $d(.,.)$, select a set U of k centers to maximize total *service*

$$\sum_{p \in P} \phi(d(U,p)),$$

where $d(U,p) = \min_{u \in U} d(u,p)$ is the distance of p to its closest center in U. The class is fairly general and includes the coverage maximization problem, which is recovered by setting $\phi(x) = \mathbb{1}\{x \leq 1\}$ in the corresponding bipartite graph (see Section B).

Clustering and coverage both seek centers that are close to the input points. Our formulation expresses this goal by giving each point a service value $\phi(d(U,p))$ based on its distance to the closest center. For example, choosing $\phi(x) = \mathbb{1}\{x \leq r\}$ gives a point service 1 if a center is within distance r, and 0 otherwise, so maximizing total service is exactly maximizing coverage. A gradually decreasing ϕ instead rewards bringing centers closer to points, expressing the clustering goal through service maximization.

One can see that the objective is submodular and hence all problems in this class admit a $1 - 1/e$ approximation algorithm. The approximation ratio $1 - 1/e$ is a natural barrier arising frequently in submodular optimization and other areas of theoretical computer science. Many works have studied whether or not the $1 - 1/e$ barrier is tight for their problems; e.g., see [CC84, Fei98, FMMM09, SVW17, AEE+17, Von13, DF23]. This raises the following natural open question.

Open Question. *Which problems in the class of submodular clustering admit an approximation ratio better than* $1 - 1/e$*?*

Intuitively, $\phi(d(U,p))$ represents the service that p receives from the set of centers U, decreasing as p moves farther away, at a rate determined by the function ϕ. By selecting an appropriate ϕ, one can model various problems of independent interest. For instance, setting $\phi(x) = \mathbb{1}\{x \leq r\}$ for some radius r leads to the same-radius version of the covering-with-disks problem. Variants of this problem, which is closely related to geometric set cover, have been extensively studied for both general metrics and the Euclidean metric [MZH83, EvL08, MR09, Var10, BP12, MRR15, CDRS18]. Submodular clustering generalizes such problems continuously, allowing the notion of coverage—interpreted as service—to decay more naturally with distance. A natural choice for ϕ is exponential decay, modeled by $\phi(x) = e^{-x}$, which aligns with signal attenuation in communication networks—a common motivation for coverage problems. Alternatively, setting $\phi(x) = \frac{1}{x^p+1}$ for $p > 0$ gives a regularized inverse-power service objective.[2]

Beyond its theoretical foundation, submodular clustering has the advantage of being inherently robust to outliers. Intuitively, if a few points in the dataset are significantly farther from the rest, they can be safely ignored at the cost of a slight decrease in service quality. This contrasts with problems such as k-means and k-median, where failing to effectively reach a point can result in an excessively large cost, often dominating the objective and degrading the overall solution quality. As a result, these methods often rely on initial preprocessing to remove outliers using ad-hoc techniques, which typically require time and resource consuming tunings [HCK+05, CG13, FKRS19, GR22, KLS18].

1.1 Our contributions

In this paper, we study submodular clustering in the most general setting of arbitrary metrics and resolve the above open problem. Our characterization is in terms of a natural parameter $\gamma_\phi := \inf_{x \geq 0} \frac{\phi(2x)}{\phi(x)}$, where $0/0$ is defined to be 1 only in ratios defining decay parameters. Our computational conventions are stated in Section 3. This parameter measures the decay rate of ϕ. Intuitively, γ_ϕ measures how much service is retained at double the distance: $\gamma_\phi = 0$ means ϕ can

[2]Note that $\phi(x) = 1/x^p$ is not feasible since $\phi(0)$ would be undefined.

decay arbitrarily fast (e.g., exponential decay), while $\gamma_\phi > 0$ means ϕ retains a constant fraction of its value when the distance doubles.

Two degenerate cases are immediate: if $\phi(0) = 0$, every solution has value zero; if $\phi(x) = 0$ for every $x > 0$, it suffices to select as many centers in $P \cap \mathcal{U}$ as the budget permits. Otherwise, $\phi(y) > 0$ for some $y > 0$. If ϕ vanishes at a finite distance, the sequence $\phi(y), \phi(2y), \phi(4y), \ldots$ has a positive term followed by a zero term, so $\gamma_\phi = 0$. The convention $0/0 = 1$ therefore accommodates service functions with zeros without changing this conclusion.

Theorem 1. *For any fixed ϕ in the model of Section 3, there exists a polynomial time algorithm for submodular clustering with an approximation ratio exceeding $1 - 1/e$ by a positive constant if and only if $\gamma_\phi > 0$, unless* P = NP.

For example, for $\phi(x) = (x^p + 1)^{-1}$, with fixed $p > 0$, we have $\gamma_\phi = \frac{1}{2^p}$, so an approximation better than $1 - 1/e$ is achievable by a positive constant depending on p. The guaranteed additive improvement tends to zero as p increases. This is a regularized inverse-power service objective. For exponential decay $\phi(x) = e^{-x}$, we have $\gamma_\phi = 0$, so the $1 - 1/e$ barrier is tight.

Beyond the theoretical contribution, our results have a practical implication for the design of clustering objectives. The parameter γ_ϕ serves as a guideline for model selection: a practitioner choosing a service function (e.g., between exponential and polynomial decay) can use γ_ϕ to determine upfront whether their formulation admits approximation guarantees beyond $1 - 1/e$ or not. It should of course be noted that the worst-case guarantee may be difficult to observe experimentally, since greedy can already perform better than $1 - 1/e$ in practice.

Set $\eta_\phi := \inf_{x \geq 0} \phi(3x)/\phi(x)$. By Lemma 31, $\eta_\phi \geq \gamma_\phi^2$, so the following bound implies the additive $c\gamma_\phi^4$ improvement of Theorem 1.

Theorem 2. *There exists a constant $c > 0$ such that, for every fixed service function ϕ in the model of Section 3, the standard greedy algorithm runs in polynomial time and satisfies*

$$f(\mathtt{alg}_k) \geq (1 - 1/e + c\eta_\phi^2)\mathrm{OPT},$$

where $\mathtt{alg}_k$ is its output and OPT *is the optimal value.*

For hardness, set $\gamma_{\phi,\mathbb{Q}} := \inf_{x \in \mathbb{Q}_{>0}} \phi(2x)/\phi(x)$, again with $0/0 = 1$. Section B shows that $\gamma_\phi \leq \gamma_{\phi,\mathbb{Q}} \leq \sqrt{\gamma_\phi}$.

Theorem 3. *Fix a service function ϕ satisfying the computational conventions of Section 3, with $\gamma_{\phi,\mathbb{Q}} < 1$. For every constant $0 < \epsilon < (1 - \gamma_{\phi,\mathbb{Q}})/e$, it is NP-hard to approximate submodular clustering with ratio $1 - \frac{1}{e} + \frac{\gamma_{\phi,\mathbb{Q}}}{e} + \epsilon$, when k is part of the input, even when $P = \mathcal{U}$.*

This implies hardness beyond $1 - 1/e + \sqrt{\gamma_\phi}/e$. The proofs are in Sections A and B. Closing the quantitative gap between our approximation and hardness bounds is an interesting direction for future work.

We note that the problem is exactly solvable if $\gamma_\phi = 1$: then $\phi(x) = b$ for all $x > 0$, while $a := \phi(0)$ may exceed b. Indeed, $\phi(2x) = \phi(x)$ and monotonicity imply constancy on $(0, \infty)$. For nonempty U, the objective is $b|P| + (a - b)|U \cap P|$, so it suffices to choose as many centers in $P \cap \mathcal{U}$ as the budget permits. For all other ϕ, we further obtain the following as a consequence of the above bounds.

Theorem 4. *For any ϕ such that $\gamma_\phi \neq 1$, submodular clustering does not admit a PTAS unless* P = NP.

Setting $\phi(x) = \mathbb{1}\{x \leq r\}$ with $r > 0$, the above theorem shows that, even for the special case of covering with disks, there is no PTAS for submodular clustering unless P = NP. This stands in stark contrast to the Euclidean setting, where [CDRS18] provide a PTAS, further highlighting the challenges arising from the general metric setting studied in our paper.

Remark 5. *The choice of constant 2 in the definition of γ_ϕ does not affect Theorem 1. We show in Appendix C that $\inf_{x \geq 0} \frac{\phi(2x)}{\phi(x)} > 0$ if and only if $\inf_{x \geq 0} \frac{\phi(cx)}{\phi(x)} > 0$ for any constant $c > 1$.*

2 Related work

Clustering Clustering plays a pivotal role in computer science, serving as a fundamental tool for various applications such as data mining, machine learning, and optimization. In particular, clustering

algorithms like k-means, k-median, and k-center have garnered significant attention due to their wide-ranging applications and theoretical foundations [CAEMN22, CAGH$^+$22, HS85, CGTS99, JV01, Che06, AV07, MMK08, LS13, BPR$^+$17, CAGH$^+$22].

Additionally, many works have considered variants of the clustering problem such as clustering with fairness [BPR$^+$17, CKLV17, MV20, LLS$^+$21], differentially private clustering [SCL$^+$16, GKM20, CKM$^+$21], Hierarchical clustering [HG05, MC12, BBD$^+$17], and Correlation Clustering [BBC04, DEFI06, BCMT22, CALN22]. For Submodular clustering, while variants of the problem are commonly considered in the ML literature, most works are more experimental in nature [NJB05, NKI10, DG14, WIW$^+$15, DS19]. Backurs and Har-Peled [BHP20] study the same service objective in $\mathbb{R}^d$ and obtain a $(1-\varepsilon)$-approximation in time $n^{\varepsilon^{-O(d)}}$ for arbitrary non-increasing ϕ. Our contribution is a characterization of when $1 - 1/e$ can be beaten in general metrics.

Since real-world data often contain outlier data points, and the objective in k-median, k-means, and k-center is susceptible to outliers, many works have considered approaches to deal with the issue, both from a practical and theoretical perspective [FKRS19, IQM$^+$20, CAFS21, SKR$^+$23]. These works are typically based on augmentations of the original clustering objective to make it less susceptible to noise. In contrast, the submodular clustering approach we study is naturally robust to outliers given its objective function.

Submodular optimization Submodular optimization is a classical subfield in theoretical computer science with applications in various fields such as machine learning, economics, operations research, and data summarization. The classical problem in submodular optimization literature is submodular maximization under a cardinality constraint: given a set of elements V and a monotone submodular function $f : 2^V \to \mathbb{R}^{\geq 0}$, the goal is to find a set $S \subseteq V$ of size $|S| \leq k$ that maximizes f. For this problem, it has long been known that the standard greedy algorithm obtains a $1 - 1/e$ approximation factor [NWF78] and many works have considered variants of the problem in various computational models [FMV11, CCPV11, CGQ15, KMZ$^+$19]. Beginning with [Fei98], several works have studied the hardness of approximation beyond $1 - 1/e$ under increasingly relaxed complexity assumptions [FLT04, FV10, KPV13], and in the presence of structure in the problem.

Both our positive and negative results are directly related to these works. The positive result identifies an important practical setting where beating $1 - 1/e$ is possible. As such, the techniques in the paper may be applicable to other problems with a similar submodular formulation captured with a metric space. In addition, the negative result extends this line of work, in particular by identifying a new important setting where $1 - 1/e$ is tight.

Service maximization Theoretical problems with a service maximization nature such as geometric coverage and facility location have long been studied in the theoretical computer science community [GK98, JMM$^+$03, JLW$^+$18, CDRS18]. In particular, the Euclidean version of the submodular clustering problem and the specific case where $\phi(x) = \mathbb{1}\{x \leq 1\}$ corresponds to the problem of geometric covering with disks, for which [CDRS18] provide a PTAS. For the general metric setting however, our hardness result implies that the problem does not admit a PTAS in this case. This stands in contrast with the result for the specific Euclidean setting and hence highlights the difficulty rising from the general metric setting studied in our paper.

3 Preliminaries

Problem setting We assume that the centers and points lie in an arbitrary metric space (M, d) where d denotes the distance function. We use $P \subseteq M$ to denote the set of points we are trying to cluster and $\mathcal{U} \subseteq M$ to denote all choices for the cluster centers. Our algorithmic results do not make any assumptions on the relationship between $\mathcal{U}$ and P while our hardness results hold even when $P = \mathcal{U}$.

We use $\phi : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ to denote the service function, which we assume to be monotone and non-increasing.

The sets P and $\mathcal{U}$ are finite and explicitly given, and ϕ is fixed independently of the input size. Polynomial-time statements assume an input representation in which the required distance, service, and marginal-gain computations and comparisons are polynomial-time. Our hardness constructions use rational distances, encoded in binary. The hardness statements concern deterministic algorithms.

We assume $1 \leq k \leq |\mathcal{U}|$; the cases $k = 0$ or $k > |\mathcal{U}|$ are immediate. By monotonicity, we may take the optimal solution and the greedy output to have exactly k centers.

For any two points $p, q \in M$, we define $f(p, q) := \phi(d(p, q))$ to be the amount p *serves* q. Given any two sets X and Y, we define

$$f(X, Y) = \sum_{y \in Y} \phi(d(X, y)),$$

where $d(X, y) := \min_{x \in X} d(x, y)$ for $X \neq \varnothing$. We define $f(\varnothing, Y) = 0$. When either of the two sets X and Y are singletons $\{ x \}$ or $\{ y \}$, we will abuse notation by writing $f(x, y)$ instead of $f(\{ x \}, \{ y \})$. For any fixed Y, the function $f(X, Y)$ is submodular in X because each term $\phi(d(X, y))$ is submodular in X (see below for a formal definition), which in turn holds because ϕ is non-increasing. The submodular clustering problem is to pick a set $U \subseteq \mathcal{U}$ of size at most k that maximizes $f(U, P)$.

Submodular functions A set function $f : 2^{\mathcal{U}} \to \mathbb{R}$ is *monotone* if $f(A) \leq f(B)$ for all $A \subseteq B \subseteq \mathcal{U}$. It is *submodular* if it satisfies the diminishing returns property: for all $A \subseteq B \subseteq \mathcal{U}$ and all $u \in \mathcal{U} \setminus B$,

$$f(A \cup \{u\}) - f(A) \;\geq\; f(B \cup \{u\}) - f(B).$$

Equivalently, for all $A, B \subseteq \mathcal{U}$, $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$. In (monotone) submodular maximization under the cardinality constraint, we are given a nonnegative monotone submodular function f and a parameter k, and the goal is to output a set S of size at most k that maximizes $f(.)$. It is well-known that the standard greedy algorithm, which repeatedly adds the element with maximal *marginal gain* with respect to the current set, obtains a $1 - 1/e$ approximation ratio. Moreover, this ratio cannot be improved for general submodular functions unless P = NP.

Notation For any three sets of points X, Y, Z, we define

$$f(X, Y \mid Z) := f(X \cup Z, Y) - f(Z, Y)$$

to be the marginal gain of X with respect to Z for the function $f(., Y)$. We will often omit the second argument of $f(., .)$ when it is set to P, i.e., we define $f(X) = f(X, P)$ and $f(X \mid Y) = f(X, P \mid Y)$.

We use $\texttt{alg} \subseteq U$ to denote the output of the greedy algorithm. We use u_t to denote the element chosen by greedy at time t and use $\texttt{alg}_t = \{ u_1, \ldots, u_t \}$ to denote the first t elements added to $\texttt{alg}$. By definition of greedy, u_t maximizes the marginal gain among all remaining centers, i.e., $u_t \in \arg\max_{u \in U \setminus \texttt{alg}_{t-1}} f(u \mid \texttt{alg}_{t-1})$. We use $\Delta_t = f(u_t \mid \texttt{alg}_{t-1})$ to denote this marginal gain. We write $f_t(X, Y) := f(X, Y \mid \texttt{alg}_{t-1})$ and $f_t(X) := f(X \mid \texttt{alg}_{t-1})$ and define gap_t as $f(\texttt{opt}) - f(\texttt{alg}_t)$.

We use $\texttt{opt} \subseteq U$ to denote the optimal solution to the clustering problem and use $\text{OPT} = f(\texttt{opt})$ to denote its value. For any center $u^* \in \texttt{opt}$, we use $\texttt{zone}(u^*) \subseteq P$ to denote all points for which u^* is the closest center in $\texttt{opt}$, i.e.,

$$\texttt{zone}(u^*) = \left\{ p \in P : u^* = \arg\min_{v^* \in \texttt{opt}} d(p, v^*) \right\},$$

where we assume that $\arg\min$ breaks ties in a fixed but otherwise arbitrary manner (e.g., lexicographic). For any point $p \in M$, we use $u^*_{\min}(p) = \arg\min_{u^* \in \texttt{opt}} d(p, u^*)$ to denote the closest point in $\texttt{opt}$ to p. We use $\texttt{zone}^{\max}_t$ to denote $\texttt{zone}(u^*_{\min}(u_t))$.

4 Overview of proof

We here sketch the main ideas used in our proofs. Due to space constraints, the full proofs are deferred to the Appendix.

Our analysis provides a fundamental understanding of when and how the standard greedy algorithm can exceed the classical $1 - 1/e$ barrier. We introduce a structured decomposition of the sources of tightness in the usual analysis and show that, in the submodular clustering setting, each source can be overcome by leveraging the geometry of the metric and the decay behavior of the service function ϕ. In particular, we isolate three distinct sources of potential slack: independence from OPT,

concentration of service within a single zone, and slack recoverable through the triangle inequality. We develop tailored arguments for each case. This perspective yields improved approximation guarantees for a broad class of decay functions and may also be applicable to other problems involving metric-based submodular objectives.

We start with an overview of our positive result. For the intuition in this overview, take $\phi(x) = 1/(x+1)$ and write $\eta = \inf_{x \geq 0} \phi(3x)/\phi(x) = 1/3$. Integer roundings are suppressed in this overview. The full proofs handle general ϕ. We first briefly recap the main intuition behind the standard analysis of the greedy algorithm. Throughout our proofs, we refer to the individual steps in this analysis and show how we refine them for our specific problem. We then discuss the main ideas behind our improved approximation ratio and present an outline of the proof. Finally, we discuss the main ideas behind the hardness of approximation (upper bound) for the problem.

Recap: standard analysis Let $\texttt{alg} = \{ u_1, \ldots, u_k \}$ denote the output of the greedy algorithm, where u_t is the element added at time step t, and let $\texttt{alg}_t$ denote $\{ u_1, \ldots, u_t \}$. Let gap_t denote $f(\texttt{opt}) - f(\texttt{alg}_t)$. We divide the proof into 4 steps, where we successively lower bound the marginal gain of u_t with respect to $\texttt{alg}_{t-1}$, i.e., $f(u_t \mid \texttt{alg}_{t-1})$.

1. *Greedy* step: $f(u_t \mid \texttt{alg}_{t-1})$ is higher than the marginal of any element in $\texttt{opt}$ (with respect to $\texttt{alg}_{t-1}$). This is because, by definition, the greedy algorithm picks the element with the highest marginal gain. Formally,
$$f(u_t \mid \texttt{alg}_{t-1}) \geq \max_{u^* \in \texttt{opt}} f(u^* \mid \texttt{alg}_{t-1}).$$
2. *Average* step: The maximum marginal gain in $\texttt{opt}$ is greater than or equal to the average marginal gain. Formally,
$$\max_{u^* \in \texttt{opt}} f(u^* \mid \texttt{alg}_{t-1}) \geq \frac{1}{k} \sum_{u^* \in \texttt{opt}} f(u^* \mid \texttt{alg}_{t-1})$$
because the maximum of any set is higher than its average.
3. *Spread* step: By submodularity of f, the sum of marginal gains across the elements of $\texttt{opt}$ is at least the marginal gain of $\texttt{opt}$. Formally
$$\sum_{u^* \in \texttt{opt}} f(u^* \mid \texttt{alg}_{t-1}) \geq f(\texttt{opt} \mid \texttt{alg}_{t-1}).$$
4. *Gap* step: The marginal gain of $\texttt{opt}$ with respect to $\texttt{alg}_{t-1}$ is at least the difference between $f(\texttt{alg}_{t-1})$ and $f(\texttt{opt})$ because f is monotone. Formally,
$$f(\texttt{opt} \mid \texttt{alg}_{t-1}) \geq \text{gap}_{t-1}.$$
because $f(\texttt{opt} \cup \texttt{alg}_{t-1}) \geq f(\texttt{opt})$.

The above four steps together imply that at each iteration t, the marginal gain of the greedy algorithm satisfies
$$f(u_t \mid \texttt{alg}_{t-1}) \geq \frac{1}{k} \text{gap}_{t-1},$$
which yields the classical $1 - 1/e$ approximation ratio after k iterations. This bound is tight for general monotone submodular functions, as each inequality in the analysis can be simultaneously tight.

To see this, consider the unit-weight square $[0, k] \times [0, k]$ with value equal to covered area (a monotone submodular function). We are given k *column* sets $C_i = [i-1, i] \times [0, k]$, each covering the full height of the square over an x-interval of width 1. We are also given k *row* sets $R_i = [0, k] \times \left[i - 1,\ i - 1 + (1 - \frac{1}{k})^{i-1}\right]$, i.e., row i covers the full width but only a decreasing fraction of height. Clearly, selecting all k columns covers the entire square and achieves value k^2. However, if greedy begins by selecting rows in order, then after selecting $R_1, \ldots, R_{t-1}$, the remaining uncovered area inside any column is exactly $k \cdot (1 - \frac{1}{k})^{t-1}$, so the marginal gain of any column C_i is $k(1 - \frac{1}{k})^{t-1}$. On the other hand, the marginal gain of the next row is also $\text{area}(R_t) = k(1 - \frac{1}{k})^{t-1}$, so at each iteration there is a tie and greedy may continue selecting rows. After k steps, the total area covered by the rows is $\sum_{i=1}^{k} k\left(1 - \frac{1}{k}\right)^{i-1} \approx (1 - 1/e)\, k^2$, showing that without additional structure, greedy cannot be guaranteed to beat $1 - 1/e$.

Identifying leverage beyond $1 - 1/e$. To surpass this barrier, one must exploit additional structure beyond submodularity. In the submodular clustering setting, this structure arises from the underlying metric space and the decay behavior of the service function ϕ. Our analysis shows that these properties fundamentally alter the tightness of the classical inequalities. In particular, while the steps in the standard analysis may remain tight, they cannot all remain tight together. Indeed, the geometric constraints imposed by the triangle inequality and the decay behavior of the service function ϕ introduce structural interactions between the different steps. Our analysis makes this interaction explicit by isolating the precise steps at which slack can be recovered and by showing that at least one such recovery must occur for a large number of greedy iterations whenever $\gamma_\phi > 0$.

Improving the lower bound on the marginal gain $f(u_t \mid \texttt{alg}_{t-1})$ from $\frac{1}{k}\text{gap}_{t-1}$ to $\frac{1+\Omega(1)}{k}\text{gap}_{t-1}$ at each iteration improves the bound on $\text{gap}_k/\text{gap}_0$ to

$$(1 - \frac{1+\Omega(1)}{k})^k \leq e^{-k\frac{1+\Omega(1)}{k}} \leq 1/e - \Omega(1),$$

which means the approximation ratio is at least $1 - 1/e + \Omega(1)$. In fact, because we always have the bound $f(u_t \mid \texttt{alg}_{t-1}) \geq \frac{1}{k}\text{gap}_{t-1}$, even if we show that $f(u_t \mid \texttt{alg}_{t-1}) \geq \frac{1+\Omega(1)}{k}\text{gap}_{t-1}$ holds for at least $\Omega(k)$ distinct values of t, we can still prove an approximation ratio of $1 - 1/e + \Omega(1)$. In fact, we do not actually need to improve the marginal gain guarantee at every iteration of the greedy algorithm, and it suffices to show that for a constant fraction of iterations $t \in [k]$, the standard bound $f(u_t \mid \texttt{alg}_{t-1}) \geq \frac{1}{k}\,\text{gap}_{t-1}$ can be strengthened by a multiplicative constant factor depending only on γ_ϕ. Since the gap decreases multiplicatively over time, such local improvements compound across iterations, yielding a strictly better final approximation ratio.

We start by focusing on the Spread step of the analysis and use the triangle inequality to prove that under some optimistic assumptions, the bounds in this step can be improved. These assumptions should be viewed as a conceptual device rather than a restriction of generality. The full analysis proceeds by showing that whenever any of these assumptions fails for a significant fraction of greedy iterations, the algorithm already achieves a strictly better-than-$1 - 1/e$ approximation by alternative arguments. Consequently, the optimistic scenario represents the only remaining case to analyze, and it is precisely in this regime that the triangle inequality and decay properties of ϕ yield additional slack.

Partition the set of points P into k *zones* based on the closest center in $\texttt{opt}$ to the point; formally,

$$\texttt{zone}(u^*) = \left\{\, p \in P : u^* = \arg\min_{v^* \in \texttt{opt}} d(v^*, p) \,\right\},$$

where we assume that ties in $\arg\min$ are broken in some fixed but otherwise arbitrary manner (e.g. lexicographically). By definition of zones, we have

$$f(\texttt{opt} \mid \texttt{alg}_{t-1}) = \sum_{u^* \in \texttt{opt}} f(u^*, \texttt{zone}(u^*) \mid \texttt{alg}_{t-1}).$$

Therefore, in order for the Spread step to be tight, we need to have $f(u^* \mid \texttt{alg}_{t-1}) = f(u^*, \texttt{zone}(u^*) \mid \texttt{alg}_{t-1})$. In other words, each center in $\texttt{opt}$ should only serve the points in its own zone. More precisely, the inequality in the Spread step ignores the amount

$$\frac{1}{k} \sum_{u^* \neq v^*} f(u^*, \texttt{zone}(v^*) \mid \texttt{alg}_{t-1})$$

discarded by the Spread step when lower bounding Δ_t. We denote this ignored amount by $\texttt{excess}_t$ and will show that it can actually be quite large because of the triangle inequality in the metric space.

Focus on the time step $t = 0.01k$. Let u^* denote the closest center in $\texttt{opt}$ to u_t. The marginal gain of the center u_t chosen by the algorithm can be decomposed into

$$f(u_t \mid \texttt{alg}_{t-1}) = \sum_{v^* \in \texttt{opt}} f(u_t, \texttt{zone}(v^*) \mid \texttt{alg}_{t-1}).$$

Consider any center $v^* \neq u^* \in \texttt{opt}$ and some point $p \in \texttt{zone}(v^*)$. For now, assume that p is closer to its optimal center v^* than to u_t.

Optimistic assumption 1: $d(v^*, p) < d(u_t, p)$.

This is an optimistic condition on $p \in \texttt{zone}(v^*)$; optimality of $\texttt{opt}$ alone does not imply this pointwise distance comparison. Applying the triangle inequality on (u_t, p, v^*), we conclude that

$$d(u_t, v^*) \le d(u_t, p) + d(p, v^*) \le 2d(u_t, p),$$

which in turn implies that $d(u_t, u^*) \le 2d(u_t, p)$ since u^* was the closest center in $\texttt{opt}$ to u_t. Applying the triangle inequality again on (u^*, u_t, p) we obtain $d(u^*, p) \le 3d(u_t, p)$ which means $f(u^*, p) \ge \frac{1}{3} f(u_t, p)$ because $\phi(x) = 1/(x+1)$. This is illustrated in Figure 1. To turn this service bound into a marginal-gain bound, we also assume that p has received little service before step t.

Optimistic assumption 2: $f(\texttt{alg}_{t-1}, p) < \frac{1}{6} f(u_t, p)$.

It follows that

$$\begin{aligned} f(u^*, p \mid \texttt{alg}_{t-1}) &= f(u^*, p) - f(\texttt{alg}_{t-1}, p) \\ &\ge \frac{1}{3} f(u_t, p) - \frac{1}{6} f(u_t, p) \\ &= \frac{1}{6} f(u_t, p) \ge \frac{1}{6} f(u_t, p \mid \texttt{alg}_{t-1}). \end{aligned}$$

If this holds for all v^* and all $p \in \texttt{zone}(v^*)$, we conclude that

$$\texttt{excess}_t \ge \sum_{v^* \ne u^*} \frac{1}{6k} f(u_t, \texttt{zone}(v^*) \mid \texttt{alg}_{t-1}).$$

Finally, assume that at least half the marginal gain of u_t comes from points outside $\texttt{zone}(u^*)$.

Optimistic assumption 3: $f(u_t, \texttt{zone}(u^*) \mid \texttt{alg}_{t-1}) \le \frac{1}{2} f(u_t \mid \texttt{alg}_{t-1})$.

Under this assumption, we obtain

$$\texttt{excess}_t \ge \Omega(\frac{1}{k}) f(u_t \mid \texttt{alg}_{t-1}).$$

Other than the factor $1/k$ which we will need to remove, this is exactly what we wanted.

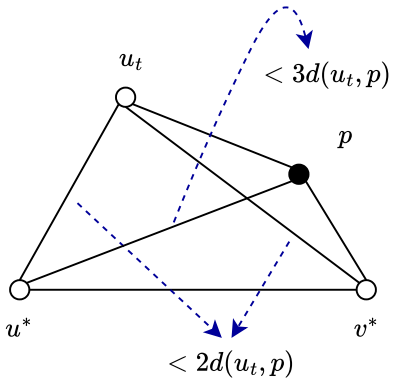


Figure 1: Geometric relationship between a greedy center u_t, its closest optimal center u^*, and a point p assigned to a different optimal center $v^* \ne u^*$. Assuming $d(v^*, p) \le d(u_t, p)$, we use the triangle inequality to show $d(u^*, p) \le 3d(u_t, p)$, which yields a constant-factor comparison between the services $f(u^*, p)$ and $f(u_t, p)$ for polynomially decaying service functions.

To remove the factor of $1/k$, we will consider the contribution of p to all $\texttt{excess}_{t'}$ rather than just $\texttt{excess}_t$. Specifically, we will show that p contributes to at least $\Omega(k)$ distinct values of $\texttt{excess}_{t'}$. Assume that $f(\texttt{alg}_{\epsilon k}, p) \le \frac{1}{6} f(u_t, p)$ for some small, but constant, value of ϵ. This means that $f(u^*, p \mid \texttt{alg}_{t'}) \ge \frac{1}{6} f(u_t, p \mid \texttt{alg}_{t-1})$ for all $t' \in [\epsilon k]$. By counting the contribution of p for all $t' \in [\epsilon k]$, we can improve the previous bounds by a factor of ϵk. To avoid double counting, for each p choose its closest greedy center among the relevant centers contributing to p. The sum of their marginal gains at p is at most the service of this one center. Section A.3.3 formalizes this argument.

Proof outline. The old, independent, and concentrated cases handle quantitative versions of assumptions 2, 1, and 3, respectively; the remaining triangle case uses the triangle argument. Set $\epsilon = c\eta$ where c is a small constant (see Section A). Let $\texttt{alg}'$ denote the set of centers in the interval $[0.005k, 0.01k)$, with $\texttt{alg}' = \{u_t : 0.005k \leq t < 0.01k\}$. We focus on this set because the relatively small values of t ensure that the marginal gains in these time steps are large. Formally, we can ensure that either for all $u_t \in \texttt{alg}'$ we have $f(u_t \mid \texttt{alg}_{t-1}) \geq 0.98\Delta^{\text{base}}$ where $\Delta^{\text{base}} = f(\texttt{opt})/k$, or the approximation ratio is at least $1 - 1/e + \Omega(1)$ (see Corollary 8).

Large number of old centers. We start with the early-time condition $f(\texttt{alg}_{\epsilon k}, p) \leq \frac{1}{6} f(u_t, p)$ corresponding to optimistic assumption 2. We categorize the points in P into two distinct classes: we say p is *young for* u_t if the assumption holds and *old for* u_t if it does not hold. We use P_t^{yo} and P_t^{old} to denote the set of all young and old points for u_t respectively. We say that u_t itself is *young* if "most" points p are young for it, where "most" is measured in terms of marginal gain; formally, u_t is young if and only if $f(u_t, P_t^{\text{yo}} \mid \texttt{alg}_{t-1}) \geq 0.9\Delta^{\text{base}}$. We say u_t is old if it is not young. We use $\texttt{alg}^{\text{old}}$ to denote the set of old centers in $\texttt{alg}'$.

In Section A.2, we show that if at least 0.1 fraction of the centers in $\texttt{alg}'$ are old, then the algorithm obtains an approximation ratio $1 - 1/e + \Omega(1)$. The main intuition behind the proof is as follows. Assume that we sum up $f(u_t, P_t^{\text{old}} \mid \texttt{alg}_{t-1})$ for all $u_t \in \texttt{alg}^{\text{old}}$. By definition of old centers and the assumption that $\left|\, \texttt{alg}^{\text{old}} \,\right| \geq 0.1 \left|\, \texttt{alg}' \,\right|$, the resulting sum should be at least $\Omega(f(\texttt{opt}))$. We show that $f(\texttt{alg}_{\epsilon k})$ is a constant fraction of this sum, where the constant is independent of ϵ. The proof is based on the fact u_t in the sum and each $p \in P_t^{\text{old}}$, $f(\texttt{alg}_{\epsilon k}, p)$ is at least $1/6 f(u_t, p)$. Combining the results we conclude that $f(\texttt{alg}_{\epsilon k})$ is $\Omega(f(\texttt{opt}))$. This improves on the standard analysis from submodular functions which only ensures $f(\texttt{alg}_{\epsilon k}) \geq \Omega(\epsilon f(\texttt{opt}))$. Since we always have a $(1 - 1/k)$ factor reduction in gap_t, the improved bound at $t = \epsilon k$ leads to an improved bound for $f(\texttt{alg}_k)$ as well (see Lemma 6).

Large number of independent centers. Given the above result, we can focus on the case where $\left|\, \texttt{alg}^{\text{yo}} \,\right| \geq 0.9 \left|\, \texttt{alg}' \,\right|$. We further split $\texttt{alg}^{\text{yo}}$ into three separate sets in order to handle the other optimistic assumptions we had made. We let $\texttt{alg}^{\text{ind}}$ denote the centers for which $f(u_t \mid \texttt{opt} \cup \texttt{alg}_{t-1}) \geq 0.01\Delta^{\text{base}}$. Intuitively, these are centers that are in some ways independent from $\texttt{opt}$ because they have a high marginal gain with respect to $\texttt{opt}$.

In Section A.3.1, we show that if $\left|\, \texttt{alg}^{\text{ind}} \,\right| \geq 0.3 \left|\, \texttt{alg}' \,\right|$, then the approximation ratio is $1 - 1/e + \Omega(1)$. The main idea behind the proof is that since these centers serve points in excess of $\texttt{opt}$, if $\texttt{alg}^{\text{ind}}$ is large then for $t \geq \lceil\, 0.01k \,\rceil$, the Gap step in the standard submodular analysis is no longer tight.

Excluding independent centers gives the service comparison corresponding to assumption 1 (Lemma 22).

Large number of concentrated centers. We further let $\texttt{alg}^{\text{con}}$ denote centers for which $f(u_t, . \mid \texttt{alg}_{t-1})$ is highly concentrated in a single zone; specifically, $f(u_t, \texttt{zone}_t^{\max} \mid \texttt{alg}_{t-1}) \geq 0.1\Delta^{\text{base}}$ where $\texttt{zone}_t^{\max}$ is the zone corresponding to the closest center to u_t in $\texttt{opt}$. We show in Section A.3.2 that if $\left|\, \texttt{alg}^{\text{con}} \,\right| \geq 0.3 \left|\, \texttt{alg}' \,\right|$, then the approximation ratio is $1 - 1/e + \Omega(1)$. Intuitively, this is because these centers disrupt the balance of $f(\texttt{opt}, \texttt{zone}(u^*) \mid \texttt{alg}_{t-1})$ across different u^*. If the number of these centers is large, then $f(\texttt{opt}, \texttt{zone}(u^*) \mid \texttt{alg}_{t-1})$ becomes much larger for some u^* than others, which makes the Average step in the standard analysis not tight.

Large number of triangle centers. Finally, the remaining centers $\texttt{alg}^{\text{tri}}$ satisfy the quantitative conditions needed for the triangle argument.

Hardness of approximation For the hardness result, we use a reduction from the Max k-cover problem. Specifically, we use the fact that it is NP-hard to choose k sets covering a maximum number of elements, even when there is a choice of k sets covering all of the elements, and when the number of elements is much higher than the number of sets. We here briefly sketch the proof that the problem does not admit a PTAS when $\phi = \frac{1}{x+1}$ (suppressing arbitrarily small additive errors in the bounds below) and refer to Section B for the main result.

Consider the bipartite graph $G = (L, R)$ corresponding to the Max k-cover problem with the extra assumptions made above, where L and R correspond to sets and elements respectively. Give distinct adjacent vertices distance 1 and distinct nonadjacent vertices distance 2; this is a finite metric. Assume for contradiction that a PTAS exists; choose a promised Max k-cover instance on which its extracted left-hand centers cover at most a $(1 - 1/e + o(1))$ fraction, as justified in Section B. Let $\texttt{alg}$ denote the output of the algorithm. We will lower bound the service from the optimal solution $\texttt{opt}$ and upper bound the service from $\texttt{alg}$ to bound the approximation ratio.

We start with the upper bound on $f(\texttt{alg})$. Since $|\,R\,| \gg |\,L\,|$, the service from $\texttt{alg}$ to L is negligible compared to its service to R, which means that $f(\texttt{alg}) \approx f(\texttt{alg}, R)$. Decompose $\texttt{alg}$ into $\texttt{alg}_L = \texttt{alg} \cap L$ and $\texttt{alg}_R = \texttt{alg} \cap R$. By assumption, we know that $\texttt{alg}_L$ has at most $(1 - 1/e + o(1))\,|\,R\,|$ neighbors in R. For any such neighbor p, we have $f(\texttt{alg}_L, p) = \phi(1)$. For any $p \in R$ that is not a neighbor of $\texttt{alg}_L$, we have $f(\texttt{alg}_L, p) \leq \phi(2)$. This allows us to upper bound $f(\texttt{alg}_L, R)$ with $|R|\,(\,\phi(1)(1 - 1/e) + \phi(2)/e\,)$. The addition of $\texttt{alg}_R$ does not affect this by much since we have $f(\texttt{alg}_R, p) = \phi(0)$ for only $k \leq |\,L\,| \ll |\,R\,|$ vertices and the baseline term $\phi(2)$ is already included for every point in R. Therefore, $f(\texttt{alg})$ is, ignoring lower order terms, upper bounded by

$$|R|\,(\,\phi(1)(1 - 1/e) + \phi(2)/e\,)\,.$$

As for the lower bound on $f(\texttt{opt})$, by assumption there exists k sets that cover all of the elements. Taking the points corresponding to these sets as centers, we conclude that

$$f(\texttt{opt}) \geq \phi(1)\,|\,R\,|\,.$$

Combining this with the upper bound on $f(\texttt{alg})$ and the fact that $\phi(2) < \phi(1)$, the claim follows.

Acknowledgments and Disclosure of Funding

The work is partially supported by DARPA expMath, ONR MURI 2024 award on Algorithms, Learning, and Game Theory, Army-Research Laboratory (ARL) grant W911NF2410052, NSF AF:Small grants 2218678, 2114269, 2347322. The authors declare no competing interests.

References

[AAH+24] Saurabh Agarwal, Bilge Acun, Basil Hosmer, Mostafa Elhoushi, Yejin Lee, Shivaram Venkataraman, Dimitris Papailiopoulos, and Carole-Jean Wu. CHAI: clustered head attention for efficient LLM inference. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024.

[ABC+16] Yossi Azar, Niv Buchbinder, T.-H. Hubert Chan, Shahar Chen, Ilan Reuven Cohen, Anupam Gupta, Zhiyi Huang, Ning Kang, Viswanath Nagarajan, Joseph Naor, and Debmalya Panigrahi. Online algorithms for covering and packing problems with convex objectives. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 148–157. IEEE Computer Society, 2016.

[ACH+24] Kyriakos Axiotis, Vincent Cohen-Addad, Monika Henzinger, Sammy Jerome, Vahab Mirrokni, David Saulpic, David P. Woodruff, and Michael Wunder. Data-efficient learning via clustering-based sensitivity sampling: Foundation models and beyond. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024.

[AEE+17] Melika Abolhassani, Soheil Ehsani, Hossein Esfandiari, MohammadTaghi Hajiaghayi, Robert D. Kleinberg, and Brendan Lucier. Beating 1-1/e for ordered prophets. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 61–71. ACM, 2017.

[AV07] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[BBC04] Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. *Machine learning*, 56:89–113, 2004.

[BBD+17] MohammadHossein Bateni, Soheil Behnezhad, Mahsa Derakhshan, MohammadTaghi Hajiaghayi, Raimondas Kiveris, Silvio Lattanzi, and Vahab Mirrokni. Affinity clustering: Hierarchical clustering at scale. *Advances in Neural Information Processing Systems*, 30, 2017.

[BCLP23] Sayan Bhattacharya, Martín Costa, Silvio Lattanzi, and Nikos Parotsidis. Fully dynamic k-clustering in $\tilde o(k)$ update time. *CoRR*, abs/2310.17420, 2023.

[BCMT22] Soheil Behnezhad, Moses Charikar, Weiyun Ma, and Li-Yang Tan. Almost 3-approximate correlation clustering in constant rounds. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731. IEEE, 2022.

[BHP20] Arturs Backurs and Sariel Har-Peled. Submodular clustering in low dimensions. *arXiv preprint arXiv:2004.05494*, 2020.

[BIO+19] Arturs Backurs, Piotr Indyk, Krzysztof Onak, Baruch Schieber, Ali Vakilian, and Tal Wagner. Scalable fair clustering. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 405–413. PMLR, 2019.

[BP12] Nikhil Bansal and Kirk Pruhs. Weighted geometric set multi-cover via quasi-uniform sampling. In Leah Epstein and Paolo Ferragina, editors, *Algorithms - ESA 2012 - 20th Annual European Symposium, Ljubljana, Slovenia, September 10-12, 2012. Proceedings*, volume 7501 of *Lecture Notes in Computer Science*, pages 145–156. Springer, 2012.

[BPR+17] Jarosław Byrka, Thomas Pensyl, Bartosz Rybicki, Aravind Srinivasan, and Khoa Trinh. An improved approximation for k-median and positive correlation in budgeted optimization. *ACM Transactions on Algorithms (TALG)*, 13(2):1–31, 2017.

[CAEMN22] Vincent Cohen-Addad, Hossein Esfandiari, Vahab Mirrokni, and Shyam Narayanan. Improved approximations for euclidean k-means and k-median, via nested quasi-independent sets. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1621–1628, 2022.

[CAFS21] Vincent Cohen-Addad, Andreas Emil Feldmann, and David Saulpic. Near-linear time approximation schemes for clustering in doubling metrics. *Journal of the ACM (JACM)*, 68(6):1–34, 2021.

[CAGH+22] Vincent Cohen-Addad, Anupam Gupta, Lunjia Hu, Hoon Oh, and David Saulpic. An improved local search algorithm for k-median. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1556–1612. SIAM, 2022.

[CALN22] Vincent Cohen-Addad, Euiwoong Lee, and Alantha Newman. Correlation clustering with sherali-adams. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 651–661. IEEE, 2022.

[CC84] Michele Conforti and Gérard Cornuéjols. Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the rado-edmonds theorem. *Discret. Appl. Math.*, 7(3):251–274, 1984.

[CCPV11] Gruia Calinescu, Chandra Chekuri, Martin Pal, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6):1740–1766, 2011.

[CDRS18] Steven Chaplick, Minati De, Alexander Ravsky, and Joachim Spoerhase. Approximation schemes for geometric coverage problems. In Yossi Azar, Hannah Bast, and Grzegorz Herman, editors, *26th Annual European Symposium on Algorithms, ESA 2018, August 20-22, 2018, Helsinki, Finland*, volume 112 of *LIPIcs*, pages 17:1–17:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[CEM+22] Vincent Cohen-Addad, Alessandro Epasto, Vahab Mirrokni, Shyam Narayanan, and Peilin Zhong. Near-optimal private and scalable k-clustering. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.

[CG13] Sanjay Chawla and Aristides Gionis. k-means–: A unified approach to clustering and outlier detection. In *Proceedings of the 2013 SIAM international conference on data mining*, pages 189–197. SIAM, 2013.

[CGQ15] Chandra Chekuri, Shalmoli Gupta, and Kent Quanrud. Streaming algorithms for submodular function maximization. In *Automata, Languages, and Programming: 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I 42*, pages 318–330. Springer, 2015.

[CGTS99] Moses Charikar, Sudipto Guha, Éva Tardos, and David B Shmoys. A constant-factor approximation algorithm for the k-median problem. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 1–10, 1999.

[Che06] Ke Chen. On k-median clustering in high dimensions. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 1177–1185. Citeseer, 2006.

[CKLV17] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. *Advances in neural information processing systems*, 30, 2017.

[CKM+21] Edith Cohen, Haim Kaplan, Yishay Mansour, Uri Stemmer, and Eliad Tsfadia. Differentially-private clustering of easy instances. In *International Conference on Machine Learning*, pages 2049–2059. PMLR, 2021.

[CM14] Radu Curticapean and Dániel Marx. Complexity of counting subgraphs: Only the boundedness of the vertex-cover number counts. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 130–139. IEEE Computer Society, 2014.

[CN06] Julia Chuzhoy and Joseph Naor. Covering problems with hard capacities. *SIAM J. Comput.*, 36(2):498–515, 2006.

[DEFI06] Erik D Demaine, Dotan Emanuel, Amos Fiat, and Nicole Immorlica. Correlation clustering in general weighted graphs. *Theoretical Computer Science*, 361(2-3):172–187, 2006.

[DF23] Mahsa Derakhshan and Alireza Farhadi. Beating (1 - 1/e)-approximation for weighted stochastic matching. In Nikhil Bansal and Viswanath Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 1931–1961. SIAM, 2023.

[DG14] Amit Dhurandhar and Karthik Gurumoorthy. Symmetric submodular clustering with actionable constraint. *arXiv preprint arXiv:1409.6967*, 2014.

[DGKR03] Irit Dinur, Venkatesan Guruswami, Subhash Khot, and Oded Regev. A new multilayered PCP and the hardness of hypergraph vertex cover. In Lawrence L. Larmore and Michel X. Goemans, editors, *Proceedings of the 35th Annual ACM Symposium on Theory of Computing, June 9-11, 2003, San Diego, CA, USA*, pages 595–601. ACM, 2003.

[DGRS05] Kedar Dhamdhere, Vineet Goyal, R. Ravi, and Mohit Singh. How to pay, come what may: Approximation algorithms for demand-robust covering problems. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings*, pages 367–378. IEEE Computer Society, 2005.

[DS19] Ni Ding and Parastoo Sadeghi. A submodularity-based clustering algorithm for the information bottleneck and privacy funnel. In *2019 IEEE Information Theory Workshop (ITW)*, pages 1–5, 2019.

[EvL08] Thomas Erlebach and Erik Jan van Leeuwen. Approximating geometric coverage problems. In Shang-Hua Teng, editor, *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2008, San Francisco, California, USA, January 20-22, 2008*, pages 1267–1276. SIAM, 2008.

[Fei98] Uriel Feige. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.

[FKRS19] Zachary Friggstad, Kamyar Khodamoradi, Mohsen Rezapour, and Mohammad R. Salavatipour. Approximation schemes for clustering with outliers. *ACM Trans. Algorithms*, 15(2):26:1–26:26, 2019.

[FLT04] Uriel Feige, László Lovász, and Prasad Tetali. Approximating min sum set cover. *Algorithmica*, 40:219–234, 2004.

[FMMM09] Jon Feldman, Aranyak Mehta, Vahab S. Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *50th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2009, October 25-27, 2009, Atlanta, Georgia, USA*, pages 117–126. IEEE Computer Society, 2009.

[FMV11] Uriel Feige, Vahab S Mirrokni, and Jan Vondrák. Maximizing non-monotone submodular functions. *SIAM Journal on Computing*, 40(4):1133–1153, 2011.

[FNSZ23] Moran Feldman, Ashkan Norouzi-Fard, Ola Svensson, and Rico Zenklusen. The one-way communication complexity of submodular maximization with applications to streaming and robustness. *J. ACM*, 70(4):24:1–24:52, 2023.

[FV10] Uriel Feige and Jan Vondrák. The submodular welfare problem with demand queries. *Theory of Computing*, 6(1):247–290, 2010.

[GK98] Sudipto Guha and Samir Khuller. Greedy strikes back: Improved facility location algorithms. In Howard J. Karloff, editor, *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, 25-27 January 1998, San Francisco, California, USA*, pages 649–657. ACM/SIAM, 1998.

[GKM20] Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. Differentially private clustering: Tight approximation ratios. *Advances in Neural Information Processing Systems*, 33:4040–4054, 2020.

[GR22] Christoph Grunau and Václav Rozhoň. Adapting k-means algorithms for outliers. In *International Conference on Machine Learning*, pages 7845–7886. PMLR, 2022.

[HCK+05] Ville Hautamäki, Svetlana Cherednichenko, Ismo Kärkkäinen, Tomi Kinnunen, and Pasi Fränti. Improving k-means by outlier removal. In *Scandinavian conference on image analysis*, pages 978–987. Springer, 2005.

[HG05] Katherine A Heller and Zoubin Ghahramani. Bayesian hierarchical clustering. In *Proceedings of the 22nd international conference on Machine learning*, pages 297–304, 2005.

[HS85] Dorit S Hochbaum and David B Shmoys. A best possible heuristic for the k-center problem. *Mathematics of operations research*, 10(2):180–184, 1985.

[IQM+20] Sungjin Im, Mahshid Montazer Qaem, Benjamin Moseley, Xiaorui Sun, and Rudy Zhou. Fast noise removal for k-means clustering. In *International Conference on Artificial Intelligence and Statistics*, pages 456–466. PMLR, 2020.

[JLN+05] Lujun Jia, Guolong Lin, Guevara Noubir, Rajmohan Rajaraman, and Ravi Sundaram. Universal approximations for tsp, steiner tree, and set cover. In Harold N. Gabow and Ronald Fagin, editors, *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 386–395. ACM, 2005.

[JLW+18] Kai Jin, Jian Li, Haitao Wang, Bowei Zhang, and Ningye Zhang. Near-linear time approximation schemes for geometric maximum coverage. *Theor. Comput. Sci.*, 725:64–78, 2018.

[JMM+03] Kamal Jain, Mohammad Mahdian, Evangelos Markakis, Amin Saberi, and Vijay V Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing lp. *Journal of the ACM (JACM)*, 50(6):795–824, 2003.

[JV01] Kamal Jain and Vijay V Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and lagrangian relaxation. *Journal of the ACM (JACM)*, 48(2):274–296, 2001.

[KKdV24] Heasung Kim, Hyeji Kim, and Gustavo de Veciana. Clustered federated learning via gradient-based partitioning. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024.

[KKT03] David Kempe, Jon M. Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In Lise Getoor, Ted E. Senator, Pedro M. Domingos, and Christos Faloutsos, editors, *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 24 - 27, 2003*, pages 137–146. ACM, 2003.

[KLS18] Ravishankar Krishnaswamy, Shi Li, and Sai Sandeep. Constant approximation for k-median and k-means with outliers via iterative rounding. In *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*, pages 646–659, 2018.

[KMZ+19] Ehsan Kazemi, Marko Mitrovic, Morteza Zadimoghaddam, Silvio Lattanzi, and Amin Karbasi. Submodular streaming in all its glory: Tight approximation, minimum memory and low adaptive complexity. In *International Conference on Machine Learning*, pages 3311–3320. PMLR, 2019.

[KPV13] Michael Kapralov, Ian Post, and Jan Vondrák. Online submodular welfare maximization: Greedy is optimal. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1216–1225. SIAM, 2013.

[KSG08] Andreas Krause, Ajit Paul Singh, and Carlos Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *J. Mach. Learn. Res.*, 9:235–284, 2008.

[LB11] Hui Lin and Jeff A. Bilmes. A class of submodular functions for document summarization. In Dekang Lin, Yuji Matsumoto, and Rada Mihalcea, editors, *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA*, pages 510–520. The Association for Computer Linguistics, 2011.

[LKG+07] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne M. VanBriesen, and Natalie S. Glance. Cost-effective outbreak detection in networks. In Pavel Berkhin, Rich Caruana, and Xindong Wu, editors, *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, California, USA, August 12-15, 2007*, pages 420–429. ACM, 2007.

[LLS+21] Bo Li, Lijun Li, Ankang Sun, Chenhao Wang, and Yingfan Wang. Approximate group fairness for clustering. In *International conference on machine learning*, pages 6381–6391. PMLR, 2021.

[LS13] Shi Li and Ola Svensson. Approximating k-median via pseudo-approximation. In *proceedings of the forty-fifth annual ACM symposium on theory of computing*, pages 901–910, 2013.

[LTK+22] Ekdeep Singh Lubana, Chi Ian Tang, Fahim Kawsar, Robert P. Dick, and Akhil Mathur. Orchestra: Unsupervised federated learning via globally consistent clustering. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato, editors, *International Conference on Machine Learning, ICML 2022,*

17-23 July 2022, Baltimore, Maryland, USA, volume 162 of *Proceedings of Machine Learning Research*, pages 14461–14484. PMLR, 2022.

[MC12] Fionn Murtagh and Pedro Contreras. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1):86–97, 2012.

[MMK08] Richard Matthew McCutchen and Samir Khuller. Streaming algorithms for k-center clustering with outliers and with anonymity. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques: 11th International Workshop, APPROX 2008, and 12th International Workshop, RANDOM 2008, Boston, MA, USA, August 25-27, 2008. Proceedings*, pages 165–178. Springer, 2008.

[MR09] Nabil H. Mustafa and Saurabh Ray. PTAS for geometric hitting set problems via local search. In John Hershberger and Efi Fogel, editors, *Proceedings of the 25th ACM Symposium on Computational Geometry, Aarhus, Denmark, June 8-10, 2009*, pages 17–22. ACM, 2009.

[MRR15] Nabil H. Mustafa, Rajiv Raman, and Saurabh Ray. Quasi-polynomial time approximation scheme for weighted geometric set cover on pseudodisks and halfspaces. *SIAM J. Comput.*, 44(6):1650–1669, 2015.

[MV20] Sepideh Mahabadi and Ali Vakilian. Individual fairness for k-clustering. In *International conference on machine learning*, pages 6586–6596. PMLR, 2020.

[MZH83] Nimrod Megiddo, Eitan Zemel, and S Louis Hakimi. The maximum coverage location problem. *SIAM Journal on Algebraic Discrete Methods*, 4(2):253–261, 1983.

[NJB05] Mukund Narasimhan, Nebojsa Jojic, and Jeff A Bilmes. Q-clustering. *Advances in Neural Information Processing Systems*, 18, 2005.

[NKI10] Kiyohito Nagano, Yoshinobu Kawahara, and Satoru Iwata. Minimum average cost clustering. *Advances in Neural Information Processing Systems*, 23, 2010.

[NWF78] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical programming*, 14:265–294, 1978.

[OR10] Krzysztof Onak and Ronitt Rubinfeld. Maintaining a large matching and a small vertex cover. In Leonard J. Schulman, editor, *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 457–464. ACM, 2010.

[SCL+16] Dong Su, Jianneng Cao, Ninghui Li, Elisa Bertino, and Hongxia Jin. Differentially private k-means clustering. In *Proceedings of the sixth ACM conference on data and application security and privacy*, pages 26–37, 2016.

[SKR+23] Joachim Spoerhase, Kamyar Khodamoradi, Benedikt Riegel, Bruno Ordozgoiti, and Aristides Gionis. A constant-factor approximation algorithm for reconciliation k-median. In *International Conference on Artificial Intelligence and Statistics*, pages 1719–1746. PMLR, 2023.

[SLS25] Jonathan Scott, Christoph H. Lampert, and David Saulpic. Differentially private federated k-means clustering with server-side data. In *Forty-second International Conference on Machine Learning*, 2025.

[STT07] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for lovasz-schrijver LP relaxations of vertex cover and max cut. In David S. Johnson and Uriel Feige, editors, *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 302–310. ACM, 2007.

[SVW17] Maxim Sviridenko, Jan Vondrák, and Justin Ward. Optimal approximation for submodular and supermodular optimization with bounded curvature. *Math. Oper. Res.*, 42(4):1197–1218, 2017.

[TA23] Henry Ling-Hei Tsang and Thomas D. Ahle. Clustering the sketch: Dynamic compression for embedding tables. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

[TIWB14] Sebastian Tschiatschek, Rishabh K. Iyer, Haochen Wei, and Jeff A. Bilmes. Learning mixtures of submodular functions for image collection summarization. In Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 1413–1421, 2014.

[Var10] Kasturi R. Varadarajan. Weighted geometric set cover via quasi-uniform sampling. In Leonard J. Schulman, editor, *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 641–648. ACM, 2010.

[Von13] Jan Vondrák. Symmetry and approximability of submodular maximization problems. *SIAM J. Comput.*, 42(1):265–304, 2013.

[WAT+23] Fulton Wang, Julius Adebayo, Sarah Tan, Diego Garcia-Olano, and Narine Kokhlikyan. Error discovery by clustering influence embeddings. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

[WIW+15] Kai Wei, Rishabh K Iyer, Shengjie Wang, Wenruo Bai, and Jeff A Bilmes. Mixed robust/average submodular partitioning: Fast algorithms, guarantees, and applications. *Advances in Neural Information Processing Systems*, 28, 2015.

A Improved approximation ratio

In this section, we prove Theorem 2. Recall that $\eta_\phi = \inf_{x \geq 0} \phi(3x)/\phi(x)$. The case $\eta_\phi = 0$ follows from the standard greedy bound, so assume $\eta_\phi > 0$. Put $c_1 = \frac{1}{2} \cdot 0.1 \cdot 0.004 \cdot 0.08$ and $\epsilon_\phi = (c_1/2)\eta_\phi$. We may assume $k \geq \max\{2000, 2/\epsilon_\phi\}$: otherwise Lemma 33 gives an additive $\Omega(\eta_\phi)$ improvement. Thus all subsequent rounding estimates, including $\lceil \epsilon_\phi k \rceil - 1 \geq \epsilon_\phi k/2$, hold. We use the normalization $f(\varnothing, P) = 0$ from Section 3. If $\mathrm{OPT} = 0$, or any prefix already attains OPT, greedy is optimal; otherwise all gap denominators below are positive.

Let $\mathcal{I} := \{t \in \mathbb{Z} : \lceil 0.005k \rceil \leq t < \lceil 0.01k \rceil\}$ denote the interval of iterations considered below.

We start by proving some preliminary tools in Section A.1 that are based on the properties of submodular functions. These tools establish that if either $f(\mathtt{alg}_t)$ or Δ_t deviate significantly from the values predicted in the standard submodular analysis, then the approximation ratio is already better than $1 - 1/e$. We then focus on the centers in time steps $t \in \mathcal{I}$ since for these tools we can ensure that Δ_t is approximately $\frac{\mathrm{OPT}}{k}$ (see Corollary 8). We group these centers into two groups of old and young centers (Definition 9), and separately handle the case where a large number of the centers are old (Section A.2) and a larger number are young (Section A.3).

A.1 Preliminary tools

If the standard submodular analysis is tight, then gap_t is exactly $(1 - 1/k)^t \mathrm{gap}_0$. The following lemma shows that if at any time step gap_t is significantly less than this, i.e., it is smaller by a constant factor, then the final approximation ratio is better than $1 - 1/e$.

Lemma 6. *If some time step t satisfies $\mathrm{gap}_t \leq (1 - \delta)(1 - 1/k)^t \mathrm{OPT}$ where $\delta \in (0, 1)$, then the approximation ratio is $\geq 1 - (1 - \delta)/e$. Consequently, if $\mathrm{gap}_t \leq (1 - 1/k)^t \mathrm{OPT} - c\mathrm{OPT}$ for some c, then the approximation ratio is $\geq 1 - 1/e + c/e$.*

Proof. For the first part, the proof follows from the fact that regardless of the value of gap at time t, it shrinks by a factor of $\leq 1 - 1/k$ in each step because $\Delta_t \geq \frac{1}{k}\mathrm{gap}_{t-1}$ by the standard analysis. Consequently, by step $t = k$, the gap is at most $(1 - \delta)(1 - 1/k)^k \mathrm{OPT}$. The claim now follows from the definition of gap_k and the fact that $\mathrm{gap}_0 = \mathrm{OPT}$. For the second part, $(1 - 1/k)^t \leq 1$ gives $(1 - 1/k)^t - c \leq (1 - c)(1 - 1/k)^t$, so the first part applies with $\delta = c$ (the case $c = 0$ is the standard bound). □

If the standard analysis is tight, then the marginal gain of u_t with respect to $\mathtt{alg}_{t-1}$ equals $(1 - 1/k)^{t-1} \frac{\mathrm{OPT}}{k}$. Let Δ_t^{tight} denote these values. Our next result states if for a sufficiently large t, specifically any $t \geq \Omega(k)$, the value Δ_t deviates significantly from Δ_t^{tight}, then the algorithm's approximation ratio is strictly better than $1 - 1/e$. Formally, we prove the following lemma.

Lemma 7. *Set $\Delta_t^{\text{tight}} = (1 - 1/k)^{t-1} \frac{\mathrm{OPT}}{k}$. Let $\epsilon, \delta \in (0, 1)$ be arbitrary parameters and assume that $k \geq 3/\min\{\epsilon, \delta\}$. If there exists $t \geq \epsilon k$ such that $\Delta_t \geq (1 + \delta)\Delta_t^{\text{tight}}$ or $\Delta_t \leq (1 - \delta)\Delta_t^{\text{tight}}$ then the approximation ratio is $1 - 1/e + \Omega(\delta \min(\epsilon, \delta))$.*

Proof. We first give a sketch of the proof, where we assume that δ and ϵ are relatively large, say 0.9 for simplicity. In both cases, we prove that by time t, the value of gap_t is smaller by a constant factor, let's say 0.9, compared to what we anticipate based on the tight example. This would prove the Lemma given Lemma 6.

First, let's examine the scenario where Δ_t is too small at time t. Since Δ_t is at least a $1/k$ fraction of gap_{t-1}, it indicates that the gap at time t is smaller than what we expect, thus the proof follows from Lemma 6. We next analyze the case where Δ_t is too large.

In the tight example, Δ_t^{tight} starts at OPT/k at $t = 1$ and reduces to approximately $\mathrm{OPT}/(ke)$ by $t = k$. Hence, the value of Δ_t effectively corresponds to $\Delta_{t'}^{\text{tight}}$ for some $t' < t - \Omega(k)$ (if $\Delta_t > \Delta_0^{\text{tight}}$, simply take $t' = 0$). We note however that Δ_t consistently decreases with t due to the nature of our algorithm. Specifically, if $f(u_t \mid \mathtt{alg}_{t-1}) < f(u_{t'} \mid \mathtt{alg}_{t'-1})$ for $t < t'$, then $f(u_t \mid \mathtt{alg}_{t-1}) < f(u_{t'} \mid \mathtt{alg}_{t-1})$, thus, the greedy approach would select $u_{t'}$ over u_t at time step t.

Consequently, for the $\Omega(k)$ time steps $\tilde{t} \in [t', t]$, the marginal gain $\Delta_{\tilde{t}}$ surpasses $\Delta_{\tilde{t}}^{\text{tight}}$, and for the $\Omega(k)$ time steps in $[\frac{t+t'}{2}, t]$, it exceeds it by a factor of $(1 + \Omega(1))$ due to:

$$\frac{\Delta_{t'}^{\text{tight}}}{\Delta_{(t+t')/2}^{\text{tight}}} = (1 - 1/k)^{(t'-t)/2} = 1 + \Omega(1)$$

Consequently, we gather an extra $\Omega(1)\Delta_t^{\text{tight}} = \Omega(\text{OPT}/k)$ value for $\Omega(k)$ steps, resulting in $f(\texttt{alg}_t)$ being $\Omega(\text{OPT})$ higher than expected, thus gap_t is a constant factor smaller than anticipated. Therefore, the proof follows from the outlined sketch as before.

Formally, if $\Delta_t \leq (1-\delta)\Delta_t^{\text{tight}}$, then

$$\begin{aligned}
\text{gap}_{t-1} &= f(\texttt{opt}) - f(\texttt{alg}_{t-1}) && \textit{(Definition of gap)} \\
&\leq f(\texttt{opt} \cup \texttt{alg}_{t-1}) - f(\texttt{alg}_{t-1}) && \textit{(Monotonicity of } f\textit{)} \\
&\leq f(\texttt{opt} \mid \texttt{alg}_{t-1}) && \textit{(Definition of } f(. \mid .)\textit{)} \\
&\leq k\Delta_t && \textit{(Greedy choice of algorithm)} \\
&\leq k(1-\delta)\Delta_t^{\text{tight}} && \textit{(Assumption on } \Delta_t\textit{)} \\
&\leq (1-\delta)(1-1/k)^{t-1}\text{OPT} && \textit{(Definition of } \Delta_t^{\text{tight}}\textit{)}.
\end{aligned}$$

Therefore, invoking Lemma 6 we conclude that the approximation ratio is at least $1 - 1/e + \delta/e$.

We now consider that case that $\Delta_t \geq (1+\delta)\Delta_t^{\text{tight}}$. Set ρ to be the smallest integer satisfying $(k/(k-1))^x \geq 1 + \delta/2$, i.e.,

$$\rho = \left\lceil \log_{k/(k-1)}(1+\delta/2) \right\rceil = \left\lceil \frac{\ln(1+\delta/2)}{\ln(k/(k-1))} \right\rceil.$$

Given the inequality $\ln(1+x) \leq x$ we have $\ln(k/(k-1)) = \ln(1 + \frac{1}{k-1}) \leq \frac{1}{k-1} \leq \frac{2}{k}$. Given the inequality $\ln(1+x) \geq x/2$ for $x \leq 1/2$, the assumption $\delta \leq 1$ implies $\ln(1+\delta/2) \geq \delta/4$. Therefore, $\rho \geq k\delta/8$. Set $t' = \max\{\, t - \rho, 0 \,\}$. Since $t \geq \epsilon k$ and $\rho \geq \Omega(k\delta)$, it follows that $t - t' \geq \Omega(k \min(\epsilon, \delta))$.

From the standard analysis of the greedy algorithm, we know that

$$\text{gap}_{t'} = f(\texttt{opt}) - f(\texttt{alg}_{t'}) \leq (1-1/k)^{t'}\text{OPT}. \tag{1}$$

Now, for all $\tilde{t} \in (t', t]$ we have

$$\begin{aligned}
\Delta_{\tilde{t}}^{\text{tight}} &= \Delta_t^{\text{tight}}(k/(k-1))^{t-\tilde{t}} && \textit{(Definition of } \Delta_t^{\text{tight}}\textit{)} \\
&\leq (k/(k-1))^{\rho-1}\Delta_t^{\text{tight}} && \textit{(Since } t' \geq t - \rho\textit{)} \\
&\leq (1+\delta/2)\Delta_t^{\text{tight}} && \textit{(Definition of } \rho\textit{)}.
\end{aligned}$$

We note however that Δ_t is non-increasing in t. This holds because if $\Delta_t > \Delta_{t'}$ for some $t > t'$, then

$$f_{t'}(u_t) \geq f_t(u_t) = \Delta_t > \Delta_{t'} = f_{t'}(u_{t'}).$$

Therefore, u_t would have been picked at time t' instead of $u_{t'}$, which is a contradiction. It follows that $\Delta_{\tilde{t}} \geq \Delta_t \geq (1+\delta)\Delta_t^{\text{tight}}$. Combining these two results we conclude that

$$\Delta_{\tilde{t}} \geq \Delta_{\tilde{t}}^{\text{tight}} + (\delta\Delta_t^{\text{tight}})/2. \tag{2}$$

Therefore,

$$\begin{aligned}
\text{gap}_t &= f(\texttt{opt}) - f(\texttt{alg}_t) && \textit{(Definition of gap)} \\
&= f(\texttt{opt}) - f(\texttt{alg}_{t'}) - \sum_{\tilde{t}=t'+1}^{t} \Delta_{\tilde{t}} && \textit{(Definition of } \Delta_{\tilde{t}}\textit{)} \\
&\leq (1-1/k)^{t'}\text{OPT} - \sum_{\tilde{t}=t'+1}^{t} \Delta_{\tilde{t}}^{\text{tight}} - \sum_{\tilde{t}=t'+1}^{t} \delta/2 \cdot \Delta_t^{\text{tight}} && \textit{(Equations (1) and (2))}
\end{aligned}$$

We note however that for any $\tilde{t}$,

$$\begin{aligned}(1-1/k)^{\tilde{t}-1}\text{OPT} - \Delta_{\tilde{t}}^{\text{tight}} &= (1-1/k)^{\tilde{t}-1}\text{OPT} - (1-1/k)^{\tilde{t}-1}\text{OPT}/k \quad \textit{(Definition of } \Delta_{\tilde{t}}^{\text{tight}}\textit{)} \\ &= (1-1/k)^{\tilde{t}}\text{OPT}\end{aligned}$$

It follows that

$$\begin{aligned}&(1-1/k)^{t'}\text{OPT} - \sum_{\tilde{t}=t'+1}^{t} \Delta_{\tilde{t}}^{\text{tight}} \\ &= (1-1/k)^{t'}\text{OPT} - \sum_{\tilde{t}=t'+1}^{t} \Big((1-1/k)^{\tilde{t}-1}\text{OPT} - (1-1/k)^{\tilde{t}}\text{OPT} \Big) \\ &= (1-1/k)^{t}\text{OPT} \qquad \textit{(Telescoping sum)}.\end{aligned}$$

Therefore,

$$\begin{aligned}\text{gap}_t &\le (1-1/k)^{t}\text{OPT} - \sum_{\tilde{t}=t'+1}^{t} \delta\Delta_t^{\text{tight}}/2 \\ &\le (1-1/k)^{t}\text{OPT} - |\, t-t' \,| \, \frac{\delta\text{OPT}}{2ke} \qquad \textit{(Since } \Delta_t^{\text{tight}} \ge \text{OPT}/ke\textit{)} \\ &= (1-1/k)^{t}\text{OPT} - \Omega(\delta \min\{\,\epsilon, \delta\,\})\text{OPT} \qquad \textit{(Since } |\, t-t' \,| \ge \Omega(k \min\{\,\epsilon,\delta\,\})\textit{)},\end{aligned}$$

which finishes the proof given Lemma 6. □

As an immediate corollary, we obtain the following result which effectively upper and lower bounds Δ_t for $t \in \mathcal{I}$.

Corollary 8. *Define* $\Delta^{\text{base}} := \frac{\text{OPT}}{k}$. *If there exists* $t \in \mathcal{I}$ *such that* $\Delta_t/\Delta^{\text{base}} \notin (0.98, 1)$, *then the approximation ratio is* $1 - 1/e + \Omega(1)$.

Proof. Note that $\Delta_t^{\text{tight}} = \frac{1}{k}(1-\frac{1}{k})^{t-1}\text{OPT}$. Therefore,

$$\begin{aligned}\frac{\Delta_t^{\text{tight}}}{\Delta^{\text{base}}} &= (1-\frac{1}{k})^{t-1} \\ &\in [(1-\frac{1}{k})^{0.01k}, (1-\frac{1}{k})^{0.005k-1}] \qquad \textit{(Since } t \in \mathcal{I}\textit{)} \\ &\in (0.99, 0.996)\end{aligned}$$

where the last inequality holds because of the Bernoulli inequality, $(1-x)^r \ge 1-rx$ for every real number $r \ge 1$ and $x \in [0,1]$, and because of the inequality $(1+x)^r \le e^{rx}$, using $k \ge 2000$ so that $e^{-0.005+1/k} < 0.996$.

Therefore, we have $\Delta^{\text{base}} < \frac{100}{99}\Delta_t^{\text{tight}}$. Hence, $\Delta_t \le 0.98\Delta^{\text{base}}$ implies that $\Delta_t < \frac{98}{99}\Delta_t^{\text{tight}}$. Therefore, if $\Delta_t \le 0.98\Delta^{\text{base}}$, the proof follows from Lemma 7. Similarly, since $\Delta^{\text{base}} \ge \frac{1}{0.996}\Delta_t^{\text{tight}}$, if $\Delta_t \ge \Delta^{\text{base}}$ then we have $\Delta_t \ge \frac{1}{0.996}\Delta_t^{\text{tight}}$, and again the claim follows from Lemma 7. □

With the above corollary established, we proceed with the proof. Let $\texttt{alg}' := \texttt{alg}_{\lceil 0.01k \rceil - 1} \backslash \texttt{alg}_{\lceil 0.005k \rceil - 1}$ denote the set of all centers chosen by the algorithm in the interval $\mathcal{I}$. Set $\epsilon_\phi = \frac{c_1}{2}\eta_\phi$ where $c_1 = \frac{1}{2} \cdot 0.1 \cdot 0.004 \cdot 0.08$; the reason behind this choice will become clear later on. We drop the dependence on ϕ when it is clear from context. We will analyze these centers and show, using a case by case analysis, that the algorithm obtains an "extra" $\Omega(\eta_\phi^2\text{OPT})$ value during the interval $[0, 0.02k)$, where "extra" is in comparison to the lower bounds in the standard submodular analysis. This extra amount leads to an improved approximation ratio given Lemma 6.

Consider a center $u_t \in \texttt{alg}'$. Its marginal gain Δ_t can be decomposed as $\sum_{p \in P} f(u_t, p \mid \texttt{alg}_{t-1})$. We start by classifying the centers u_t into two groups based on the points that mainly contribute to Δ_t.

Definition 9 (young and old centers)**.** *We say a point p is* young *for u_t if at time ϵk, only a $\eta/2$ fraction from $f(u_t, p)$ is served and let P_t^{yo} denote the set of all young points. Formally,*

$$P_t^{\text{yo}} := \left\{ p : f(\texttt{alg}_{\lceil \epsilon k \rceil - 1}, p) \leq \eta/2 f(u_t, p) \right\}.$$

We say that a center u_t is young if its marginal gain from young points is at least $0.9\Delta^{\text{base}}$ and let $\texttt{alg}^{\text{yo}}$ denote the set of all young centers. Formally,

$$\texttt{alg}^{\text{yo}} := \left\{ u_t \in \texttt{alg}' : f_t(u_t, P_t^{\text{yo}}) \geq 0.9\Delta^{\text{base}} \right\}. \tag{3}$$

We similarly use P_t^{old} and $\texttt{alg}^{\text{old}}$ to denote the set of old points (i.e., points that are not young) for u_t and the set of old centers (i.e., centers that are not young) respectively.

We will separately analyze the case where most of the centers in $\texttt{alg}'$ are old (Section A.2) and the case where most of them are young (Section A.3). We first consider the case where most centers are old.

A.2 Large number of old centers

In this section, we prove the following lemma.

Lemma 10. *Let $\epsilon = \frac{c_1}{2}\eta$ where $c_1 = \frac{1}{2} \cdot 0.1 \cdot 0.004 \cdot 0.08$. If $\left| \texttt{alg}^{\text{old}} \right| \geq 0.1 \left| \texttt{alg}' \right|$ then the approximation ratio of the algorithm is $\geq 1 - 1/e + \Omega(\eta)$.*

Consider a bipartite graph where the vertices on the left are $\texttt{alg}^{\text{old}}$ and the vertices on the right are $P_\cup^{\text{old}} := \cup_{u_t \in \texttt{alg}^{\text{old}}} P_t^{\text{old}}$. For each center $u_t \in \texttt{alg}^{\text{old}}$ and $p \in P_t^{\text{old}}$, we put an edge between u_t and p with the weight $f_t(u_t, p)$. We will separately bound the sum of these weights from the right and left sides to prove the approximation ratio. We start by bounding the weights from the left hand side. For each center $u_t \in \texttt{alg}^{\text{old}}$, we use the assumption $u_t \in \texttt{alg}^{\text{old}}$ to show that $\sum_{p \in P_t^{\text{old}}} f_t(u_t, p) \geq (0.98 - 0.9)\Delta^{\text{base}} = 0.08\Delta^{\text{base}}$. Therefore, the sum of the edges is at least $\Omega(\text{OPT})$.

Now let us bound this sum from the right side. For each $p \in P_\cup^{\text{old}}$, we claim that the sum of the edges connected to it is at most $\frac{2}{\eta} f(\texttt{alg}_{\epsilon k}, p)$. Let $\texttt{alg}^{\text{old}}(p)$ denote the neighbors of p in the graph, i.e., $\left\{ u_t \in \texttt{alg}^{\text{old}} : p \in P_t^{\text{old}} \right\}$. We first prove that the sum of the edges connected to p is upper bounded by $\max_{u_{t'} \in \texttt{alg}^{\text{old}}(p)} f(u_{t'}, p)$. This is because each $u_{t'} \in \texttt{alg}^{\text{old}}(p)$ contributes its marginal gain to this sum, which means the sum is at most $f(\texttt{alg}^{\text{old}}(p), p)$. By definition, we have $f(\texttt{alg}^{\text{old}}(p), p) = \max_{u_{t'} \in \texttt{alg}^{\text{old}}(p)} f(u_{t'}, p)$. Assume this maximum is attained by u_t. By definition of old points, $p \in P_t^{\text{old}}$ implies that $f(u_t, p) \leq \frac{2}{\eta} f(\texttt{alg}_{\epsilon k}, p)$. This proves the desired claim. Summing over P we conclude that the sum of the edges in the graph is at most $\frac{2}{\eta} f(\texttt{alg}_{\epsilon k})$.

Combining the upper and lower bounds and keeping track of the constants, we conclude that $f(\texttt{alg}_{\epsilon k}) \geq c_1 \eta_\phi \text{OPT}$. The lower bound under standard analysis however is $\approx \frac{c_1}{2}\eta_\phi \text{OPT}$ because we've added only ϵk centers, each center has marginal gain $\Delta_t \approx \text{OPT}/k$, and $\epsilon = c_1\eta_\phi/2$. This means that we have $\Omega(\eta_\phi \text{OPT})$ more value than we expect under the standard analysis, which allows us to finish the proof using Lemma 6.

We now proceed with a formal proof.

Lemma 11. *For any $p \in P_\cup^{\text{old}}$,*

$$\sum_{u_{t'} \in \texttt{alg}^{\text{old}}(p)} f_{t'}(u_{t'}, p) < \frac{2}{\eta} f(\texttt{alg}_{\lceil \epsilon k \rceil - 1}, p)$$

Proof. If all the marginal gains in the sum are zero, the claim is immediate: $p \in P_\cup^{\text{old}}$ implies $f(\texttt{alg}_{\lceil \epsilon k \rceil - 1}, p) > 0$. Otherwise, let u_t denote the last (i.e., largest t) center in $\texttt{alg}^{\text{old}}(p)$ such that $f(u_t, p \mid \texttt{alg}_{t-1}) > 0$.

We know that $f(\texttt{alg}_t, p) = \max_{u \in \texttt{alg}_t} f(u, p)$, and $f(\texttt{alg}_{t-1}, p) = \max_{u \in \texttt{alg}_{t-1}} f(u, p)$. Therefore, the nonzero marginal gain implies that

$$f(\texttt{alg}_t, p) = \max_{u \in \texttt{alg}_t} f(u, p) = f(u_t, p) \tag{4}$$

By the definition of $f(.,. \mid .)$, we have

$$f(\texttt{alg}_t, p) = \sum_{t'=1}^{t} f(u_{t'}, p \mid \texttt{alg}_{t'-1}). \tag{5}$$

By combining (4) and (5) we get the following:

$$\begin{aligned}
f(u_t, p) &= \sum_{t'=1}^{t} f(u_{t'}, p \mid \texttt{alg}_{t'-1}) && \textit{(Definition of marginal gain)} \\
&\geq \sum_{t'=1}^{t} f(u_{t'}, p \mid \texttt{alg}_{t'-1}) \mathbb{1}\left\{u_{t'} \in \texttt{alg}^{\text{old}}(p)\right\} && \textit{(Non-negativity of marginal gain)} \\
&= \sum_{u_{t'} \in \texttt{alg}^{\text{old}}(p)} f(u_{t'}, p \mid \texttt{alg}_{t'-1})
\end{aligned}$$

However, by definition of P_t^{old}, we have $f(u_t, p) < \frac{2}{\eta} f(\texttt{alg}_{\lceil \epsilon k \rceil - 1}, p)$, finishing the proof. □

Proof of Lemma 10. We calculate $\sum_{u_t \in \texttt{alg}^{\text{old}}} \sum_{p \in P_t^{\text{old}}} f_t(u_t, p)$ in two ways.

For each center $u_t \in \texttt{alg}^{\text{old}}$, we know by the definition of old centers (Definition 9) that $f_t(u_t, P_t^{\text{yo}}) < 0.9\Delta^{\text{base}}$. Therefore,

$$\begin{aligned}
\sum_{p \in P_t^{\text{old}}} f_t(u_t, p) &= f_t(u_t, P_t^{\text{old}}) && \textit{(Definition of } f(.,.)\textit{)} \\
&= f_t(u_t) - f_t(u_t, P_t^{\text{yo}}) && \textit{(Definition of } P^{\text{old}}\textit{)} \\
&> f_t(u_t) - 0.9\Delta^{\text{base}} && \textit{(Since } u_t \in \texttt{alg}^{\text{old}}\textit{)} \\
&= \Delta_t - 0.9\Delta^{\text{base}} && \textit{(Definition of } \Delta_t\textit{)} \\
&> 0.98\Delta^{\text{base}} - 0.9\Delta^{\text{base}} \\
&= 0.08\Delta^{\text{base}},
\end{aligned}$$

where for the final inequality we have used Corollary 8; specifically, we assumed that $\Delta_t \geq 0.98\Delta^{\text{base}}$ since otherwise the claim follows from the Corollary.

Since we have assumed that more than 0.1 fraction of centers in $\texttt{alg}'$ are old, we obtain the following bound.

$$\sum_{u_t \in \texttt{alg}^{\text{old}}} \sum_{p \in P_t^{\text{old}}} f_t(u_t, p) > 0.1 \,|\, \texttt{alg}' \,|\, 0.08\Delta^{\text{base}}$$

The left hand side of the above inequality can be rewritten as follows because of double counting:

$$\sum_{u_t \in \texttt{alg}^{\text{old}}} \sum_{p \in P_t^{\text{old}}} f_t(u_t, p) = \sum_{p \in P_{\cup}^{\text{old}}} \sum_{u_t \in \texttt{alg}^{\text{old}}(p)} f_t(u_t, p)$$

We can further upper bound this as

$$\begin{aligned}
\sum_{p \in P_{\cup}^{\text{old}}} \sum_{u_t \in \texttt{alg}^{\text{old}}(p)} f_t(u_t, p) &< \sum_{p \in P_{\cup}^{\text{old}}} \frac{2}{\eta} f(\texttt{alg}_{\lceil \epsilon k \rceil - 1}, p) && \textit{(Lemma 11)} \\
&\leq \sum_{p \in P} \frac{2}{\eta} f(\texttt{alg}_{\lceil \epsilon k \rceil - 1}, p) && \textit{(non-negativity of } f\textit{)} \\
&= \frac{2}{\eta} f(\texttt{alg}_{\lceil \epsilon k \rceil - 1})
\end{aligned}$$

By combining the two bounds, we get:

$$f(\texttt{alg}_{\lceil \epsilon k \rceil - 1}) > \frac{\eta}{2} \cdot 0.1 \cdot |\, \texttt{alg}' \,| \cdot 0.08 \cdot \Delta^{\text{base}} \geq 0.1 \cdot 0.004 \cdot 0.08 \cdot \frac{\eta}{2} \cdot \text{OPT} = c_1 \eta \text{OPT},$$

Therefore, $\text{gap}_{\lceil \epsilon k \rceil - 1} \leq \text{OPT}(1 - c_1\eta)$

We note however that

$$(1 - 1/k)^{\lceil \epsilon k \rceil - 1} \geq 1 - \frac{\lceil \epsilon k \rceil - 1}{k} \geq 1 - \frac{\epsilon k}{k} = 1 - \frac{c_1}{2}\eta.$$

Therefore, given Lemma 6, we have $\frac{f(\texttt{alg}_k)}{\text{OPT}} \geq 1 - 1/e + \frac{c_1}{2e}\eta$, finishing the proof. □

A.3 Large number of young centers

We now focus on the case where most of the centers in $\texttt{alg}'$ are young; specifically, we assume that $|\,\texttt{alg}^{\text{yo}}\,| \geq 0.9\,|\,\texttt{alg}'\,|$. By definition, all these centers satisfy $f_t(u_t, P_t^{\text{yo}}) \geq 0.9\Delta^{\text{base}}$. We further classify the centers into three groups based on the "source" of the marginal gain $f_t(u_t, P_t^{\text{yo}})$, i.e., the points that contribute to this marginal gain.

Recall from Section 3 that $u^*_{\min}(p) = \arg\min_{u^* \in \texttt{opt}} d(u^*, p)$ denotes the closest center to p in $\texttt{opt}$ and that $\texttt{zone}(u^*) := \{\, p : u^*_{\min}(p) = u^* \,\}$ denotes the set of points in P for which the closest center in $\texttt{opt}$ is u^*. We further recall that $\texttt{zone}_t^{\max}$ denotes $\texttt{zone}(u^*_{\min}(u_t))$. The classification is as follows.

1. We say u_t is *independent* if

   $$f(u_t \mid \texttt{opt} \cup \texttt{alg}_{t-1}) \geq 0.01\Delta^{\text{base}}.$$

   We use $\texttt{alg}^{\text{ind}}$ to denote the set of all independent centers.

2. We say u_t is *concentrated* if it is not independent and

   $$f_t(u_t, \texttt{zone}_t^{\max}) \geq 0.1\Delta^{\text{base}}.$$

   We use $\texttt{alg}^{\text{con}}$ to denote the set of all concentrated centers.

3. Finally, if u_t is not independent and is not concentrated, we say it is *triangle* and use $\texttt{alg}^{\text{tri}}$ to denote the set of all triangle centers.

Since 90% of the centers in $\texttt{alg}'$ are young and all young centers fall into one of the above three groups, at least one group is large (formally, it contains 30% of the centers in $\texttt{alg}'$). We next separately analyze the case where each group is large and show why we get the desired approximation ratio.

A.3.1 Large number of independent centers

The following lemma shows that if the number of independent centers is large, then we get the desired approximation ratio

Lemma 12. *If* $\left|\,\texttt{alg}^{\text{ind}}\,\right| \geq 0.3\,|\,\texttt{alg}'\,|$ *then the approximation ratio is* $\geq 1 - 1/e + \Omega(1)$.

Proof. We first give a sketch of the proof. By definition, independent centers have a lot of marginal gain from points for which they service more than $\texttt{opt}$. Therefore, the large number of independent centers implies that $f(\texttt{opt} \cup \texttt{alg}_{t-1})$ is significantly larger than $f(\texttt{opt})$ for $t \geq 0.01k$. Specifically, since 0.01 fraction of the marginal gain of each center is not served by $\texttt{opt}$, and we have at least $\Omega(k)$ independent centers in the interval $\mathcal{I}$, we conclude $f(\texttt{opt} \cup \texttt{alg}_t) - f(\texttt{opt}) \geq \Omega(\text{OPT})$ for any $t \geq 0.01k$ (Equation 6). This allows us to refine our lower bounds on Δ_t in the standard submodular analysis because this analysis relied on the inequality $f(\texttt{opt} \cup \texttt{alg}_t) \geq f(\texttt{opt})$ (this is equivalent to the inequality $f(\texttt{opt} \mid \texttt{alg}_t) \geq \text{gap}_t$ used in the proof) which in our case is no longer tight.

We now proceed with a formal proof. Let $\{ u_{t_1}, \dots, u_{t_s} \} = \texttt{alg}^{\text{ind}}$ where $t_1 < t_2 < \cdots < t_s$. Define $\texttt{alg}^{\text{ind}}_i := \{ u_{t_1}, \dots, u_{t_i} \}$.

$$\begin{aligned}
f(\texttt{alg}^{\text{ind}} \mid \texttt{opt}) &= \sum_{i=1}^{s} \big(f(\texttt{alg}^{\text{ind}}_i \mid \texttt{opt}) - f(\texttt{alg}^{\text{ind}}_{i-1} \mid \texttt{opt}) \big) && \textit{(Telescoping sum)} \\
&= \sum_{i=1}^{s} f(u_{t_i} \mid \texttt{opt} \cup \texttt{alg}^{\text{ind}}_{i-1}) && \textit{(Definition of } f(. \mid .)\textit{)} \\
&\geq \sum_{i=1}^{s} f(u_{t_i} \mid \texttt{opt} \cup \texttt{alg}_{t_i - 1}) && \textit{(Submodualrity and } \texttt{alg}^{\text{ind}}_{i-1} \subseteq \texttt{alg}_{t_i-1}\textit{)} \\
&\geq 0.01 \,\big|\, \texttt{alg}^{\text{ind}} \,\big|\, \Delta^{\text{base}} && \textit{(Definition of } \texttt{alg}^{\text{ind}}\textit{)} \\
&\geq 0.01 \cdot 0.3 \cdot 0.004k \cdot \Delta^{\text{base}} && \textit{(Assumption } \big|\, \texttt{alg}^{\text{ind}} \,\big| \geq 0.3 \,|\, \texttt{alg}' \,|\textit{)},
\end{aligned}$$

which is $\geq \Omega(\text{OPT})$ because $k\Delta^{\text{base}} = \text{OPT}$.

Fix any value $t \geq \lceil\, 0.01k \,\rceil$. Since $\texttt{alg}^{\text{ind}} \subseteq \texttt{alg}' \subseteq \texttt{alg}_{t-1}$, we conclude

$$f(\texttt{alg}_{t-1} \mid \texttt{opt}) \geq c' \text{OPT}. \tag{6}$$

for some constant c'. Refining the original analysis of the greedy algorithm from Section 4, for all $t \geq \lceil\, 0.01k \,\rceil$,

$$\begin{aligned}
k\Delta_t &\geq f(\texttt{opt} \cup \texttt{alg}_{t-1}) - f(\texttt{alg}_{t-1}) && \textit{(Spread step)} \\
&= \big(f(\texttt{opt} \cup \texttt{alg}_{t-1}) - f(\texttt{opt}) \big) + \big(f(\texttt{opt}) - f(\texttt{alg}_{t-1}) \big) \\
&= f(\texttt{opt}) - f(\texttt{alg}_{t-1}) + f(\texttt{alg}_{t-1} \mid \texttt{opt}) && \textit{(Definition of } f(. \mid .)\textit{)} \\
&\geq f(\texttt{opt}) - f(\texttt{alg}_{t-1}) + c' \text{OPT} && \textit{(Equation (6))} \\
&\geq (1 + c') \big(f(\texttt{opt}) - f(\texttt{alg}_{t-1}) \big) && \textit{(Since } f(\texttt{alg}_{t-1}) \geq 0 \textit{).}
\end{aligned}$$

By definition of Δ_t, this implies

$$f(\texttt{opt}) - f(\texttt{alg}_t) = f(\texttt{opt}) - f(\texttt{alg}_{t-1}) - \Delta_t \leq \left(1 - \frac{1 + c'}{k} \right) \big(f(\texttt{opt}) - f(\texttt{alg}_{t-1}) \big) .$$

Since $f(\texttt{opt}) - f(\texttt{alg}_t) \leq (1 - \frac{1}{k}) \big(f(\texttt{opt}) - f(\texttt{alg}_{t-1}) \big)$ holds for $t < \lceil\, 0.01k \,\rceil$, we obtain

$$\begin{aligned}
\frac{f(\texttt{opt}) - f(\texttt{alg}_k)}{f(\texttt{opt})} &= \prod_{t < \lceil\, 0.01k \,\rceil} \frac{f(\texttt{opt}) - f(\texttt{alg}_t)}{f(\texttt{opt}) - f(\texttt{alg}_{t-1})} \prod_{t \geq \lceil\, 0.01k \,\rceil} \frac{f(\texttt{opt}) - f(\texttt{alg}_t)}{f(\texttt{opt}) - f(\texttt{alg}_{t-1})} && \textit{(Telescoping product)} \\
&\leq \left(1 - \frac{1}{k} \right)^{\lceil\, 0.01k \,\rceil - 1} \cdot \left(1 - \frac{1 + c'}{k} \right)^{k - \lceil\, 0.01k \,\rceil + 1} \\
&\leq (1/e)^{1 + c'(k - \lceil\, 0.01k \,\rceil + 1)/k} && \textit{(Since } 1 - x \leq e^{-x}\textit{)} \\
&\leq (1/e)^{1 + 0.99c'} = (1/e) - \Omega(1),
\end{aligned}$$

which implies $f(\texttt{alg}) \geq (1 - 1/e + \Omega(1))\text{OPT}$ by rearranging. □

A.3.2 Large number of concentrated centers

Our main result in this section is that if $|\, \texttt{alg}^{\text{con}} \,|$ is large, then the approximation ratio is better than $1 - 1/e$. Formally, we prove the following lemma.

Lemma 13. *If* $|\, \texttt{alg}^{\text{con}} \,| \geq 0.3 \,|\, \texttt{alg}' \,|$ *then the approximation ratio is* $\geq 1 - 1/e + \Omega(1)$.

Throughout the rest of this subsection, assume $|\texttt{alg}^{\text{con}}| \geq 0.3|\texttt{alg}'|$ and $\Delta_t \in (0.98\Delta^{\text{base}}, \Delta^{\text{base}})$ for $t \in \mathcal{I}$; the complementary marginal case already proves Lemma 13 by Corollary 8.

We start with some definitions that are used in the proofs. Define $\texttt{rem}_t := f(\texttt{opt} \mid \texttt{alg}_{t-1})$ and $\texttt{rem}_t(u^*) := f(\texttt{opt}, \texttt{zone}(u^*) \mid \texttt{alg}_{t-1})$. Intuitively, $\texttt{rem}_t(u^*)$ captures the amount of service "remaining" from the zone corresponding to u^*, i.e., the amount of service we would gain within this

zone by adding u^* to $\texttt{alg}_{t-1}$. We note that $\texttt{rem}_t(u^*)$ and $\texttt{rem}_t$ are both decreasing in t since $f(.)$ is submodular.

Given Corollary 8, for all $t \in \mathcal{I}$ we can assume that $\Delta_t \leq \Delta^{\text{base}}$ since otherwise the Lemma holds trivially. By the greedy choice of the algorithm, this implies

$$\max_{u^*} \texttt{rem}_t(u^*) \leq \Delta^{\text{base}} \qquad \text{for all } t \geq \lceil 0.005k \rceil. \tag{7}$$

For later times, the same bound follows from the monotonicity of $\texttt{rem}_t(u^*)$. For any $u_t \in \texttt{alg}^{\text{con}}$, we recall from Section 3 that $u^*_{\min}(u_t)$ denotes the closest point to u_t in $\texttt{opt}$.

Proof sketch We first show that if $u_t \in \texttt{alg}^{\text{con}}$ and $u^*_{\min}(u_t) = u^*$, then $\texttt{rem}_t(u^*) - \texttt{rem}_{t+1}(u^*) \geq 0.09\Delta^{\text{base}}$ (Lemma 14). In other words, $\texttt{rem}_t(u^*)$ decreases significantly after taking u_t. Let $\texttt{opt}^{\text{con}} \subseteq \texttt{opt}$ denote the set of all such u^* for all $u_t \in \texttt{alg}^{\text{con}}$. Together with Equation (7), this decrease has two important implications. Firstly, any u^* can be $u^*_{\min}(u_t)$ only $O(1)$ times which means the size of $\texttt{opt}^{\text{con}}$ is at least $\Omega(k)$ (Lemma 15). Secondly, since $\texttt{rem}_t$ is decreasing in t, we have $\texttt{rem}_t(u^*) \leq 0.91\Delta^{\text{base}}$ for $u^* \in \texttt{opt}^{\text{con}}$ and $t \geq 0.01k$ (Lemma 16).

Focus on time $t = \lceil 0.02k \rceil$. If the average value for $\texttt{rem}_t(u^*)$, i.e., $\texttt{rem}_t/k$, is less than $0.97\Delta^{\text{base}}$, then we can show that the algorithm has already served more than the lower bounds obtained in the submodular analysis, which in turn allows us to get a better approximation ratio than $1 - 1/e$ using Lemma 6 (see Lemma 17). Otherwise, for all $t \in [0.01k, 0.02k)$, the value $\texttt{rem}_t(u^*)$ for $u^* \in \texttt{opt}^{\text{con}}$ is significantly (specifically, at least $(0.97 - 0.91)\Delta^{\text{base}} = 0.06\Delta^{\text{base}}$) less than the average value $\texttt{rem}_t/k$. Since $|\,\texttt{opt}^{\text{con}}\,| \geq \Omega(k)$, this means that there is a large "deficit" of $\Omega(k\Delta^{\text{base}})$ in $\texttt{rem}_t$ from these centers (compared to the average) which must be offset by an equally large "surplus" from $\texttt{opt} \backslash \texttt{opt}^{\text{con}}$ (Lemma 18). This in turn means that $\texttt{rem}_t(u^*) \geq \texttt{rem}_t/k + \Omega(\Delta^{\text{base}})$ for some $u^* \in \texttt{opt} \backslash \texttt{opt}^{\text{con}}$. By the greedy choice of u_t, this means that $\Delta_t \geq \texttt{rem}_t(u^*) \geq (1 + \Omega(1))\text{gap}_t/k$, which improves on the $\Delta_t \geq \text{gap}_t/k$ lower bound used in the standard submodular analysis (Lemma 19). Specifically, this means that during $t = [0.01k, 0.02k)$, the difference $f(\texttt{opt}) - f(\texttt{alg}_t)$ decreases with the rate of $1 - \frac{1+\Omega(1)}{k}$ instead of $1 - \frac{1}{k}$, leading to a bound of $1 - 1/e + \Omega(1)$ on the overall approximation ratio (Lemma 20)).

We now proceed with a formal proof. Define $\texttt{opt}^{\text{con}} := \{\, u^*_{\min}(u_t) : u_t \in \texttt{alg}^{\text{con}} \,\}$. We start by establishing the mentioned properties for $\texttt{opt}^{\text{con}}$.

Lemma 14. *If $u_t \in \texttt{alg}^{\text{con}}$ and $u^*_{\min}(u_t) = u^*$,*

$$\texttt{rem}_t(u^*) - \texttt{rem}_{t+1}(u^*) \geq 0.09\Delta^{\text{base}}.$$

Proof. Let Z denote $\texttt{zone}(u^*)$ for simplicity. Since $u_t \in \texttt{alg}^{\text{con}}$, the assumption $u^*_{\min}(u_t) = u^*$ implies $f(u_t, Z \mid \texttt{alg}_{t-1}) \geq 0.1\Delta^{\text{base}}$. In other words, when focusing on $\texttt{zone}(u^*)$, the center u_t has $0.1\Delta^{\text{base}}$ marginal gain with respect to $\texttt{alg}_{t-1}$. By the assumption $u_t \notin \texttt{alg}^{\text{ind}}$, we know that at most $0.01\Delta^{\text{base}}$ of this gain is from areas that are not served by $\texttt{opt}$. Therefore, the remaining $0.09\Delta^{\text{base}}$ of the gain must be from areas served by $\texttt{opt}$, which means that adding u_t to $\texttt{alg}_{t-1}$ reduces the marginal gain of $\texttt{opt}$ for the points in Z by $0.09\Delta^{\text{base}}$ as claimed.

Formally,

$$\begin{aligned} \texttt{rem}_{t+1}(u^*) &= f(\texttt{opt}, Z \mid \texttt{alg}_t) && \textit{(Definition of } \texttt{rem}\textit{)} \\ &= f(\texttt{opt} \cup \texttt{alg}_t, Z) - f(\texttt{alg}_t, Z) && \textit{(Definition of } f(. \mid .)\textit{)}, \end{aligned} \tag{8}$$

Since u_t is in $\texttt{alg}^{\text{con}}$, we conclude that it is in $\texttt{alg}^{\text{yo}}$ but not in $\texttt{alg}^{\text{ind}}$ which by definition of $\texttt{alg}^{\text{ind}}$ implies $f(u_t \mid \texttt{opt} \cup \texttt{alg}_{t-1}) \leq 0.01\Delta^{\text{base}}$. We can therefore upper bound the first term as

$$\begin{aligned} f(\texttt{opt} \cup \texttt{alg}_t, Z) &= f(\texttt{opt} \cup \texttt{alg}_{t-1}, Z) + f(u_t, Z \mid \texttt{opt} \cup \texttt{alg}_{t-1}) && \textit{(Since } \texttt{alg}_t = \texttt{alg}_{t-1} \cup \{\, u_t \,\}\textit{)} \\ &\leq f(\texttt{opt} \cup \texttt{alg}_{t-1}, Z) + f(u_t \mid \texttt{opt} \cup \texttt{alg}_{t-1}) && \textit{(Since } Z \subseteq P\textit{)} \\ &< f(\texttt{opt} \cup \texttt{alg}_{t-1}, Z) + 0.01\Delta^{\text{base}} && \textit{(Since } u_t \notin \texttt{alg}^{\text{ind}}\textit{)}. \end{aligned} \tag{9}$$

As for the second term, we can lower bound it as

$$\begin{aligned} f(\texttt{alg}_t, Z) &= f(\texttt{alg}_{t-1}, Z) + f(u_t, Z \mid \texttt{alg}_{t-1}) && \textit{(Since } \texttt{alg}_t = \texttt{alg}_{t-1} \cup \{ u_t \}\textit{)} \\ &= f(\texttt{alg}_{t-1}, Z) + f_t(u_t, Z) && \textit{(Definition of } f_t\textit{)} \\ &= f(\texttt{alg}_{t-1}, Z) + f_t(u_t, \texttt{zone}_t^{\max}) && \textit{(Since } Z = \texttt{zone}(u^*) \textit{ and } u^* = u^*_{\min}(u_t)\textit{)} \\ &\geq f(\texttt{alg}_{t-1}, Z) + 0.1\Delta^{\text{base}} && \textit{(Since } u_t \in \texttt{alg}^{\text{con}}\textit{)}. \end{aligned} \tag{10}$$

Subtracting the above Equations we obtain

$$\begin{aligned} \texttt{rem}_{t+1}(u^*) &= f(\texttt{opt} \cup \texttt{alg}_t, Z) - f(\texttt{alg}_t, Z) && \textit{(Equation (8))} \\ &\leq f(\texttt{opt} \cup \texttt{alg}_{t-1}, Z) + 0.01\Delta^{\text{base}} - f(\texttt{alg}_{t-1}, Z) - 0.1\Delta^{\text{base}} && \textit{(Equations (9) and (10))} \\ &= \texttt{rem}_t(u^*) - 0.09\Delta^{\text{base}} && \textit{(Definition of } \texttt{rem}_t\textit{)}, \end{aligned}$$

as desired. □

As an immediate corollary, we obtain a lower bound on the size of $|\, \texttt{opt}^{\text{con}} \,|$.

Lemma 15. *If we have* $|\, \texttt{alg}^{\text{con}} \,| \geq 0.3 \,|\, \texttt{alg}' \,|$*, then we have* $|\, \texttt{opt}^{\text{con}} \,| \geq \Omega(k)$.

Proof. We first give a sketch of the proof. We first claim that each $u^* \in \texttt{opt}$ can be $u^*_{\min}(u_t)$ for only a constant number of $u_t \in \texttt{alg}^{\text{con}}$. To prove this, we use Lemma 14 which implies that each time $u^* = u^*_{\min}(u_t)$, the value $\texttt{rem}_t(u^*)$ decreases by $0.09\Delta^{\text{base}}$. The initial value of $\texttt{rem}_t(u^*)$ for $t = 0.005k$ is at most Δ^{base} because of Equation (7). The final value is also at least 0 because f is monotone. Therefore, each u^* equals $u^*_{\min}(u_t)$ only a constant number of times as claimed. Since $|\, \texttt{alg}^{\text{con}} \,| \geq 0.3 \,|\, \texttt{alg}' \,| \geq \Omega(k)$, this proves the lemma.

We now proceed with a formal proof. We first show that for all $u^* \in \texttt{opt}$,

$$|\, \{\, u_t \in \texttt{alg}^{\text{con}} : u^* = u^*_{\min}(u_t) \,\} \,| \leq O(1), \tag{11}$$

Assume that $\{\, u_t \in \texttt{alg}^{\text{con}} : u^* = u^*_{\min}(u_t) \,\} = \{\, u_{t_1}, \ldots, u_{t_s} \,\}$ where $t_1 < t_2 < \cdots < t_s$. Since $\texttt{rem}_t$ is decreasing in t, for any $i \in [s-1]$ we have

$$\texttt{rem}_{t_i}(u^*) - \texttt{rem}_{t_{i+1}}(u^*) \geq \texttt{rem}_{t_i}(u^*) - \texttt{rem}_{t_i+1}(u^*) \geq 0.09\Delta^{\text{base}}.$$

Given Equation (7) however, $\texttt{rem}_{t_1}(u^*) \leq \Delta^{\text{base}}$. Therefore, since $\texttt{rem}_{t_s}(u^*) > 0$, we conclude that $s \leq 13 = O(1)$.

Given the above claim,

$$\begin{aligned} |\, \texttt{opt}^{\text{con}} \,| &= \sum_{u^* \in \texttt{opt}} \mathbb{1}\left\{\exists u_t \in \texttt{alg}^{\text{con}} : u^* = u^*_{\min}(u_t)\right\} && \textit{(Definition of } \texttt{opt}^{\text{con}}\textit{)} \\ &\geq \Omega(1) \sum_{u^* \in \texttt{opt}} |\, \{\, u_t \in \texttt{alg}^{\text{con}} : u^* = u^*_{\min}(u_t) \,\} \,| && \textit{(Equation (11))} \\ &= \Omega(1) \sum_{u^* \in \texttt{opt}} \sum_{u_t \in \texttt{alg}^{\text{con}}} \mathbb{1}\left\{u^* = u^*_{\min}(u_t)\right\} \\ &= \Omega(1) \sum_{u_t \in \texttt{alg}^{\text{con}}} \sum_{u^* \in \texttt{opt}} \mathbb{1}\left\{u^* = u^*_{\min}(u_t)\right\} && \textit{(Rearranging)} \\ &= \Omega(1) \sum_{u_t \in \texttt{alg}^{\text{con}}} 1 && \textit{(uniqueness of } u^*_{\min}(.)\textit{)} \\ &= \Omega(k) && \textit{(Assumption } |\, \texttt{alg}^{\text{con}} \,| \geq 0.3 \,|\, \texttt{alg}' \,|\textit{)}, \end{aligned}$$

as desired. □

Using Lemma 14, we additionally obtain the following bound on $\texttt{rem}_t(u^*)$ for $u^* \in \texttt{opt}^{\text{con}}$ and $t \geq 0.01k$.

Lemma 16. *For any* $u^* \in \texttt{opt}^{\text{con}}$ *and any* $t' \geq \lceil 0.01k \rceil$*, we have*

$$\texttt{rem}_{t'}(u^*) \leq 0.91\Delta^{\text{base}}.$$

Proof. Since $u^* \in \texttt{opt}^{\text{con}}$, there exists $u_t \in \texttt{alg}^{\text{con}}$ for which $u^* = u^*_{\min}(u_t)$. Given Equation (7), $\texttt{rem}_t(u^*) \leq \Delta^{\text{base}}$ and given Lemma 14 $\texttt{rem}_t(u^*) - \texttt{rem}_{t+1}(u^*) \geq 0.09\Delta^{\text{base}}$. It follows that $\texttt{rem}_{t+1}(u^*) \leq 0.91\Delta^{\text{base}}$ which proves the claim since $t' \geq t+1$ and $\texttt{rem}_t$ is decreasing in t. □

We now handle the case of $\texttt{rem}_{\lceil 0.02k \rceil} \leq 0.97\text{OPT}$ as mentioned in the sketch.

Lemma 17. *If at $t = \lceil 0.02k \rceil$ we have $\texttt{rem}_t \leq 0.97\text{OPT}$, then the approximation ratio is $\geq 1 - 1/e + \Omega(1)$.*

Proof. Since $\texttt{rem}_{\lceil 0.02k \rceil} \leq 0.97\text{OPT}$, we conclude that $f(\texttt{opt}) - f(\texttt{alg}_{\lceil 0.02k \rceil - 1}) \leq 0.97\text{OPT}$. Put $m = \lceil 0.02k \rceil - 1$. Bernoulli's inequality gives $(1-1/k)^m \geq 1 - m/k > 0.98$. Hence $\text{gap}_m \leq (1-1/k)^m\text{OPT} - 0.01\text{OPT}$, and Lemma 6 applies at time m. □

Given this claim, we can focus on the case that $\texttt{rem}_{\lceil 0.02k \rceil} \geq 0.97\text{OPT}$. We first show that there is a surplus in $\texttt{rem}_t(u^*)$ from $\texttt{opt} \setminus \texttt{opt}^{\text{con}}$.

Lemma 18. *If $\texttt{rem}_{\lceil 0.02k \rceil} \geq 0.97\text{OPT}$, then for all $t \in [0.01k, 0.02k)$,*

$$\sum_{u^* \in \texttt{opt} \setminus \texttt{opt}^{\text{con}}} \left(\texttt{rem}_t(u^*) - \texttt{rem}_t/k \right) \geq \Omega(k)\Delta^{\text{base}},$$

Proof. For all $t \in [0.01k, 0.02k)$,

$$0 = \sum_{u^*} \left(\texttt{rem}_t(u^*) - \texttt{rem}_t/k \right) \qquad \textit{(Since } \texttt{rem}_t = \sum \texttt{rem}_t(u^*)\textit{)}$$

$$= \sum_{u^* \in \texttt{opt}^{\text{con}}} \left(\texttt{rem}_t(u^*) - \texttt{rem}_t/k \right) + \sum_{u^* \in \texttt{opt} \setminus \texttt{opt}^{\text{con}}} \left(\texttt{rem}_t(u^*) - \texttt{rem}_t/k \right),$$

which implies

$$\sum_{u^* \in \texttt{opt} \setminus \texttt{opt}^{\text{con}}} \left(\texttt{rem}_t(u^*) - \texttt{rem}_t/k \right) = \sum_{u^* \in \texttt{opt}^{\text{con}}} \left(\texttt{rem}_t/k - \texttt{rem}_t(u^*) \right)$$

We know that $\texttt{rem}_t(u^*) \leq 0.91\Delta^{\text{base}}$ for $u^* \in \texttt{opt}^{\text{con}}$ from Lemma 16 and that $\texttt{rem}_t/k \geq \texttt{rem}_{\lceil 0.02k \rceil}/k \geq 0.97\Delta^{\text{base}}$ by assumption. Plugging in these bounds we obtain

$$\sum_{u^* \in \texttt{opt}^{\text{con}}} \left(\texttt{rem}_t/k - \texttt{rem}_t(u^*) \right) \geq |\, \texttt{opt}^{\text{con}} \,| \left(0.97\Delta^{\text{base}} - 0.91\Delta^{\text{base}} \right)$$

$$\geq \Omega(k) \cdot (0.97 - 0.91)\Delta^{\text{base}} \qquad \textit{(Lemma 15)},$$

which finishes the proof. □

The following lemma implies an improved lower bound on Δ_t compared to the bound obtained in the standard analysis of the greedy algorithm, which is $(f(\texttt{opt}) - f(\texttt{alg}_{t-1}))/k$.

Lemma 19. *If $\texttt{rem}_{\lceil 0.02k \rceil} \geq 0.97\text{OPT}$, then for all $t \in [0.01k, 0.02k)$,*

$$\Delta_t \geq \frac{(1 + \Omega(1))}{k} \left(f(\texttt{opt}) - f(\texttt{alg}_{t-1}) \right).$$

Proof. By the greedy choice of the algorithm, Δ_t is larger than the maximum value of $\texttt{rem}_t(u^*)$, which in turn is larger than the average value of $\texttt{rem}_t(u^*)$ over $\texttt{opt} \setminus \texttt{opt}^{\text{con}}$. Given Lemma 18, this is at least $\Omega(1)\Delta^{\text{base}}$ larger than $\texttt{rem}_t/k$. Given Equation (7) however, $\texttt{rem}_t/k \leq \Delta^{\text{base}}$ which implies $\Delta_t \geq (1 + \Omega(1))\texttt{rem}_t/k$. The claim now follows from the fact that $\texttt{rem}_t \geq f(\texttt{opt}) - f(\texttt{alg}_{t-1})$.

Formally,

$$\Delta_t \geq \max_{u^* \in \texttt{opt} \setminus \texttt{opt}^{\text{con}}} \texttt{rem}_t(u^*) \qquad \textit{(Greedy choice of algorithm)}$$

$$= \texttt{rem}_t/k + \max_{u^* \in \texttt{opt} \setminus \texttt{opt}^{\text{con}}} \left(\texttt{rem}_t(u^*) - \texttt{rem}_t/k \right)$$

$$\geq \texttt{rem}_t/k + \frac{1}{k} \sum_{u^* \in \texttt{opt} \setminus \texttt{opt}^{\text{con}}} \left(\texttt{rem}_t(u^*) - \texttt{rem}_t/k \right) \qquad \textit{(Since } \max \geq \text{average}\textit{)}$$

$$\geq \texttt{rem}_t/k + \Omega(1)\Delta^{\text{base}} \qquad \textit{(Lemma 18)}.$$

Given Equation (7) we have $\texttt{rem}_t/k \le \max_{u^*} \texttt{rem}_t(u^*) \le \Delta^{\text{base}}$. It follows that

$$\begin{aligned}
\Delta_t &\ge (1+\Omega(1))\texttt{rem}_t/k. \\
&= \frac{(1+\Omega(1))}{k}\left(f(\texttt{opt}\cup\texttt{alg}_{t-1}) - f(\texttt{alg}_{t-1})\right) && \textit{(Definition of } \texttt{rem}_t\textit{)} \\
&\ge \frac{(1+\Omega(1))}{k}\left(f(\texttt{opt}) - f(\texttt{alg}_{t-1})\right) && \textit{(Monotonicity of } f\textit{).}
\end{aligned}$$

□

Lemma 20. *If* $\texttt{rem}_{\lceil 0.02k\rceil} \ge 0.97\text{OPT}$*, then the approximation ratio is* $\ge 1-1/e+\Omega(1)$.

Proof. The main idea behind the proof is as follows. During the time intervals $[0, 0.01k)$ and $[0.02k, k]$, the decrease rate of gap_t is $(1-1/k)$ as proved in the standard submodular analysis. For $[0.01k, 0.02k)$ however, we have an improved decrease rate of $1-(1+\Omega(1))/k$ because of Lemma 19. Since the length of the interval $[0.01k, 0.02k]$ is $\Omega(k)$, this implies an $\Omega(1)$ improvement.

Formally, given Lemma 19, for $t \in [0.01k, 0.02k)$ we have

$$f(\texttt{alg}_t) - f(\texttt{alg}_{t-1}) = f(u_t \mid \texttt{alg}_{t-1}) = \Delta_t \ge \frac{(1+\Omega(1))}{k}\left(f(\texttt{opt}) - f(\texttt{alg}_{t-1})\right).$$

It follows that $f(\texttt{opt}) - f(\texttt{alg}_t) \le \left(1 - \frac{1+\Omega(1)}{k}\right)\left(f(\texttt{opt}) - f(\texttt{alg}_{t-1})\right)$. As before however, for all t we know that

$$f(\texttt{opt}) - f(\texttt{alg}_t) \le (1-1/k)\left(f(\texttt{opt}) - f(\texttt{alg}_{t-1})\right),$$

Therefore,

$$\begin{aligned}
&\frac{f(\texttt{opt}) - f(\texttt{alg}_k)}{f(\texttt{opt}) - f(\texttt{alg}_0)} \\
&= \prod_{t=1}^{\lceil 0.01k\rceil - 1} \frac{f(\texttt{opt}) - f(\texttt{alg}_t)}{f(\texttt{opt}) - f(\texttt{alg}_{t-1})} \prod_{t=\lceil 0.01k\rceil}^{\lceil 0.02k\rceil - 1} \frac{f(\texttt{opt}) - f(\texttt{alg}_t)}{f(\texttt{opt}) - f(\texttt{alg}_{t-1})} \prod_{t=\lceil 0.02k\rceil}^{k} \frac{f(\texttt{opt}) - f(\texttt{alg}_t)}{f(\texttt{opt}) - f(\texttt{alg}_{t-1})} \\
&\le (1/e)^{1+\Omega(1)\frac{\lceil 0.02k\rceil - \lceil 0.01k\rceil}{k}} \\
&\le (1/e)^{1+\Omega(1)} \\
&\le 1/e - \Omega(1).
\end{aligned}$$

Therefore, $f(\texttt{alg}_k) \ge \text{OPT}(1-1/e+\Omega(1))$, finishing the proof. □

Putting it all together Lemma 13 follows from either Lemma 17 or Lemma 20 depending on the value of $\texttt{rem}_{\lceil 0.02k\rceil}$.

A.3.3 Large number of triangle centers

In this section, we prove the following lemma which establishes that if $\left|\,\texttt{alg}^{\text{tri}}\,\right|$ is large, then the approximation ratio exceeds $1-1/e$.

Lemma 21. *If* $\left|\,\texttt{alg}^{\text{tri}}\,\right| \ge 0.3\,|\,\texttt{alg}'\,|$ *then the approximation ratio is* $\ge 1-1/e+\Omega(\eta^2)$.

Proof sketch Define $\overline{P}_t$ as

$$\overline{P}_t := \left\{\, p \in P_t^{\text{yo}} : f(u_t, p) \le 1.1 f(\texttt{opt}, p) \wedge p \notin \texttt{zone}_t^{\max} \,\right\}$$

and recall from Section 3 that $u^*_{\min}(p)$ denotes the closest center $u^* \in \texttt{opt}$ to p.

We divide the proof into four parts. The first three parts establish some tools, which the final part uses to prove Lemma 21.

In Part 1, we show that for triangle centers, the marginal gain from the points in $\overline{P}_t$ is already high; specifically, $f_t(u_t, \overline{P}_t) \ge 0.6\Delta^{\text{base}}$ (Lemma 23) for $u_t \in \texttt{alg}^{\text{tri}}$. We establish this by using the fact that $u_t \in \texttt{alg}^{\text{yo}}$ but $u_t \notin \texttt{alg}^{\text{ind}}$ to show that the marginal gain from points satisfying $f(u_t, p) > 1.1 f(\texttt{opt}, p)$ is at most $0.15\Delta^{\text{base}}$ (Lemma 22). This implies Lemma 23 given the lower

bound of $0.9\Delta^{\text{base}}$ on $f_t(u_t, P_t^{\text{yo}})$ and the upper bound of $0.1\Delta^{\text{base}}$ on $f_t(u_t, \texttt{zone}_t^{\max})$ which follows from $u_t \notin \texttt{alg}^{\text{con}}$.

In Part 2, we show that for any center u_t and any $p \in \overline{P}_t$, the center $u^*_{\min}(u_t) \in \texttt{opt}$ serves at least $\eta/1.1$ fraction of what u_t serves from p. Formally, $f(u^*_{\min}(u_t), p) \geq \frac{\eta}{1.1} f(u_t, p)$ (Lemma 25). To prove this, we use the triangle inequality together with the fact that $f(u_t, p) \leq 1.1 f(\texttt{opt}, p)$.

In Part 3, we show that if the centers in opt service a lot in each other's zones, then the algorithm obtains a better bound than $1 - 1/e$. To better explain this, let us focus on the time step $t = 0$ for simplicity. Recall that the standard submodular analysis relies on the inequality $\sum_{u^* \in \texttt{opt}} f(u^*) \geq f(\texttt{opt})$ (this is the Spread step in the proof). By definition of zones, however, it is easy to see that $f(\texttt{opt})$ can be decomposed as $\sum_{u^* \in \texttt{opt}} f(u^*, \texttt{zone}(u^*))$. Therefore, for this inequality to be tight, we need to have $f(u^*, \texttt{zone}(v^*)) = 0$ for all $u^* \neq v^*$. In other words, this bound does not account for the amount each center in opt may service other zones. Defining $\texttt{excess}_t$ to be a $1/k$ fraction of the sum of this "inter zone" services at time t (see Equation (16)), and at the end of the algorithm, we have basically collected $\sum \texttt{excess}_t$ on top of the standard bound of $(1 - 1/e)$OPT (Lemma 29).

With the above three parts established, the fourth part proves the main lemma as follows. Consider a bipartite graph where the vertices on the left are the triangle centers $\texttt{alg}^{\text{tri}}$ and the vertices in the right are all points in $\overline{P}_t$ for the vertices on the left, i.e., $\overline{P}_\cup := \cup_{u_t \in \texttt{alg}^{\text{tri}}} \overline{P}_t$. For any $u_t \in \texttt{alg}^{\text{tri}}$ and $p \in \overline{P}_t$, we connect u_t and p with an edge that has weight $f_t(u_t, p)$. Let $\overline{\texttt{alg}}(p) := \left\{ u_t \in \texttt{alg}^{\text{tri}} : p \in \overline{P}_t \right\}$ denote the set of neighbors for any $p \in \overline{P}_\cup$, and let $u^{\text{tri}}_{\min}(p) := \arg\min_{u \in \overline{\texttt{alg}}(p)} d(u, p)$ denote the neighbor with minimum distance. Note that each $p \in \overline{P}_\cup$ has at least one neighbor by definition of $\overline{P}_\cup$ and as such $u^{\text{tri}}_{\min}$ is always defined. We will upper and lower bound the sum of the edges in terms of $\sum \texttt{excess}_t$ and OPT respectively, allowing us to improve the $1 - 1/e$ bound using the analysis from Part 3.

For the upper bound, consider any vertex $p \in \overline{P}_\cup$ and let $u^{\text{tri}}_{\min}$ denote the closest center to p in $\overline{\texttt{alg}}(p)$. We will first show that the sum of the edges connected to p is at most $f(u^{\text{tri}}_{\min}, p)$ (Lemma 28). Intuitively, this holds because each edge represents an increase in how much p is served, while $f(u^{\text{tri}}_{\min}, p)$ represents the final amount of service. Choose t such that $u^{\text{tri}}_{\min} = u_t$ and set $u^* = u^*_{\min}(u^{\text{tri}}_{\min})$. Since there is an edge between $u^{\text{tri}}_{\min}$ and p, it must be the case that $p \in \overline{P}_t$, which in turn implies that p is not in $\texttt{zone}(u^*)$. Given Part 2, $f(u^*, p)$ is at least $\eta/1.1$ of $f(u^{\text{tri}}_{\min}, p)$. Since $p \in P_t^{\text{yo}}$ however, at time ϵk we have $f(\texttt{alg}_{\lceil \epsilon k \rceil - 1}, p) \leq \eta/2 f(u^{\text{tri}}_{\min}, p)$. Therefore, for any $t' \in [1, \epsilon k)$ we have $f_{t'}(u^*, p) \geq \eta(1/1.1 - 1/2) f(u^{\text{tri}}_{\min}, p)$, which means p contributes $\Omega(\eta f(u^{\text{tri}}_{\min}, p)/k)$ to each $\texttt{excess}_{t'}$. Since the interval has size $\Omega(\eta k)$, the point p contributes $\Omega(\eta^2 f(u^{\text{tri}}_{\min}, p))$ to the $\sum \texttt{excess}_{t'}$ (Lemma 27). In other words, the contribution of p is at least an $\Omega(\eta^2)$ fraction of the sum of the edges connected to it. Summing over p, we conclude that the sum of the edges in the graph is at most $O\left(\eta^{-2}\right)\left(\sum_{t'} \texttt{excess}_{t'}\right)$.

For the lower bound, we note that the sum of edges connected to each u_t is by definition $f_t(u_t, \overline{P}_t)$, which is at least $0.6\Delta^{\text{base}}$ given Part 1. Therefore, since $\left|\texttt{alg}^{\text{tri}}\right|$ is $\Omega(k)$ by assumption, the sum of all of the edges in the graph is $\Omega(k\Delta^{\text{base}}) = \Omega(\text{OPT})$. Given the upper and lower bounds on the sum of the edges, we conclude that $\sum_{t'} \texttt{excess}_{t'} \geq \Omega(\eta^2 \text{OPT})$ (Lemma 29), which in turn implies that the approximation ratio is $1 - 1/e + \Omega(\eta^2)$. The remainder of this section formalizes this sketch.

Part 1 The next two lemmas form one step: bound the gain from points where greedy serves substantially more than opt, then subtract this gain and the closest optimal zone's contribution to lower-bound the gain on $\overline{P}_t$.

Lemma 22. *If* $u_t \in \texttt{alg}^{\text{yo}} \backslash \texttt{alg}^{\text{ind}}$ *then*

$$\sum_{p \in P} f_t(u_t, p) \mathbb{1}\left\{ f(u_t, p) \geq 1.1 f(\texttt{opt}, p) \right\} < 0.15\Delta^{\text{base}}.$$

Proof. The main intuition behind the proof is as follows. For any point p such that $f(u_t, p) \geq 1.1 f(\texttt{opt}, p)$, the amount u_t serves that is beyond opt is substantial compared to the amount it serves overall; specifically, it is at least $0.1/1.1 \approx 1/10$ of the amount served. Therefore, if $0.15\Delta^{\text{base}}$ marginal gain of u_t comes from these points, then $\approx 0.015\Delta^{\text{base}}$ marginal gain is not served by opt, which contradicts the assumption that $u_t \notin \texttt{alg}^{\text{ind}}$.

Formally, assume for contradiction that

$$\sum_{p \in P} f_t(u_t, p) \mathbb{1}\left\{ f(u_t, p) \geq 1.1 f(\texttt{opt}, p) \right\} \geq 0.15 \Delta^{\text{base}}. \tag{12}$$

Define $P_t^{\text{ind}} \subseteq P$ as

$$P_t^{\text{ind}} := \left\{ p \in P : f(u_t, p) \geq 1.1 f(\texttt{opt}, p) \right\}.$$

We claim that

$$\forall p \in P_t^{\text{ind}} : f(u_t, p \mid \texttt{opt} \cup \texttt{alg}_{t-1}) \geq \frac{1}{11} f_t(u_t, p). \tag{13}$$

If $f(\texttt{opt}, p) \leq f(\texttt{alg}_{t-1}, p)$, then

$$\begin{aligned}
&f(u_t, p \mid \texttt{opt} \cup \texttt{alg}_{t-1}) \\
&= f(\texttt{opt} \cup \texttt{alg}_{t-1} \cup \{ u_t \}, p) - f(\texttt{opt} \cup \texttt{alg}_{t-1}, p) && \textit{(Definition of } f(. \mid .)\textit{)} \\
&= f(\texttt{alg}_{t-1} \cup \{ u_t \}, p) - f(\texttt{alg}_{t-1}, p) && \textit{(Since } f(\texttt{opt}, p) \leq f(\texttt{alg}_{t-1}, p)\textit{)} \\
&= f(u_t, p \mid \texttt{alg}_{t-1}) && \textit{(Definition of } f(. \mid .)\textit{)} \\
&= f_t(u_t, p) && \textit{(Definition of } f_t\textit{).}
\end{aligned}$$

If $f(\texttt{opt}, p) \geq f(\texttt{alg}_{t-1}, p)$, then

$$\begin{aligned}
&f(u_t, p \mid \texttt{opt} \cup \texttt{alg}_{t-1}) \\
&= f(\texttt{opt} \cup \texttt{alg}_{t-1} \cup \{ u_t \}, p) - f(\texttt{opt} \cup \texttt{alg}_{t-1}, p) && \textit{(Definition of } f(. \mid .)\textit{)} \\
&= f(\texttt{opt} \cup \{ u_t \}, p) - f(\texttt{opt}, p) && \textit{(Since } f(\texttt{alg}_{t-1}, p) \leq f(\texttt{opt}, p)\textit{)} \\
&\geq f(u_t, p) - f(\texttt{opt}, p) \\
&\geq \frac{1}{11} f(u_t, p) && \textit{(Since } p \in P_t^{\text{ind}}\textit{)} \\
&\geq \frac{1}{11} \left(f(u_t, p) - f(\texttt{alg}_{t-1}, p) \right) \\
&= \frac{1}{11} f_t(u_t, p),
\end{aligned}$$

which proves Equation (13) in this case as well.

Summing over all p we obtain

$$\begin{aligned}
f(u_t \mid \texttt{opt} \cup \texttt{alg}_{t-1}) &= \sum_{p \in P} f(u_t, p \mid \texttt{opt} \cup \texttt{alg}_{t-1}) && \textit{(Definition of } f(.)\textit{)} \\
&\geq \sum_{p \in P_t^{\text{ind}}} f(u_t, p \mid \texttt{opt} \cup \texttt{alg}_{t-1}) && \textit{(Monotonicity of } f\textit{)} \\
&\geq \frac{1}{11} \sum_{p \in P_t^{\text{ind}}} f_t(u_t, p) && \textit{(Equation (13))} \\
&\geq \frac{1}{11} \cdot 0.15 \Delta^{\text{base}} && \textit{(Equation (12))} \\
&\geq 0.01 \Delta^{\text{base}},
\end{aligned}$$

contradicting the assumption $u_t \in \texttt{alg}^{\text{yo}} \backslash \texttt{alg}^{\text{ind}}$. □

We now use the above lemma to obtain a lower bound on the marginal gain from points $\overline{P}_t$.

Lemma 23. *$f_t(u_t, \overline{P}_t) \geq 0.65 \Delta^{\text{base}}$ for any $u_t \in \texttt{alg}^{\text{tri}}$.*

Proof. For any $u_t \in \texttt{alg}^{\text{tri}}$,

$$\begin{aligned}
f_t(u_t, \overline{P}_t) &\geq f_t(u_t, P_t^{\text{yo}}) - f_t\left(u_t, \left\{ p \in P_t^{\text{yo}} : f(u_t, p) \geq 1.1 f(\texttt{opt}, p) \right\} \right) - f_t(u_t, \texttt{zone}_t^{\max}) \\
&\geq 0.9 \Delta^{\text{base}} - 0.15 \Delta^{\text{base}} - 0.1 \Delta^{\text{base}} \\
&= 0.65 \Delta^{\text{base}},
\end{aligned}$$

where in the second inequality we have used the fact that $u_t \in \texttt{alg}^{\text{yo}}$ and Lemma 22 for bounding the first and second terms respectively, and we have used the fact that $u_t \notin \texttt{alg}^{\text{con}}$ for bounding the third term. □

Part 2 We first prove the following lemma that encapsulates the main property we want from the triangle inequality.

Lemma 24. *Let u^*, v^* be two centers in* opt *and assume that $d(u,u^*) \le d(u,v^*)$ for some center u. For any point p in* $\texttt{zone}(v^*)$*,*

$$d(u^*,p) \le 3\max\{\, d(v^*,p), d(u,p)\,\}. \tag{14}$$

Proof. Set $x := \max\{\, d(v^*,p), d(u,p)\,\}$. By the triangle inequality on u,p,v^*, we have $d(u,v^*) \le 2x$. Since u^* was closer than v^* to u, this further means that $d(u,u^*) \le 2x$. Invoking the triangle inequality on u^*,u,p finishes the proof. We refer to Figure 1 for an illustration. □

Lemma 25. *Let u^*, v^* be two centers in* opt *and assume that $d(u,u^*) \le d(u,v^*)$ for some center u. For any point p in* $\texttt{zone}(v^*)$*,*

$$f(u^*,p) \ge \eta \min\{\, f(v^*,p), f(u,p)\,\}. \tag{15}$$

Additionally, any u_t and $p \in \overline{P}_t$ satisfy $f(u^*_{\min}(u_t),p) \ge \frac{\eta}{1.1} f(u_t,p)$.

Proof. Equation (15) follows from Lemma 24 together with the fact $\phi(3x) \ge \eta\phi(x)$ for all x by definition of η. Formally,

$$\begin{aligned}
f(u^*,p) &= \phi(d(u^*,p)) && \textit{(Definition of } f(.,.)\textit{)}\\
&\ge \phi(3\max\{\, d(v^*,p), d(u,p)\,\}) && \textit{(Lemma 24)}\\
&\ge \eta\phi(\max\{\, d(v^*,p), d(u,p)\,\}) && \textit{(Definition of } \eta\textit{)}\\
&= \eta\min\{\, \phi(d(v^*,p)), \phi(d(u,p))\,\} && \textit{(Since } \phi \textit{ is non-increasing)}\\
&= \eta\min\{\, f(v^*,p), f(u,p)\,\} && \textit{(Definition of } f(.,.)\textit{)}
\end{aligned}$$

For the second part, set $u = u_t$, $u^* = u^*_{\min}(u_t)$ and $v^* = u^*_{\min}(p)$. Since $p \in \overline{P}_t$, we have $u^* \ne v^*$. Since $p \in \overline{P}_t$, we also have $f(\texttt{opt},p) \ge 1/1.1 f(u_t,p)$. Therefore,

$$\begin{aligned}
f(u^*,p) &\ge \eta\min\{\, f(u_t,p), f(u^*_{\min}(p),p)\,\} && \textit{(Equation (15))}\\
&= \eta\min\{\, f(u_t,p), f(\texttt{opt},p)\,\} && \textit{(Definition of } u^*_{\min}\textit{)}\\
&\ge \frac{\eta}{1.1} f(u_t,p) && \textit{(Since } f(u_t,p) \le 1.1 f(\texttt{opt},p)\textit{)},
\end{aligned}$$

as desired. □

Part 3 Define $\texttt{excess}_t$ as

$$\texttt{excess}_t := \frac{1}{k}\left(\sum_{u^* \ne v^* \in \texttt{opt}} f_t(u^*, \texttt{zone}(v^*))\right). \tag{16}$$

The following lemma formalizes the sketch from the beginning of the section.

Lemma 26. *For any $t' \in [k]$,*

$$f(\texttt{opt}) - f(\texttt{alg}_{t'}) \le (1-1/k)^{t'} f(\texttt{opt}) - 1/4 \sum_{t \le t'} \texttt{excess}_t$$

Proof. The main intuition behind the proof is as follows. Since the algorithm chooses its centers greedily, the marginal gain Δ_t is at least $\max_{u^* \in \texttt{opt}} f_t(u^*)$. In the standard analysis of submodular functions, this would further be lower bounded by $\frac{1}{k} f_t(\texttt{opt})$. However, this bound is only tight if $f_t(\texttt{opt}) = \sum_{u^* \in \texttt{opt}} f_t(u^*)$, which only happens if each u^* only serves points in its own zone. Given the definition of $\texttt{excess}_t$, we can actually refine this bound to obtain $f_t(u_t) \ge 1/k f_t(\texttt{opt}) + \texttt{excess}_t$ (see Equation (17)). This in turn implies that for any t, not only does the gap $f(\texttt{opt}) - f(\texttt{alg}_t)$ decrease by a factor of $1-1/k$, but it also has an extra decrease of $\texttt{excess}_t$ (see Equation (19)). Intuitively, this means we should be able to collect the extra $\sum \texttt{excess}_t$ at the end of the algorithm as an extra served amount. The extra contribution from iteration t is discounted by $(1-1/k)^{t'-t}$ in the bound at a later time t'. Since $(1-1/k)^k \approx 1/e$ however, this only affects this extra amount by a

constant factor. In the lemma statement, we use $1/4$ instead of $1/e$ to ensure the result holds for all $k \geq 2$.

At any time t,

$$\begin{aligned}
f_t(u_t) &\geq \max_{u^* \in \texttt{opt}} f_t(u^*) && \textit{(Greedy choice of algorithm)} \\
&= \max_{u^* \in \texttt{opt}} \left(f_t(u^*, \texttt{zone}(u^*)) + \sum_{v^* \in \texttt{opt} \setminus \{ u^* \}} f_t(u^*, \texttt{zone}(v^*)) \right) && \textit{(Since } \texttt{zone}(.) \textit{ partitions } P\textit{)} \\
&\geq \frac{1}{k} \left(\sum_{u^* \in \texttt{opt}} f_t(u^*, \texttt{zone}(u^*)) + \sum_{u^* \neq v^* \in \texttt{opt}} f_t(u^*, \texttt{zone}(v^*)) \right) && \textit{(Since } \max \geq \text{average }\textit{)} \\
&= \frac{1}{k} f_t(\texttt{opt}) + \frac{1}{k} \left(\sum_{u^* \neq v^* \in \texttt{opt}} f_t(u^*, \texttt{zone}(v^*)) \right) && \textit{(Definition of } \texttt{zone}\textit{)},
\end{aligned}$$

where we note (to avoid confusion) that the sum $\sum_{v^* \in \texttt{opt} \setminus \{ u^* \}}$ in the second line is only over v^* (and thus contains $k-1$ terms), while the sum $\sum_{v^* \neq u^*}$ in the third line is over both u^* and v^* (and thus contains $k(k-1)$ terms).

Plugging in definition of $\texttt{excess}_t$, we can rewrite this as

$$f_t(u_t) \geq \frac{1}{k} f_t(\texttt{opt}) + \texttt{excess}_t, \tag{17}$$

By monotonicity of f however, we can further bound the first term on the right hand side as

$$\frac{1}{k} f_t(\texttt{opt}) = \frac{1}{k} \left(f(\texttt{opt} \cup \texttt{alg}_{t-1}) - f(\texttt{alg}_{t-1}) \right) \geq \frac{1}{k} \left(f(\texttt{opt}) - f(\texttt{alg}_{t-1}) \right). \tag{18}$$

We can therefore bound the decrease in $f(\texttt{opt}) - f(\texttt{alg}_t)$ in each step as

$$\begin{aligned}
f(\texttt{opt}) - f(\texttt{alg}_t) &= f(\texttt{opt}) - f(\texttt{alg}_{t-1}) - f_t(u_t) && \textit{(Definition of } f_t\textit{)} \\
&\leq (1 - 1/k) \left(f(\texttt{opt}) - f(\texttt{alg}_{t-1}) \right) - \texttt{excess}_t && \textit{(Equations (17) and (18))}
\end{aligned} \tag{19}$$

We claim using induction that

$$f(\texttt{opt}) - f(\texttt{alg}_{t'}) \leq (1 - 1/k)^{t'} f(\texttt{opt}) - \sum_{t=1}^{t'} (1 - 1/k)^{t'-t} \texttt{excess}_t.$$

For $t' = 0$, the claim holds trivially. Assuming it holds for $t' - 1$,

$$\begin{aligned}
&f(\texttt{opt}) - f(\texttt{alg}_{t'}) \\
&\leq (1 - 1/k) \left(f(\texttt{opt}) - f(\texttt{alg}_{t'-1}) \right) - \texttt{excess}_{t'} && \textit{(Equation (19))} \\
&\leq (1 - 1/k) \left((1 - 1/k)^{t'-1} f(\texttt{opt}) - \sum_{t=1}^{t'-1} (1 - 1/k)^{t'-t-1} \texttt{excess}_t \right) - \texttt{excess}_{t'} && \textit{(Induction hypothesis)} \\
&\leq (1 - 1/k)^{t'} f(\texttt{opt}) - \sum_{t=1}^{t'} (1 - 1/k)^{t'-t} \texttt{excess}_t
\end{aligned}$$

The lemma now follows from $\texttt{excess}_t \geq 0$ and the fact that $(1 - 1/k)^{t'} \geq (1 - 1/k)^k \geq 1/4$. $\square$

Part 4 We proceed to lower bound $\sum \texttt{excess}_t$. For each p, define

$$\texttt{excess}_t(p) := 1/k \sum_{u^* \neq u^*_{\min}(p)} f_t(u^*, p).$$

It is clear that $\texttt{excess}_t = \sum_p \texttt{excess}_t(p)$. Recall that $u^{\text{tri}}_{\min}(p) = \arg\min_{u \in \overline{\texttt{alg}}(p)} d(u,p)$ denotes the center with minimum distance to p. We first use Lemma 25 to show that during the time $[0, \epsilon k]$, the algorithm served an extra $\Omega(\eta^2 f(u^{\text{tri}}_{\min}(p), p))$ compared to the bound obtained using standard submodular analysis. We then use double counting to show that $\sum_p f(u^{\text{tri}}_{\min}(p), p)$ is $\Omega(\sum_{u_t \in \texttt{alg}^{\text{tri}}} f_t(u_t, \overline{P}_t)) = \Omega(\text{OPT})$.

Lemma 27. *For any $p \in \overline{P}_\cup$*

$$\sum_{t'=1}^{\lceil \epsilon k \rceil - 1} \texttt{excess}_{t'}(p) \geq \Omega\left(\eta^2 f(u^{\text{tri}}_{\min}(p), p) \right).$$

Proof. Choose $t \in \mathcal{I}$ such that $u_t = u^{\text{tri}}_{\min}(p)$ and set $u^* = u^*_{\min}(u_t)$. By definition of u_t, we have $p \in \overline{P}_t$, which implies $p \notin \texttt{zone}^{\max}_t$ and $f(u_t, p) \leq 1.1 f(\texttt{opt}, p)$. We note that $p \notin \texttt{zone}^{\max}_t$ can be rewritten as $u^* \neq u^*_{\min}(p)$.

The main idea behind the proof is as follows. Lemma 25 implies that $f(u^*, p) \geq \frac{\eta}{1.1} f(u_t, p)$. In other words, the amount u^* serves p is a sizeable chunk of the amount u_t serves p. Since $p \in P^{\text{yo}}_t$ however, at most $\eta/2 f(u_t, p)$ is served by time ϵk. This means that for all $t' \in [1, \epsilon k)$, $f_{t'}(u^*, p) \geq (\eta/1.1 - \eta/2) f(u_t, p)$. Since $u^* \neq u^*_{\min}(p)$, this means that for every integer $t' \in [1, \epsilon k)$, an $\Omega(\eta/k)$ fraction of $f(u_t, p)$ appears in $\texttt{excess}_{t'}(p)$. Summing over t finishes the proof.

We now proceed with the formal proof. The assumption $p \in \overline{P}_t$ implies

$$f(\texttt{alg}_{\lceil \epsilon k \rceil - 1}, p) \leq \frac{\eta}{2} f(u_t, p) \quad \text{and} \quad f(u^*, p) \geq \frac{\eta}{1.1} f(u_t, p), \tag{20}$$

where the first inequality follows from $p \in P^{\text{yo}}_t$ and the second inequality follows from Lemma 25. Therefore, for any $t' \in [1, \epsilon k)$

$$\begin{aligned}
\texttt{excess}_{t'}(p) &= \frac{1}{k} \sum_{v^* \neq u^*_{\min}(p)} f_{t'}(v^*, p) && \textit{(Definition of } \texttt{excess}_{t'}(p)\textit{)} \\
&\geq \frac{1}{k} f_{t'}(u^*, p) && \textit{(Since } u^* \neq u^*_{\min}(p)\textit{)} \\
&= \frac{1}{k} \left(f(u^*, p) - f(\texttt{alg}_{t'-1}, p) \right) && \textit{(Definition of } f_{t'}\textit{)} \\
&\geq \frac{1}{k} \left(f(u^*, p) - f(\texttt{alg}_{\lceil \epsilon k \rceil - 1}, p) \right) && \textit{(Monotonicity of } f \textit{ and } t' \leq \epsilon k\textit{)} \\
&\geq \frac{\eta/1.1 - \eta/2}{k} f(u_t, p) && \textit{(Equation (20)).}
\end{aligned}$$

Since $\lceil \epsilon k \rceil - 1 \geq \epsilon k/2$ and $\epsilon = \Omega(\eta)$, summing over t' we obtain $\sum_{t'=1}^{\lceil \epsilon k \rceil - 1} \texttt{excess}_{t'}(p) \geq \Omega(\eta^2 f(u_t, p))$, which finishes the proof since $u_t = u^{\text{tri}}_{\min}(p)$. □

We now prove that $f(u^{\text{tri}}_{\min}(p), p)$ is at least the sum of edges connected to p in the bipartite graph described in the proof sketch.

Lemma 28. *For any $p \in \overline{P}_\cup$,*

$$f(u^{\text{tri}}_{\min}(p), p) \geq \sum_{u_t \in \overline{\texttt{alg}}(p)} f_t(u_t, p).$$

Proof. The main idea behind the proof is as follows. Each center $u_t \in \overline{\texttt{alg}}(p)$ partially serves p, with any point satisfying $f_t(u_t, p) > 0$ increasing this partial amount by some strictly positive value. The lemma simply states that the amount served at the end (i.e., $f(u^{\text{tri}}_{\min}(p), p)$) is at least the sum of the increases in partial service over a subset of t (i.e., t satisfying $u_t \in \overline{\texttt{alg}}(p)$).

Formally, let $\{ u_{\tau_1}, \ldots, u_{\tau_s} \}$ denote the set $\left\{ u_t \in \overline{\texttt{alg}}(p) : f_t(u_t, p) > 0 \right\}$, where $\tau_1 < \tau_2 < \cdots < \tau_s$. If $s = 0$, the right-hand side is zero and the claim follows from nonnegativity; assume henceforth that $s > 0$. Since $f_{\tau_i}(u_{\tau_i}, p) > 0$, we must have $f(u_{\tau_1}, p) < f(u_{\tau_2}, p) < \cdots < f(u_{\tau_s}, p)$. Define

$\tau_0 = 0$ and let $f(u_0, p)$ denote 0 (we define the term for simplifying the algebra and do not actually need u_0). It follows that

$$\begin{aligned}
f(u^{\text{tri}}_{\min}(p), p) &\geq f(u_{\tau_s}, p) && \textit{(Definition of } u^{\text{tri}}_{\min}(.)\textit{)} \\
&= \sum_{i=1}^{s} \big(f(u_{\tau_i}, p) - f(u_{\tau_{i-1}}, p) \big) && \textit{(Telescoping sum)} \\
&\geq \sum_{i=1}^{s} \big(f(u_{\tau_i}, p) - f(\texttt{alg}_{\tau_i - 1}, p) \big) && \textit{(Since } u_{\tau_{i-1}} \in \texttt{alg}_{\tau_i - 1}\textit{)} \\
&= \sum_{i=1}^{s} \big(f(\texttt{alg}_{\tau_i}, p) - f(\texttt{alg}_{\tau_i - 1}, p) \big) && \textit{(Since } f_{\tau_i}(u_{\tau_i}, p) > 0\textit{)} \\
&= \sum_{i=1}^{s} f_{\tau_i}(u_{\tau_i}, p) && \textit{(Definition of } f_{\tau_i}(.)\textit{)} \\
&= \sum_{u_t \in \overline{\texttt{alg}}(p)} f_t(u_t, p). && \square
\end{aligned}$$

Combining Lemmas 27 and 28 and summing over all p we obtain the following lemma.

Lemma 29. *If* $\left| \texttt{alg}^{\text{tri}} \right| \geq 0.3 \left| \texttt{alg}' \right|$*, then*

$$\sum_{t=1}^{\lceil \epsilon k \rceil - 1} \texttt{excess}_t \geq \Omega(\eta^2 \text{OPT}).$$

Proof. By Lemma 27, any point $p \in \overline{P}_{\cup}$ contributes $\Omega(\eta^2) f(u^{\text{tri}}_{\min}(p), p)$ to the above sum. By Lemma 28, this is at least $\Omega(\eta^2) \sum_{u_t \in \overline{\texttt{alg}}(p)} f_t(u_t, p)$. Therefore, if we divide this sum among all $u_t \in \overline{\texttt{alg}}(p)$ proportional to the value $f_t(u_t, p)$, each u_t receives $\Omega(\eta^2) f_t(u_t, p)$. Since this holds for all p satisfying $u_t \in \overline{\texttt{alg}}(p)$ and since $u_t \in \overline{\texttt{alg}}(p)$ is equivalent to $p \in \overline{P}_t$, it follows that each $u_t \in \texttt{alg}^{\text{tri}}$ receives $\Omega(\eta^2) \sum_{p \in \overline{P}_t} f_t(u_t, p)$, which is at least $\Omega(\eta^2) \cdot 0.65 \Delta^{\text{base}}$ given Lemma 23. Since this holds for all $u_t \in \texttt{alg}^{\text{tri}}$ and $\left| \texttt{alg}^{\text{tri}} \right| \geq 0.3 \left| \texttt{alg}' \right| \geq 0.3 \cdot 0.004k = \Omega(k)$, the claim follows.

Formally,

$$\begin{aligned}
\sum_{t=1}^{\lceil \epsilon k \rceil - 1} \texttt{excess}_t &= \sum_{t=1}^{\lceil \epsilon k \rceil - 1} \sum_{p} \texttt{excess}_t(p) \\
&\geq \Omega(\eta^2) \sum_{p \in \overline{P}_{\cup}} f(u^{\text{tri}}_{\min}(p), p) && \textit{(Lemma 27)} \\
&\geq \Omega(\eta^2) \sum_{p \in \overline{P}_{\cup}} \sum_{u_t \in \overline{\texttt{alg}}(p)} f_t(u_t, p) && \textit{(Lemma 28)} \\
&\geq \Omega(\eta^2) \sum_{u_t \in \texttt{alg}^{\text{tri}}} \sum_{p \in \overline{P}_t} f_t(u_t, p) && \textit{(Double counting)} \\
&\geq \Omega(\eta^2) \left| \texttt{alg}^{\text{tri}} \right| (0.65 \Delta^{\text{base}}) && \textit{(Lemma 23)} \\
&= \Omega(\eta^2) \cdot 0.3k \cdot 0.65 \Delta^{\text{base}} && \textit{(Since } \left| \texttt{alg}^{\text{tri}} \right| \geq \Omega(k)\textit{)} \\
&= \Omega(\eta^2 k \Delta^{\text{base}}),
\end{aligned}$$

which is $\Omega(\eta^2 \text{OPT})$ since $\Delta^{\text{base}} = \text{OPT}/k$. $\square$

Putting everything together Combining the analysis from the above parts we obtain the proof of Lemma 21.

Proof of Lemma 21. Invoking Lemma 26 with $t' = k$ we obtain

$$f(\texttt{alg}_k) \geq (1-1/e)\text{OPT} + 1/4 \sum_{t=1}^{k} \texttt{excess}_t \geq (1-1/e)\text{OPT} + 1/4 \sum_{t=1}^{\lceil \epsilon k \rceil - 1} \texttt{excess}_t,$$

where the second inequality uses the non-negativity of $\texttt{excess}_t$. Invoking Lemma 29 to bound $\sum_{t=1}^{\lceil \epsilon k \rceil - 1} \texttt{excess}_t$ finishes the proof. □

A.4 Putting everything together

Putting together Lemmas 10, 12, 21, 13 we obtain Theorem 2.

Proof of Theorem 2. If $\left| \texttt{alg}^{\text{old}} \right| \geq 0.1 \left| \texttt{alg}' \right|$, then the theorem follows from Lemma 10. Otherwise, we have $\left| \texttt{alg}^{\text{yo}} \right| \geq 0.9 \left| \texttt{alg}' \right|$. At least one of the three sets $\texttt{alg}^{\text{ind}}, \texttt{alg}^{\text{tri}}, \texttt{alg}^{\text{con}}$ must have size at least $\frac{1}{3} \left| \texttt{alg}^{\text{yo}} \right| \geq 0.3 \left| \texttt{alg}' \right|$ because $\texttt{alg}^{\text{tri}}$ was defined as $\texttt{alg}^{\text{yo}} \setminus (\texttt{alg}^{\text{ind}} \cup \texttt{alg}^{\text{con}})$. If it is $\texttt{alg}^{\text{ind}}$, the theorem follows from Lemma 12. If it is $\texttt{alg}^{\text{tri}}$, the claim follows from Lemma 21. Finally, if it is $\texttt{alg}^{\text{con}}$, the claim follows from Lemma 13. □

B Asymptotically tight hardness

In this section, we provide a hardness of approximation result for the submodular clustering problem, where the upper bound on the approximation ratio depends on $\gamma_{\phi,\mathbb{Q}}$, defined below. The hardness result simultaneously shows that the problem does not admit a PTAS when $\gamma_\phi < 1$ and that the $1 - 1/e$ approximation ratio cannot be improved when $\gamma_\phi = 0$. Recall the rational-distance parameter

$$\gamma_{\phi,\mathbb{Q}} := \inf_{x \in \mathbb{Q}_{>0}} \frac{\phi(2x)}{\phi(x)},$$

with the same convention $0/0 = 1$ as for γ_ϕ. We have

$$\gamma_\phi \leq \gamma_{\phi,\mathbb{Q}} \leq \sqrt{\gamma_\phi}. \tag{21}$$

The first inequality follows by restricting the infimum to rational arguments. For the second, put $b = \gamma_{\phi,\mathbb{Q}}$ and, for any real $x > 0$, choose a rational $q \in (x/2, x)$. Monotonicity and the definition of b give

$$\phi(2x) \geq \phi(4q) \geq b\phi(2q) \geq b^2\phi(q) \geq b^2\phi(x).$$

Taking the infimum over x, with the stated convention when $\phi(x) = 0$, gives $b^2 \leq \gamma_\phi$; the ratio at $x = 0$ is 1 and does not affect this inequality. Thus $\gamma_{\phi,\mathbb{Q}} = 0$ if and only if $\gamma_\phi = 0$, and $\gamma_{\phi,\mathbb{Q}} < 1$ if and only if $\gamma_\phi < 1$. Theorem 3 therefore implies the stated hardness bound. We now prove it, fixing ϕ and ϵ as in its statement. Our proof is based on a reduction from the hardness of approximating Max k-Cover beyond $1 - 1/e$, even when the optimum is assumed to cover the entirety of the ground set. Specifically, we use the following hardness result by [KPV13].

Theorem 30. *For every fixed $0 < \epsilon < 1/e$, unless* P = NP*, no polynomial-time algorithm can do the following on every input bipartite graph $G = (L, R, E)$ and budget k: given the promise that some $S^* \subseteq L$ with $|S^*| \leq k$ satisfies $N(S^*) = R$, output $S \subseteq L$ with $|S| \leq k$ and $|N(S)|/|R| > 1 - 1/e + \epsilon$. Here $N(S)$ is the neighbor set of S, and k is part of the input.*

This formulation follows from the Max k-cover gap-hardness result stated in Section 2 of [KPV13], by taking its error parameter sufficiently small and k sufficiently large.

Given the above result, we formalize our hardness as follows. Write $\gamma = \gamma_{\phi,\mathbb{Q}}$ and $\varepsilon = \epsilon$. Set $\zeta = 1 - 1/e + \varepsilon/4$ and $\xi = 1 - 1/e + \frac{\gamma}{e} + \varepsilon$. Set a positive rational x with $\phi(x) > 0$ such that $\frac{\phi(2x)}{\phi(x)} \leq \gamma + \varepsilon/4$; such a rational x exists by the definition of $\gamma_{\phi,\mathbb{Q}}$, since $\gamma + \varepsilon/4 < 1$. Assume that there exists an algorithm $\mathcal{A}$ that solves the submodular clustering problem with approximation ratio ξ. For any instance of the bipartite graph above, we reduce it to a submodular clustering instance and show that obtaining an approximation ratio ξ in the reduced instance would imply an approximation ratio better than ζ in the original instance, which is not possible unless P = NP.

The reduction is as follows. We first copy all of the vertices on the right hand side $C \left\lceil \frac{\max\{k, |L|\}}{|R|} \right\rceil$ times where C is set to a large integer constant; specifically, $C \geq \frac{4}{\varepsilon}\left(1 + \frac{\phi(0)}{\phi(x)}\right)$. Let $G' = (L, R', E')$ be the graph. Let $M' = L \cup R'$ with $d(v, v) = 0$, $d(v, w) = x$ for edges of G', and $d(v, w) = 2x$ for every other pair of distinct vertices. Since all positive distances lie in $[x, 2x]$, the triangle inequality holds, regardless of whether G' is connected. Set $P = \mathcal{U} = M'$. By construction, $|R'| \geq C \max\{k, |L|\}$. It is easy to check that this representation of M' (given as a weighted complete graph) is still polynomial in the size of G. We give the metric M' as input to $\mathcal{A}$ and denote its output with $\texttt{alg}$. Set $\texttt{alg}_L = \texttt{alg} \cap L$ to be the vertices of this $\texttt{alg}$ that are in L. Since ϕ and ε are fixed, x and C are constants, so this is a polynomial-time procedure for producing $\texttt{alg}_L$. By Theorem 30, unless P = NP, there must therefore exist a promised, fully coverable instance on which $|N_G(\texttt{alg}_L)|/|R| \leq \zeta$. Fix such an instance, where $N_G(S)$ denotes the neighbor set in G. By construction of G' this further implies

$$\frac{|N_{G'}(\texttt{alg}_L)|}{|R'|} \leq \zeta. \tag{22}$$

We use this result to upper bound the approximation ratio of $\texttt{alg}$ for the original problem.

Since the distance between any two adjacent vertices is x,

$$f(\texttt{opt}) \geq f(S^*) \geq |N_{G'}(S^*)|\,\phi(x) = |R'|\,\phi(x). \tag{23}$$

We proceed to the upper bound $f(\texttt{alg})$. We say a vertex p is covered with value y if $f(\texttt{alg}, p) = y$. There are at most k selected vertices, each contributing $\phi(0)$. Other vertices can also have service $\phi(0)$ when ϕ has a plateau; we only separate the selected vertices. Excluding the selected vertices, the vertices on the left side are covered with value at most $\phi(x)$. As for the vertices on the right side, those that are neighbors of $\texttt{alg}_L$ are covered with $\phi(x)$, while all other unselected vertices are covered with at most $\phi(2x)$. It follows that

$$\begin{aligned} f(\texttt{alg}) &\leq k\phi(0) + |L|\,\phi(x) + |N_{G'}(\texttt{alg}_L)|\,\phi(x) + (|R'| - |N_{G'}(\texttt{alg}_L)|)\,\phi(2x) \\ &\leq k\phi(0) + |L|\,\phi(x) + |R'|\,\zeta\phi(x) + |R'|\,(1-\zeta)\phi(2x), \end{aligned}$$

where for the inequality we have used Equation (22) and the fact that $\phi(2x) < \phi(x)$. Since $|R'| \geq C \max\{k, |L|\}$, dividing both sides by $\phi(x)\,|R'|$ we obtain

$$\begin{aligned} \frac{f(\texttt{alg})}{\phi(x)\,|R'|} &\leq \frac{\phi(0)}{C\phi(x)} + \frac{1}{C} + \zeta + (1-\zeta)\frac{\phi(2x)}{\phi(x)} \\ &\leq \frac{\varepsilon}{4} + \zeta + (1-\zeta)\frac{\phi(2x)}{\phi(x)} && \textit{(Definition of } C\textit{)} \\ &\leq \frac{\varepsilon}{4} + (1 - 1/e) + \frac{\varepsilon}{4} + \frac{\phi(2x)}{e\phi(x)} && \textit{(Definition of } \zeta\textit{)} \\ &\leq 1 - 1/e + \frac{\varepsilon}{2} + \frac{\varepsilon}{4e} + \frac{\gamma}{e} && \textit{(Definition of } x\textit{)}. \end{aligned}$$

Together with Equation (23), we obtain $\frac{f(\texttt{alg})}{f(\texttt{opt})} < \xi$, contradicting the assumption that $\mathcal{A}$ achieves approximation ratio ξ.

C Omitted proofs

For any $c > 1$, define $\gamma_{\phi,c} = \inf_{x \geq 0} \frac{\phi(cx)}{\phi(x)}$. We drop the dependence on ϕ throughout the proof.

Lemma 31. *γ_c is decreasing in c. Additionally, any $c > 1$ satisfies $\gamma_{c^2} \geq \gamma_c^2$.*

Proof. The fact that γ_c is decreasing in c follows immediately from the fact that ϕ was assumed to be decreasing. As for $\gamma_{c^2} \geq \gamma_c^2$, by definition of γ,

$$\begin{aligned}
\inf_{x\geq 0} \frac{\phi(c^2 x)}{\phi(x)} &= \inf_{x\geq 0} \left(\frac{\phi(c^2 x)}{\phi(cx)} \frac{\phi(cx)}{\phi(x)} \right) \\
&\geq \inf_{x\geq 0} \left(\frac{\phi(c^2 x)}{\phi(cx)} \right) \inf_{x\geq 0} \left(\frac{\phi(cx)}{\phi(x)} \right) \qquad \textit{(Since } \inf(f\cdot g) \geq \inf(f)\cdot\inf(g)\textit{)} \\
&\geq \inf_{x\geq 0} \left(\frac{\phi(cx)}{\phi(x)} \right)^2,
\end{aligned}$$

as desired. □

Lemma 32. *For any two values $c, c' > 1$, we have $\gamma_{\phi,c} > 0$ if and only if $\gamma_{\phi,c'} > 0$. Additionally, $\gamma_{\phi,c} < 1$ if and only if $\gamma_{\phi,c'} < 1$.*

Proof. Assume without loss of generality that $c > c'$. We first show that $\gamma_c > 0$ if and only if $\gamma_{c'} > 0$. Since $c > c'$, we have $\gamma_c \leq \gamma_{c'}$, it is clear that $\gamma_c > 0$ implies $\gamma_{c'} > 0$. As for the other direction, taking t to be a power of 2 large enough to ensure $c'^t > c$, we have $\gamma_{c'}^t \leq \gamma_{c'^t} \leq \gamma_c$. Therefore, $\gamma_{c'} > 0$ implies $\gamma_c > 0$ as well, finishing the proof. Similarly, if $\gamma_{c'} < 1$, then $\gamma_c < 1$ as well. If $\gamma_c < 1$, then $\gamma_{c'}^t \leq \gamma_c < 1$, finishing the proof. □

Lemma 33. *Any $k \geq 1$ satisfies $\left(1 - \frac{1}{k}\right)^k \leq \frac{1}{e} - \frac{1}{10ek}$.*

Proof. For $k = 1$ the inequality is immediate. Assume $k \geq 2$. Taking logarithms from both sides, we need to show that $k \ln\left(1 - \frac{1}{k}\right) \leq \ln\left(\frac{1}{e} - \frac{1}{10ek}\right)$. The right hand side can be rewritten as

$$\ln(1/e - 1/(10ek)) = \ln(1/e(1 - 1/(10k))) = -1 + \ln(1 - 1/(10k)).$$

Using the Taylor expansion $\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots$, this becomes equivalent to

$$-1 - \frac{1}{2k} - \frac{1}{3k^2} - \cdots \leq -1 - \frac{1}{10k} - \frac{1}{200k^2} - \ldots,$$

this holds however because the left hand side is smaller than the right hand side term by term. Specifically, if we ignore the -1 on both sides, the i-th term in the left hand side is $-\frac{1}{(i+1)k^i}$, while in the right hand side the i-th term is $-\frac{1}{i\cdot 10^i k^i}$. □

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the main result (Theorem 1).

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The results are theoretical and we do not claim practical superiority over heuristic methods, which may perform better in practice depending on the problem instance.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: All assumptions are stated in the theorem statements. Full proofs are provided in the appendix, with proof sketches in the main body.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [N/A]

   Justification: The paper contains no experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [N/A]

Justification: The paper contains no experiments requiring code or data.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [N/A]

   Justification: The paper contains no experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A]

   Justification: The paper contains no experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [N/A]

   Justification: The paper contains no experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The paper presents purely theoretical work and fully conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: The paper presents foundational theoretical work with no direct path to societal impact.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release any models or datasets.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use any assets (code, data, or models).

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not introduce any new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs were used only for writing and editing assistance and do not form part of the core methodology.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.