arXiv:2303.01400v1 [cs.CG] 2 Mar 2023

# C几何相交图中的聚类核心集

SayanBandyapadhyay * Fedor V.Fomin † TanmayInamdar ††

**摘要**

设计核心集——即以较小空间保存数据、使解的代价保持在（$1 \pm \epsilon$）近似因子内的概要——是研究以中心为基础的 $k$聚类问题（如 $k$均值或 $k$中位数）的一个重要方向。Feldman 和 Langberg [STOC'11] 证明，对于一般度量中的 $k$聚类 $n$ 点，可以构造大小关于 $n$呈对数依赖的核心集。此外，在一般度量中，这种关于 $n$ 的依赖性是不可避免的。近期该领域的大量研究致力于针对特殊度量，构造大小与 $n$ 无关的核心集，例如 $d$维欧几里得空间 [Huang、Vishnoi、STOC'20], 倍增度量 [Huang、Jiang、Li、Wu、FOCS'18], 有界树宽图的度量 [Baker、Braverman、Huang、Jiang、Krauthgamer、Wu、ICML'20], 或排除固定小图的图 [Braverman、Jiang、Krauthgamer、Wu、SODA'21]。

本文首次为几何相交图所诱导的度量下的 $k$聚类构造了大小与 $n$无关的核心集。例如，对于欧几里得加权单位圆盘图（UDG）和单位正方形图（USG）中的 $k$聚类，我们得到了大小为$k^{\frac{2}{\log_{\epsilon} \mathcal{O}(1)}}k$ 的核心集。这些构造源自一条一般定理，该定理指出了图度量的两种规范性质，只要满足这两种性质，就足以得到大小与 $n$无关的核心集。我们的定理证明建立在 Cohen-Addad、Saulpic 和 Schwiegelshohn 的近期工作 [STOC '21], 之上；该工作表明，只要存在一个称为质心集的有趣中心集合，就能得到规模较小的核心集。我们工作的主要技术贡献，是证明满足两种典型几何性质的图存在这样一个小规模的质心集。粗略地说，我们利用了这样一个事实：对于距离较近的点，几何相交图的度量与欧几里得度量"相似"；而对于相距较远的点，则与平面图的最短路径度量"相似"。构造小规模质心集的主要技术难点，在于如何结合这两种截然不同的度量。

The 新的核心集构造有助于设计首个针对基于中心的簇 问题的近似算法，适用于单位圆盘图和单位正方形图，并且在 $k$ 和 $\epsilon$ 下是固定参数可解的（固定参数近似方案） .

## 1 引言

聚类是最重要的数据分析技术之一，其目标是将数据集划分为若干组，使每组都包含相似的数据点。数据点之间的相似性由距离函数刻画，而通过最小化代理代价函数，可以找到数据点最自然的聚类。在这项工作中，我们研究了广受关注的 $(k, z)$-聚类问题。

*波特兰州立大学，美国俄勒冈州。

†卑尔根大学，挪威。

‡本研究成果得到了挪威研究理事会通过 BWCA 项目（资助编号：314528）以及欧洲研究理事会（ERC）通过 LOPPRE 项目（参考编号：819416）提供的资助。

**$(k,z)$-聚类**。给定度量空间 $(\Omega,\mathsf{d})$ 中的点集 $P$，以及两个正整数 $k$ 和$z$，求 $\Omega$ 中由 $k$ 个点（或中心）组成的集合 $C$，使以下成本函数最小化：

$$\mathsf{cost}(C)=\sum_{p\in P}\mathsf{cost}(p,C)$$

其中 $\mathsf{cost}(p,C)=(\mathsf{d}(p,C))^z$，且 $\mathsf{d}(p,C)=\min_{c\in C}\mathsf{d}(p,c)$。

两个受到广泛研究的聚类问题—— $k$-均值和 $k$-中位数聚类，分别是 $(k,z)$-聚类在 $z=2$ 和 $z=1$ 条件下的特例。处理大型数据以进行分析的一种常用方法，是将数据缩减方案作为预处理步骤。在聚类问题中，一种这样的数据预处理方法是构造一种称为核心集的对象。

**核心集**。非正式地说，$(k,z)$-聚类的 $\epsilon$-核心集是数据的一个小型摘要，它能够针对任意一组 $k$ 个中心，近似地（误差在 $(1\pm\epsilon)$ 倍以内）保留聚类的代价（我们通常将“$\epsilon$-核心集”简称为“核心集”）。因此，在核心集点上得到的任何中心解集，都可以直接用作原始数据集的解。下面给出正式定义。

**定义 1** ($\epsilon$-核心集)。在度量空间 $(\Omega,\mathsf{d})$ 中，对于点集 $P$ 上的$(k,z)$-聚类问题，核心集是一个带权子集 $Y$，它取自 $\Omega$，权重为 $\omega:Y\to\mathbb{R}^+$，并满足对于任意满足$|\mathcal{S}|=k$的集合 $\mathcal{S}\subseteq\Omega$，

$$\left|\sum_{p\in P}\mathsf{cost}(p,\mathcal{S})-\sum_{p\in Y}\omega(p)\mathsf{cost}(p,\mathcal{S})\right|\le\epsilon\cdot\sum_{p\in P}\mathsf{cost}(p,\mathcal{S}).$$

Feldman 和 Langberg [20] 证明，对于任意一般度量中的 $n$ 个点，可以在时间 $\tilde{\mathcal{O}}(nk)$ 内构造一个大小为$\mathcal{O}(\epsilon^{-2z}k\log k\log n)$ 的核心集，其中 $\tilde{\mathcal{O}}()$ 符号隐藏了一个多对数因子。众所周知，上述界中对 $\log n$ 的依赖无法避免[3, 18]。然而，对于若干特殊度量，可以构造大小与数据规模无关的核心集。欧几里得空间方面已有大量研究，最终得到了界 $\mathcal{O}(k\cdot(\log k)^{\mathcal{O}(1)}\cdot 2^{\mathcal{O}(z\log z)}\epsilon^{-2}\cdot\min\{\epsilon^{-z},k\})$ [18],，该界与数据规模和空间维度无关。此外，这一问题也在其他特殊设置中得到研究，例如倍增度量 [28], 树宽有界图中的最短路径度量 [3], 以及排除固定小图的图 [9]。Cohen-Addad 等人 [18] 近期提出了一个统一框架，涵盖了所有这些结果。我们指出，由于核心集中的不同（加权）点的数量通常要少得多（有时甚至与 $n$无关），因此它们自然适用于流式处理等非顺序场景 [14, 27]。

我们要指出的是，关于图度量中小型核心集的所有已知结果，都高度利用了图的稀疏性，例如有界树宽 [3]或排除某个固定禁用子式 [9]。这其中有充分的理由。在完全图中，通过为边设置适当的权重，可以表示任意 一般度量。因此，如果一个图族包含规模很大的团，那么在此类图中的聚类就与一般度量中的聚类同样困难。

在这项工作中，我们关注具有最短路径度量的边加权几何相交图的核心集构造。由一组几何对象构成的几何相交图中，每个对象对应一个顶点，每对交集非空的对象对应一条边。（我们指出，为了设计算法，我们并不需要显式使用这些对象或其几何表示。只要给定边权重，仅使用图表示就足够了。）特别是，几何相交图是自组织通信和无线传感器网络中广泛研究的一种模型 [4, 32, 35, 38, 41]。值得注意的是，

聚类是此类网络中常见的拓扑管理方法。分组节点可作为子程序，用于以分布式方式执行各种任务以及进行资源管理；有关无线传感器网络中不同聚类方法的概述，请参见综述 [42] 。

我们的工作受到以下问题的启发：“是否可以利用几何相交图的性质，构造大小不依赖于数据规模的核心集？”一般来说，这个问题的答案是否。这是因为几何相交图可能包含大型团。即使对象简单如单位正方形，相应的相交图也可能是一个团；并且，正如我们已经指出的，通过为团的边设置合适的权重，可以表示任意度量。因此，在带有任意边权重的几何相交图中构造核心集，其难度与在一般度量中构造核心集相同。因此，我们需要以某种方式限制边权重，才能为几何相交图获得非平凡的核心集。作为一个说明性示例，我们来看看欧几里得加权单位圆盘图——这是一类经过深入研究的几何相交图。

**欧几里得加权单位圆盘图度量。**单位圆盘图（UDG）的定义如下：平面上有一组半径为 1 的闭圆盘，且顶点与这些圆盘的中心之间存在一一对应关系；当且仅当两个对应中心所在的圆盘相交时，这两个顶点之间才有一条边。边的权重等于两个对应中心之间的欧几里得距离。欧几里得加权单位圆盘图已在计算几何领域得到广泛研究 [11, 25]。除了实际应用方面的动机，单位圆盘图在理论上也颇有意义。一方面，由于嵌入在平面上，它们在“放大视野”时类似于平面图，但局部可能包含大型团。另一方面，它们所诱导的度量融合了几何和图设置：局部呈欧几里得特性，而整体上属于图度量。得益于后一特性，单位圆盘图度量可用于精细化聚类；而仅使用纯粹的欧几里得距离，只能找出由空间凸划分所诱导的簇（见图 1）。

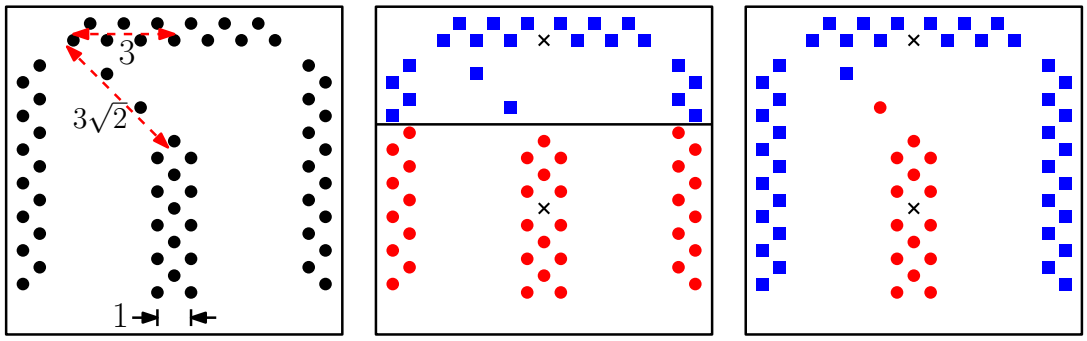


图 1：（左）二维点集。（中）基于欧几里得距离的二均值聚类。中心用叉号表示。两个簇中的点分别用圆盘（红色）和正方形（蓝色）表示。（右）在单位圆盘图上进行二均值聚类——由于 $3\sqrt{2}$ 对角线路径距离，所有蓝色点（正方形）到上方中心的距离都更近。

## 1.1 我们的结果。

现在，我们通过指出图 $G$ 的两种典型几何性质，来将我们对欧几里得加权单位圆盘图“混合”特性的直觉形式化；这两种性质足以用于构造

小规模核心集。为便于阐述，我们先固定一些记号。对于 $G$的任意子图 $H$，我们分别用 $V(H)$ 和 $E(H)$ 表示其顶点集和边集。对于顶点集 $V' \subseteq V$，我们用 $G[V']$ 表示由 $V'$诱导的 $G$ 子图。对于 $G$的子图 $H$ 以及 $u$、$v \in V(H)$，令 $\pi_H(u,v)$ 表示 $u$ 与 $v$之间的一条最短路径（根据 $G$ 中限制在 $H$上的边权重计算），且该路径使用 $H$的边；令 $\mathsf{d}_H(u,v)$ 表示 $\pi_H(u,v)$ 的权重，即沿 $\pi_H(u,v)$ 的边权重之和。对于图中的任意路径 $\pi$，令 $|\pi|$ 表示 $\pi$上的边数。注意，$\mathsf{d}_H$ 是 $H$上的所谓最短路径度量。最后，对于任意一对点$p$、$q \in \mathbb{R}^2$，令 $|pq|$ 表示 $p$ 与 $q$之间的欧几里得距离（即 $\ell_2$-范数距离）。

**典型几何性质**。
(1) 局部欧几里得：存在（不一定互不相同的）常数 $c_1$、$c_2$、$c_3$、$c_4 \geq 0$，使得
以下结论成立。 $G$ 有一个嵌入 $\lambda: V(G) \to \mathbb{R}^2$ 在平面中，使得 $G$的顶点映射到平面中的点，并满足以下两个性质。

1. 对于任意两个 $u,v \in V(G)$，若 $|\lambda(u)\lambda(v)| \leq c_1$，则 $uv \in E(G)$；并且对于任意 $u',v' \in V(G)$，若$|\lambda(u')\lambda(v')|_2 > c_2$，则 $u'v' \notin E(G)$。2. 对于任意满足 $uv \in E(G)$ 的 $u,v \in V(G)$，令 $w(uv)$ 表示边 $uv$ 的权重。那么，边 $uv$ 是 $G$ 中 $u$ 和 $v$ 之间的一条最短路径。此外，$c_3 \cdot |\lambda(u)\lambda(v)| \leq w(uv) \leq c_4 \cdot |\lambda(u)\lambda(v)|$。

(2) 平面伸展子图：对于任意满足 $V' \subseteq V(G)$ 的导出子图 $G' = G[V']$，存在一个平面 $\alpha$-伸展子图 $H' = (V', E(H'))$，其中 $\alpha \geq 1$ 是某个固定值，即：(i) $H'$ 是 $G'$ 的子图（因此也是 $G$ 的子图）—— $E(H') \subseteq E(G')$；以及 (ii) 对任意 $u, v \in V'$，$\mathsf{d}_{G'}(u, v) \leq \mathsf{d}_{H'}(u, v) \leq \alpha \cdot \mathsf{d}_{G'}(u, v)$。

我们的主要结果如下定理所示。

**定理 1**（非正式表述）。考虑由任意满足典型几何性质 *(1)* 和 *(2)* 的图 $G$ 所诱导的度量空间 $(V,\mathsf{d}_G)$，以及集合 $P \subseteq V(G)$。那么，存在一个多项式时间算法，可为 $P$ 上的 $(k,z)$-聚类构造一个核心集，其大小为 $\mathcal{O}(\epsilon^{-\beta} k \log^2 k)$，其中$\beta = \mathcal{O}(z \log z)$。

定理 1 是一个实用的工具，可用于为若干有趣的几何相交图及其配套度量构造核心集。首先，观察一下我们的初始示例，即由欧几里得加权单位圆盘图 $G$ 诱导的度量满足这两项规范性质。考虑将 $G$ 嵌入到 $\mathbb{R}^2$中。注意，任意两点之间有边当且仅当这两点的欧几里得距离至多为 2，且这条边的权重恰好等于该欧几里得距离。因此，$G$ 是局部欧几里得的，其中 $c_1 = c_2 = 2$，且 $c_3 = c_4 = 1$。此外，根据 Li、Calinescu 和 Wan 的一项结果 [37]，，任意欧几里得加权单位圆盘图都存在一个常数伸展平面伸展子图（参见命题 6）。因此，$G$ 也满足平面伸展子图性质。因此，根据定理 1，我们可以为欧几里得加权单位圆盘图上的 $(k, z)$-聚类构造大小为 $\mathcal{O}(\epsilon^{-\beta} k \log^2 k)$ 的核心集。下文将进一步讨论我们框架的应用。

**$\ell_\infty$-加权单位正方形图度量**。单位正方形图（USG）与单位圆盘图相似，不同之处在于，它们是由（轴平行的）单位正方形而非单位圆盘的交集定义的 [1]。事实上，这两个图类并不相同。例如，$K_{1,5}$ 爪图可以由单位圆盘图实现，却无法由任何单位正方形图实现。（见图 2。）由于单位正方形是 $\ell_\infty$-范数下的“单位球”，因此考虑边上的 $\ell_\infty$ 权重更为自然。不难看出，局部欧几里得

[1]虽然乍看之下可能不太自然，但将 *unit square* 定义为边长为 2 的正方形会很方便。这类似于将单位圆盘定义为直径为 2 的圆盘。在这两种情况下，单位正方形图的类别都不受缩放影响。

性质对 $\ell_\infty$-加权单位正方形图成立——我们将在第 4.2 节给出形式化证明。另一方面，为了确立第二个性质，我们必须证明单位正方形图存在常数伸展平面生成器。据我们所知，此前尚无这一结果，而它本身也具有独立价值。我们将在第 5 节证明这一结果。因此， $\ell_\infty$ 加权单位正方形图也满足应用定理 1 所需的两个性质，从而能够得到一个规模较小的核心集。

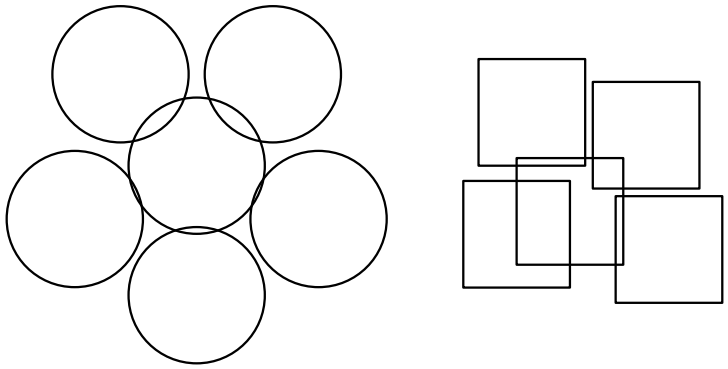

图 2：展示实现$K_{1,5}$ 的一组圆盘，以及实现 $K_{1,4}$的一组正方形。两对相交的单位正方形必须彼此包含对方的一个角，因此，中央正方形至多能与另外四个两两不相交的单位正方形相交。

**其他扩展**。 在 $\mathbb{R}^2$ 中，所有 $\ell_p$ 距离（$1 \le p \le \infty$）彼此相差至多一个 $\sqrt{2}$ 倍数。因此，我们的论证可以轻松扩展到 UDG/USG 上任意 $\ell_p$ 权重的情形，对任意$1 \le p \le \infty$ 都不会改变界限。我们在第 4.3 节中正式证明这一点。最后，我们考虑最大度有界的无权（即跳数距离）单位圆盘图中的最短路径度量。值得注意的是，根据[7]的一项结果，这些图满足平面伸展子图性质。尽管如此，我们在第 4.4 节中表明，可以修改我们的方法，为这类度量构造一个规模较小的核心集。总而言之，对于以下图度量，我们获得了大小与 $n$ 无关的 $(k, z)$-聚类核心集。

- 任意 $1 \le p \le \infty$ 的 $\ell_p$-距离加权单位圆盘图，
- 任意 $1 \le p \le \infty$ 的 $\ell_p$-距离加权 USG，
- 有界度无权单位圆盘图。

**FPT 近似方案**。 作为定理 1的推论，我们得到几何相交图上（$k$， $z$）聚类的（$1+\epsilon$）-近似算法；该算法关于$k$ 和 $\epsilon$是固定参数可解（FPT）的。值得注意的是，即使对于单位圆盘图，这样的（$1+\epsilon$）-近似此前也尚不为人所知，因为它无法由先前已知的核心集大小界推出。在我们的工作之前，对于单位圆盘图——正如一般度量中一样——已知的最佳界是 $O(k \log\, n \cdot \epsilon^{-\max(2,z)})$ [19]。我们注意到，尽管核心集将待聚类的不同点（或客户点）数量减少了，但潜在中心（或设施）的数量仍保持不变，即 $n$。因此，核心集并不能直接帮助我们枚举所有可能的 $k$ 中心集合，从中选出最佳集合。枚举这些中心集合的另一种方法，是枚举核心集点的所有可能划分（或聚类）。注意，核心集点的每一种聚类都对应于原始点的一种聚类，并且聚类代价在一个（$1 \pm \epsilon$）因子范围内保持不变。根据我们的核心集大小界

为 $\mathcal{O}$（$\epsilon^{-\beta}k\log^2 k$），不同聚类的数量仅为 $k^{\mathcal{O}(\epsilon^{-\beta}k_{\log^2 k})}$。由于时间复杂度主要由计算所有划分的聚类中心和代价所决定，该算法的总体运行时间为 $k^{\mathcal{O}(\epsilon^{-\beta}k_{\log^2 k})}\ n^{\mathcal{O}(1)}$ 时间。

**推论 1.** 对于上述列出的每一种度量，都存在一种$(1+\epsilon)$-近似算法，用于$(k,z)$-聚类，且其运行时间为 $2^{\mathcal{O}(\epsilon^{-\beta}k_{\log^3 k})}n^{\mathcal{O}(1)}$，其中 $\beta=\mathcal{O}(z\log\ z)$。

## 1.2 方法概述

我们的核心集构造基于 Cohen-Addad、Saulpic 和 Schwiegelshohn[19], 近期的一项工作，该工作提出了一个框架，可在各种情形下构造核心集。该框架的核心思想是，将核心集构造问题转化为证明某个有趣的中心集或质心集存在。具体而言，考虑任意一个由 $k$ 个中心组成的集合 $\mathcal{S}$，以及任意一个点子集$X\subseteq P$，其中这些点与 $\mathcal{S}$ 的距离相对于某个已有解 $\mathcal{A}$足够近。那么，若子集$\mathbb{C}\subseteq\Omega$ 中包含能够良好近似 $\mathcal{S}$的中心，即存在

$\tilde{\mathcal{S}}\subseteq\mathbb{C}$，使得对于每个 $p\in X$，都有 $|\text{cost}(p,\mathcal{S})-\text{cost}(p,\tilde{\mathcal{S}})|\leq\epsilon(\text{cost}(p,\mathcal{S})+\text{cost}(p,\mathcal{A}))$，则称 $\Omega$ 为 $P$ 的质心集。该框架非正式地指出，如果存在一个质心集 $\mathbb{C}$，那么就可以构造一个核心集，其大小对 $|\mathbb{C}|$呈对数依赖。这种依赖源于他们的随机化构造：为了证明对所有可能的有趣解成立的并集界，需要引入这一项，而有趣解的数量至多为 $|\mathbb{C}|^k$。通过证明小规模质心集的存在，他们为多种空间改进了核心集的大小界。

我们采用 Cohen-Addad 等人的框架来构造核心集。我们的主要技术结果表明，对于度量为 $\mathsf{d}_G$ 的图 $G$，如果满足两种典型几何性质，那么就存在一个规模较小的 $G$ 质心集。这是证明中最具挑战性的部分，需要将计算几何中的工具和技术以一种新颖的方式结合起来。一旦我们证明质心集存在，核心集的构造便可沿用[19]的步骤。为便于说明，我们考虑一个具体例子：欧几里得加权单位圆盘图。

处理单位圆盘图时遇到的首要难题，是这类图融合了欧几里得情形与图论 度量或最短路径度量。例如，考虑一个中心为 $s$ 的点簇。距离较近（即距离不超过 2）的点 $s$表现得就像欧几里得情形中的点。但距离 $s$ 较远的点 $p$，其最短路径距离可能远大于 $p$ 与 $s$之间实际的欧几里得距离，见图 1。我们首先表明，可以从概念上将这两种情况区分开来——但必须谨慎处理，因为一个簇中可能同时包含这两类点。值得注意的是，先前的研究都不必处理这种混合度量。为了处理距离较近的点集，我们利用局部欧几里得 性质。具体而言，通过叠加边长足够小的网格，并从网格的每个单元中选取一个代表点，我们就能计算出一个保留近处点之间距离的质心集。

在另一种情况下，点 $p$ 与中心 $s$之间的最短路径由多条边组成，因此我们需要处理图度量。这种情况有趣得多。其他所有在特定图度量中构造小规模质心集的研究，都利用了某些图类具有小型或性质良好的分隔器这一事实。例如，有界树宽图具有大小受树宽限制的分隔器；而排除固定禁用子式的图则具有最短路径分隔器。然而，单位圆盘图可以包含任意大的团，因此一般而言并不具有这类分隔器。由此，我们遇到了一个技术瓶颈。请注意，这是首个处理此类稠密图的研究。为克服这一挑战，我们利用另一项典型性质。我们不直接处理单位圆盘图，而是考虑它的平面伸展子图，其中的距离至多相差一个常数因子。第二项

典型性质是平面伸展子图。由于平面图具有最短路径分隔器，我们现在可以应用现有技术。然而，如果完全依赖该伸展子图，某些距离可能会在其中放大一个常数（$> 2$）倍，因此无法确保构造核心集所需的（$1 \pm \epsilon$）因子界。

因此，我们以下述方式将伸展子图用作支撑图。首先，我们利用平面伸展子图所具有的最短路径分隔器，对原始单位圆盘图进行递归分解。值得注意的是，尽管平面图分解已用于核心集相关研究，但利用这种引导式方案来分解更一般的图是一项新颖工作。接着，我们利用这一单位圆盘图的递归分解，以及用于求得该分解的最短路径分隔器，来构造质心集。在这一构造中，我们以受限的方式使用伸展子图，使其带来的误差不超过从点 $p$ 到其对应的（近似）中心的最短路径上至多一条边的权重的 $\alpha$ 倍。然而，请注意，如果这样的最短路径仅由一条边组成，那么即使这一误差也过大。为了解决这个问题，我们仅在最短路径“足够长”时才依赖平面伸展子图，也就是说，该路径包含 $\Omega(z/\epsilon)$ 条边。在这种情况下，局部欧几里得性质意味着，对于这样的“长路径”，路径的长度 与路径上的边数 相差至多一个常数因子。这意味着，使用伸展子图对一条边进行改道所引入的误差至多为路径长度的 $\epsilon$ 倍，因此可以忽略不计。

最后，如果点与中心之间的最短路径由 $\mathcal{O}(z/\epsilon)$ 条边组成，那么我们可以使用经过修改的网格单元论证来获得一个规模较小的质心集。

需要指出的是，这只是对这三种情形各自面临挑战的直观概述。质心集的实际构造，以及对每种情形下所产生误差的分析，都相当复杂。当用另一个中心 $\tilde{s} \in \mathbb{C}$ 替代中心 $s \in \mathcal{S}$ 时，我们需要确保对于最近中心为 $s$ 的点 $p$，$\mathsf{d}_G(p, \tilde{s})$ 与 $\mathsf{d}_G(p, s)$ 相比既不会过大，也不会过小，因为我们希望对绝对差值中的误差进行界定。此外，在合并针对这三种情形分别构造的三个质心（子）集时，我们必须格外谨慎，并确保在某种情形下为中心 $s$ 找到的良好替代点 $\tilde{s}$ 不会对另一种情形中所处理点的误差造成不利影响。

**相关工作**。 这里概述核心集方面的文献。更完整的文献列表请参阅 [19, 29]。核心集构造最初因一系列工作而受到关注，这些工作在低维欧几里得空间中获得了规模较小的核心集 [23, 26, 27]。Chen [14]为欧几里得空间构造了第一个维度依赖为多项式的核心集，并为一般度量构造了第一个核心集；对于 $k$-median，其规模为 $\mathcal{O}(k^2\epsilon^{-2}\log^2 n)$。此后，维度依赖性得到了进一步改善 [21, 34]。最后，在[22, 43]中，这种依赖性被消除。近期的改进请参见 [5, 9, 18, 19, 29]。

在一般度量中，$k$-中位数和 $k$-均值问题都存在多项式时间的 $\mathcal{O}(1)$-近似算法 [1, 10, 12, 13, 30, 36]。此外，还可以得到近似保证更优、且关于 $k$ 和 $\epsilon$ [16] 为固定参数可解的算法。自然地，这些问题也在专用度量中得到了研究。对于欧几里得 $k$-中位数问题 [2]和 $k$-均值问题 [17, 24]，已有多项式时间近似方案（PTAS）。其他改进见 [15, 33] 。与几何聚类类似，图设置中的聚类也得到了广泛研究。对于排除子式图 [9, 17]，已有多项式时间近似方案。此外，对于树宽有界的图 [3] 和有界高速公路维数的图 [6, 9]，也已知固定参数可解近似方案。

## 2 几何图的核心集

为做好铺垫，我们需要给出以下来自 [19的质心集定义 ].

**定义 2** （质心集）。 考虑任意度量空间 （Ω、d）、客户点集合 $P \subseteq \Omega$，以及两个正整数 $k$ 和 $z$。令 $\epsilon > 0$ 为精度参数。给定中心集合 $\mathcal{A}$，集合 $\mathbb{C}$ 是关于 $P$ 上的 （$k$，$z$）-聚类的 $\mathcal{A}$-近似质心集，并满足以下性质。

对于每个 $\mathcal{S} \subseteq \mathbb{C}$，使得对于满足代价（$p$,$\mathcal{S}$）在 $\left(\frac{10z}{\epsilon}\right)^z \cdot$ 代价（$p$，$\mathcal{A}$） 或 代价（$p$，$\tilde{\mathcal{S}}$） $\leq \left(\frac{10z}{\epsilon}\right)^z \cdot$ 代价（$p$，$\mathcal{A}$）的每个 $p \in P$，均有

$$|\mathsf{cost}(p,\mathcal{S}) - \mathsf{cost}(p,\tilde{\mathcal{S}})| \leq \frac{\epsilon}{z\log(z/\epsilon)}(\mathsf{cost}(p,\mathcal{S}) + \mathsf{cost}(p,\mathcal{A})).$$

非正式地说，质心集 $\mathbb{C}$ 是候选中心点的集合，其规模可能远小于˜Ω；在不使˜点的代价发生较大变化的情况下，可以将其中的 $\mathcal{S}$ 中心替换为 $\mathcal{S} \subseteq \mathbb{C}$ 中的 $k$ 中心，而这些中心与 $\mathcal{S}$ 或 $\mathcal{S}$ 的距离相较于 $\mathcal{A}$ 而言（相对于 d）要近得多。他们证明了，可以获得规模仅以任意此类质心集大小的对数形式增长的核心集。

**命题 1** ([19])。 考虑任意度量空间 (Ω,d)，一个点集 $P \subseteq \Omega$，其中包含 $n$ 个互不相同的点，以及两个正整数 $k$ 和 $z$。令 $\epsilon > 0$ 为精度参数。假设 $\mathcal{A}$ 是针对 $P$上$(k,z)$-聚类的一个给定的常数因子近似算法。

假设存在一个针对 $P$上$(k,z)$-聚类的 $\mathcal{A}$-近似质心集。那么，存在一个多项式时间算法，以至少 $1-\delta$ 的概率构造一个大小为

$$\mathcal{O}\left(\frac{2^{\mathcal{O}(z\log z)} \cdot \log^4(1/\epsilon)}{\min\{\epsilon^2, \epsilon^z\}}(k\log|\mathbb{C}| + \log\log(1/\epsilon) + \log(1/\delta))\right)$$

并带有正权重的核心集，用于 $P$上的$(k,z)$-聚类。

首先请注意，上述核心集框架只要求这样的质心集存在，无需显式计算它。事实上，在分析中，这样的质心集仅用于界定所计算核心集的大小。

我们的主要技术贡献是证明了一个定理，该定理保证当几何图度量满足两个典型的几何性质时，存在一个较小的质心集。

**定理 2.** *(质心集定理)* 考虑由满足前文定义的局部欧几里得和平面伸展子图性质的任意图$G = (V,E)$所诱导的度量空间 $(V,\mathsf{d}_G)$ 。再考虑点集 $X \subseteq V$ 以及两个正整数 $k$ 和 $z \geq 1$。令 $\epsilon > 0$ 为精度参数。此外，假设 $\mathcal{A}$ 是 $X$上的$(k,z)$-聚类的一个解。那么，存在一个 $\mathcal{A}$-近似质心集 $\mathbb{C}$，用于 $X$ 上的$(k,z)$-聚类，其大小为 $\exp(\mathcal{O}(\log^2|X| + z^{16}\epsilon^{-8}(\log(z/\epsilon))^8\log|X|))$。

我们将在下一节证明定理 2。随后，根据命题 1 并结合 [9], ，稍作修改以适应我们不同的核心集大小界，即可得到所需的核心集结果。为完整起见，我们将给出证明。

**定理 3**。 考虑由满足局部欧几里得和平面伸展子图性质的任意图 $G = (V,E)$ 所诱导的度量空间 $(V$、$\mathsf{d}_G)$、包含 $n$ 个互异点的集合 $P \subseteq V$，以及两个正整数 $k$ 和 $z \geq 1$。则存在一个多项式时间算法，以至少 $1-\delta$ 的概率，为 $P$上的 $(k,z)$-聚类构造一个大小为 $\mathcal{O}(\epsilon^{-\mathcal{O}(z_{\log z})}k\log^2 k\log^3(1/\delta))$的核心集，其中 $z$为常数，且 $\delta < 1/4$。

证明。为证明该定理，我们应用迭代式规模缩减算法 [9]。令 $t$ 为满足 $\log^{(t-1)} n \geq \max\{125k\epsilon^{-\rho}\log(1/\delta),\ \rho 2^{\rho+1}\}$ 的最大整数，其中 $\rho$ 为稍后设定的充分大的常数。此外，令 $X_0 = P$。对于 $1 \leq i \leq t$，令 $\epsilon_i = \epsilon/(\log^{(i)} n)^{1/\rho}$ 并令 $\delta_i = \delta/|X_{i-1}|$。令 $X_i$为将命题 1 的算法应用于客户集合 $X_{i-1}$ 所计算得到的核心集，其中 $\epsilon$为 $\epsilon_i$，$\delta$ 为 $\delta_i$，且适用于 $1 \leq i \leq t$。最后，令 $U$ 为将相同算法应用于 $X_t$ 后返回的核心集，其近似保证为 $1+\epsilon$，失败概率为 $\delta$。

首先，我们讨论连续两次迭代之间核心集大小的递减速率。令 $n_i$ 表示 $1 \leq i \leq t$的 $X_i$ 大小。那么，

$$
\begin{aligned}
n_i &= \frac{2^{\mathcal{O}(z\log z)}\cdot \log^4(1/\epsilon_i)}{\min\{\epsilon_i^2, \epsilon_i^z\}}(k\log|\mathbb{C}_i| + \log\log(1/\epsilon_i) + \log(1/\delta_i)) \quad (\mathbb{C}_i \text{ is the centroid set w.r.t. } X_i)\\
&= \frac{2^{\mathcal{O}(z\log z)}\cdot \log^4(1/\epsilon_i)}{\min\{\epsilon_i^2, \epsilon_i^z\}}(k\log^2 n_{i-1} + kz^9\epsilon_i^{-5}\log n_{i-1} + \log\log(1/\epsilon_i) + \log(1/\delta_i))\\
&\leq \epsilon_i^{-\alpha} k\log^2 n_{i-1}\log(1/\delta_i) \qquad (\alpha = \mathcal{O}(z\log z))\\
&\leq \epsilon_i^{-\alpha} k\log^2 n_{i-1}(\log n_{i-1} + \log(1/\delta))\\
&\leq \epsilon_i^{-\alpha} k\log^3 n_{i-1}\log(1/\delta)
\end{aligned}
$$

最后一个不等式成立，因为 $n_{i-1} \geq 4$ 且 $\delta < 1/4$。令 $\Gamma = k\epsilon^{-\alpha}\log(1/\delta)$。接下来，我们证明，对于 $1 \leq i \leq t$，$n_i \leq 125\Gamma(\log^{(i)} n)^4$。我们对 $i$使用归纳法。在基础情形中，$n_1 \leq \epsilon_1^{-\alpha} k\log^3 n\log(1/\delta) = k\epsilon^{-\alpha}\log(1/\delta)(\log n)^{\alpha/\rho}\log^3 n = \Gamma\log^4 n \leq 125\Gamma\log^4 n$。倒数第二个等式是通过令 $\rho$ 的值等于 $\alpha$得到的。现在，考虑任意一个 $i \geq 2$。

$$
\begin{aligned}
n_i &\leq \epsilon_i^{-\alpha} k\log^3 n_{i-1}\log(1/\delta)\\
&\leq (\epsilon/(\log^{(i)} n)^{1/\rho})^{-\alpha} k\log^3 n_{i-1}\log(1/\delta)\\
&= k\epsilon^{-\alpha}\log(1/\delta)(\log^{(i)} n)^{\alpha/\rho}\log^3 n_{i-1}\\
&= \Gamma(\log^{(i)} n)\log^3 n_{i-1} \qquad (\text{as } \rho = \alpha)\\
&\leq \Gamma(\log^{(i)} n)(\log(125\Gamma(\log^{(i-1)} n)^4))^3\\
&\leq \Gamma(\log^{(i)} n)(\log(125\Gamma) + 4\log^{(i)} n)^3\\
&\leq \Gamma(\log^{(i)} n)(\log^{(i)} n + 4\log^{(i)} n)^3 \quad (\text{by definition of } t,\ \log^{(i-1)} n \geq 125k\epsilon^{-\rho}\log(1/\delta) = 125\Gamma)\\
&= 125\Gamma(\log^{(i)} n)^4
\end{aligned}
$$

由此可得 $n_t \leq 125\Gamma(\log^{(t)} n)^4$。现在，根据 $t$的定义，$\log^{(t)} n < \max\{125\Gamma, \rho 2^{\rho+1}\}$。因此，$n_t \leq \max\{(125\Gamma)^5,\ 125\Gamma(\rho 2^{\rho+1})^4\}$。接下来，我们分析以下断言中的误差。

**断言1**。$\prod_{i=1}^{t}(1+\epsilon_i) \leq 1 + 10\epsilon$.

证明。根据定义，$X_t$ 是一个核心集，其近似保证至多为 $\Pi_{i=1}^{t}(1+\epsilon_i)$。现在，根据我们的假设，对于 $i \leq t$，$\log^{(i-1)} n \geq \rho 2^{\rho+1}$。对于这样的 $\log^{(i-1)} n$ 值，

$$
\begin{aligned}
&\log^{(i-1)} n \geq 2^{\rho} \log^{(i)} n \\
&\Rightarrow (\log^{(i-1)} n)^{1/\rho} \geq 2(\log^{(i)} n)^{1/\rho} \\
&\Rightarrow \epsilon/(\log^{(i-1)} n)^{1/\rho} \leq (1/2) \cdot (\epsilon/(\log^{(i)} n)^{1/\rho}) \\
&\Rightarrow \epsilon_i \geq 2\epsilon_{i-1} \\
&\Rightarrow \sum_{i=1}^{t} \epsilon_i \leq 2\epsilon_t
\end{aligned}
$$

由此可得

$$
\prod_{i=1}^{t}(1+\epsilon_i) \leq \exp\left(\sum_{i=1}^{t}\epsilon_i\right) \leq \exp\left(2\epsilon_t\right) \leq \exp\left(\frac{2\epsilon}{(\log^{(t)} n)^{1/\rho}}\right) \leq \exp\left(2\epsilon\right) \leq 1+10\epsilon.
$$

倒数第二个不等式由以下事实得出：$\log^{(t)} n \geq 1$，这在 $\log^{(t-1)} n \geq \rho 2^{\rho+1}$ 时成立。

□

接下来，我们分析失败概率。首先注意，我们可以假设 $X_{i-1}$ 的大小至少为 $\log^{(i-1)} n$。否则，我们可以向核心集 $X_{i-1}$ 中添加任意点并增大其大小。因此，$n_{i-1} \geq \log^{(i-1)} n$。于是，$\delta_i = \delta/n_{i-1} \leq \delta/\log^{(i-1)} n$。所以，总失败概率至多为：

$$
\sum_{i=1}^{t}\delta_i = \delta\left(\frac{1}{n} + \frac{1}{\log n} + \ldots + \frac{1}{\log^{(t-1)} n}\right) = \mathcal{O}(\delta)
$$

最后一个不等式成立，是因为 $\log^{(t-1)} n \geq \rho 2^{\rho+1} \geq 4$。

上文证明了，$X_t$ 是一个大小至多为 $\max\{(125\Gamma)^5,\ 125\Gamma(\rho 2^{\rho+1})^4\}$ 的核心集，其近似保证为 $1+\mathcal{O}(\epsilon)$，失败概率至多为 $\mathcal{O}(\delta)$。因此，通过对 $X_t$ 应用命题 1 的算法得到的集合 $U$ 是一个核心集，其近似保证为 $1+\mathcal{O}(\epsilon)$，失败概率至多为 $\mathcal{O}(\delta)$，且大小至多为

$$
\begin{aligned}
\epsilon^{-\alpha} k \log^2 |X_t| \log(1/\delta) &= \mathcal{O}(\epsilon^{-\alpha} k \log(1/\delta) \log^2 \Gamma) \\
&= \mathcal{O}(\epsilon^{-\alpha} k \log(1/\delta)(\log k + \log(1/\epsilon) + \log\log(1/\delta))^2) \\
&= \mathcal{O}(\epsilon^{-\beta} k \log^2 k \log^3(1/\delta))
\end{aligned}
$$

其中 $\beta = \mathcal{O}(\alpha) = \mathcal{O}(z \log z)$。将 $\epsilon$ 和 $\delta$ 按常数进行缩放，即可得到所需的界。最后，由于 $t+1 = \mathcal{O}(\log^* n)$，我们将命题 1 $\mathcal{O}$中的 $\mathcal{O}(n)$ 时间算法应用 $^* n$ 次。此外，为了应用该算法，我们还需要计算一个近似解 $\mathcal{A}$。为此，我们˜使用 Mettu 和 Plaxton 的 $\mathcal{O}(nk)$ 时间算法 [40]，该算法可返回一个常数近似。

□

# 3 质心集定理的证明

在本节中，我们证明主要结果：小规模质心集的存在性。回顾一下，我们给定了 $G = (V,\ E)$，这是一个有 $n$ 个顶点的连通无向边加权图。此外，$G$ 满足两项典型几何性质：(1) 局部欧几里得，以及 (2)平面伸展子图。

由于 $(V, \mathsf{d}_G)$ 是我们的度量空间，因此我们交替使用“点”和“顶点”这两个术语。 $X \subseteq V$ 是给定的点集。我们还给定了 $\mathcal{A}$，它是 $X$ 上 $(k, z)$-聚类的一个解。我们证明，对于 $X$ 上的 $(k, z)$-聚类，存在一个大小为 $\exp(\mathcal{O}(\log^2 |X| + z^{16}\epsilon^{-8}(\log(z/\epsilon))^8 \log |X|))$ 的 $\mathcal{A}$-近似质心集，并且该集合满足以下性质。

~对于每个中心集合 $k$ $\mathcal{S} \subseteq V$，都存在 $\mathcal{S} \subseteq \mathbb{C}$，使得对于每个满足 $\mathsf{cost}(p, \mathcal{S}) \leq \left(\frac{10z}{\epsilon}\right)^z \cdot \mathsf{cost}(p, \mathcal{A})$ 或 $\mathsf{cost}(p, \tilde{\mathcal{S}}) \leq \left(\frac{10z}{\epsilon}\right)^z \cdot \mathsf{cost}(p, \mathcal{A})$ 的 $p \in X$，都有

$$|\mathsf{cost}(p, \mathcal{S}) - \mathsf{cost}(p, \tilde{\mathcal{S}})| \leq \frac{\epsilon}{z \log(z/\epsilon)} (\mathsf{cost}(p, \mathcal{S}) + \mathsf{cost}(p, \mathcal{A})).$$

## 3.1 有用性质

在分析中，我们将使用局部欧几里得性质的一些推论，陈述如下。

**命题 2**（有界距离。）. 存在普适常数 $c_1', c_2'$，使得 $G$ 中任意一条边的权重至多为 $c_2'$；并且对于任意 $\tau \geq 2$ 和任意 $u, v \in V$（$G$），若 $|\pi_G(u, v)| = \tau$，则 $\mathsf{d}_G(u, v) \geq c_1'\tau$。也就是说，对于 $\tau \geq 2$，$\mathsf{d}_G(u, v) = \Theta(\tau)$。

证明。设 $c_1, c_2, c_3, c_4 \geq 0$ 为满足 $G$ 的常数，且嵌入 $\lambda$ 满足局部欧几里得 性质。考虑一条边 $uv \in E(G)$。由此可得 $|\lambda(u)\lambda(v)| \leq c_1$，这意味着 $w(uv) \leq c_4 \cdot |\lambda(u)\lambda(v)| \leq c_4 \cdot c_1$。因此，任意一条边的权重都有上界 $c_2' := c_1 \cdot c_4$。

现在，考虑一条最少跳数最短路径 $\pi(u, v) = (u = w_0, w_1, \ldots, w_\tau = v)$，其中 $\tau \geq 1$。对于任意 $0 \leq i \leq t-2$，注意到 $w_i w_{i+2} \notin E(G)$——否则我们可以通过捷径得到一条长度更短且跳数更少的路径。这意味着，$|\lambda(w_i)\lambda(w_{i+2})| > c_1$。然后，根据三角不等式可得 $|\lambda(w_i)\lambda(w_{i+1})| + |\lambda(w_{i+1})\lambda(w_{i+2})| > c_1$。另一方面，$\mathsf{d}_G(w_i, w_{i+2}) = \mathsf{d}_G(w_i, w_{i+1}) + \mathsf{d}_G(w_{i+1}, w_{i+2}) \geq c_3 \cdot (|\lambda(w_i)\lambda(w_{i+1})| + |\lambda(w_{i+1})\lambda(w_{i+2})|) > c_3 \cdot c_1$。也就是说，$\pi(u, v)$中每条长度为 2 的子路径，其长度至少为一个常数。因此，若 $\tau \geq 2$，则 $\mathsf{d}_G(u, v) \geq c_3 \cdot c_1 \cdot \lfloor \tau/2 \rfloor \geq c_1' \cdot \tau$。 □

命题 *3*（$\mu$-网的构造）。设 $0 < \mu < c_1/\sqrt{2}$，且 $t \geq 0$。那么，对于任意 $v \in V(G)$，令 $B(v, r) = \{u \in V(G) : d_G(u, v) \leq r\}$。则存在一个大小为 $\mathcal{O}(r^2/\mu^2)$ 的子集 $B' \subseteq B(v, r)$，使得对于任意 $u \in B(v, r)$，都存在 $u' \in B'$，满足 $d_G(u, u') = \mathcal{O}(\mu)$。这样的集合 $B'$ 称为 $B(v, r)$ 的 $\mu$-网。

证明。考虑由局部欧几里得性质保证的嵌入 $\lambda : V(G) \to \mathbb{R}^2$。考虑边长为 $\mu < c_1\sqrt{2}$ 的网格。由此可知，包含在同一网格单元中的任意两点所对应的顶点在 $G$ 中互为邻居。此外，对于任意邻居 $u, v \in V(G)$，可得 $|\lambda(v)\lambda(f(v))| \leq \sqrt{2}\mu \leq c_1$，这意味着 $\mathsf{d}_G(v, f(v)) = w(vf(v)) \leq c_4 \cdot |\lambda(v) f(v)| \leq c_4 \cdot \sqrt{2}\mu = \mathcal{O}(\mu)$。为构造 $B'$，我们从每个单元 $C \cap B(v, r)$ 中任意选取一个点，前提是 $C \cap B(v, r) \neq \emptyset$。

为了界定 $B'$ 的大小，我们考虑两种情况。如果 $r \leq c_1$，那么 $B(v, r)$ 中的所有顶点都是 $v$ 的邻居。因此，所有对应的点都包含在以 $\lambda(v)$ 为中心、半径为 $r/c_3$ 的圆盘内。该圆盘中边长为 $\mu$ 的网格单元数量可由 $\mathcal{O}(r^2/\mu^2)$ 上界估计，我们从每个这样的单元中取一个点加入 $B'$。如果 $c_1 < r < c_2$，那么我们同样可以用类似的论证，将单元数量上界估计为 $\mathcal{O}(c_2^2/\mu^2) = \mathcal{O}(r^2/\mu^2)$。

否则，假设 $r > c_2$。考虑一个顶点 $w$，使得 $\mathsf{d}_G(v, w) \leq r$，但 $v$ 和 $w$ 不是 $G$中的邻居。由此可得 $|\pi_G(v, w)| = \tau \geq 2$。因此，$|\lambda(v)\lambda(w)| \leq c_4\tau$。另一方面，命题 2 表明 $\mathsf{d}_G(v, w) \geq c_1'\tau$。这意味着 $|\lambda(v)\lambda(w)| \leq (c_4/c_1') \cdot r$。由此可知，与 $B(v, r)$中的一个顶点对应的点，位于以 $\lambda(v)$为中心、半径为 $\mathcal{O}(r)$的圆盘内。该圆盘中边长为 $\mu$ 的单元数量为 $\mathcal{O}(r^2/\mu^2)$。 □

命题*4*（有界度支撑图）。对于任意精度值 $0 < \mu < c_1/\sqrt{2}$，存在一个图 $H$，具有以下性质：*(i)* $V(H) \subseteq V(G)$；*(ii)* $H$的最大顶点度由多项式函数 $\Delta(1/\mu) = \mathcal{O}(1/\mu^2)$ 有界；*(iii)* 存在一个映射 $f: V(G) \to V(H)$，使得对于任意 $v \in V(G)$，$d_G(v, f(v)) = \mathcal{O}(\epsilon)$；并且对于 $G$中的任意路径 $u_1, u_2, \ldots, u_t$，$H$都包含路径 $f(u_1)\ f(u_2) \ldots, f(u_t)$。

证明。考虑一个边长为 $\mu$ 的网格单元，如前一个命题所述，这意味着属于同一网格单元的任意两个点都是 $G$中的邻居。现在，我们按如下方式构造支撑图 $H$。从每个包含 $V(G)$中某个点的网格单元中任意选取一个点，并将这个特殊点添加到 $V(H)$中。映射 $f$ 按如下方式构造：将每个顶点 $v \in V(G)$映射到包含 $v$的网格单元中的特殊点。由此可得，$|\lambda(v)\lambda(f(v))| \leq \sqrt{2}\mu \leq c_1$。同样可得，$\mathsf{d}_G(v,\ f(v)) = w(vf(v)) \leq c_4 \cdot |\lambda(v)f(v)| \leq c_4 \cdot \sqrt{2}\mu = \mathcal{O}(\mu)$。

接下来，我们定义 $H$的边。考虑两个网格单元 $C_1$ 和 $C_2$，以及分别位于其中的特殊点 $p_1$和 $p_2$。如果 $C_1$ 和 $C_2$ 包含 $G$中的两个邻居点，则向 $H$ 添加边 $p_1p_2$。因此，可能包含一条边的两个特殊点之间的最大距离至多为$c_2 + \mathcal{O}(\mu)$。由于每个网格单元中至多选取一个特殊点，因此 $H$中任意特殊点的度数均由 $\mathcal{O}(1/\mu^2)$界定。现在，考虑 $G$中的任意路径 $u_1$，$u_2$、.....、$u_t$。则对于 $1 \leq i \leq t-1$，两个特殊点 $f(u_i)$和 $f(u_{i+1})$之间的边存在于 $H$中。因此，路径 $f(u_1)\ f(u_2)$ ....., $f(u_t)$位于 $H$中。 □

在下面的分析中，我们经常需要使用以下适用于距离幂的三角不等式推广。

**引理 1**（幂的三角不等式 [39]）。令 $\epsilon > 0$ 和 $z \geq 1$。

*1.* 对任意非负实数 $x$、$y$，$(x+y)^z \leq (1+\epsilon)^{z-1}x^z + \left(\frac{1+\epsilon}{\epsilon}\right)^{z-1} y^z$.

*2.* 设（Ω，d）为一个度量空间，且 $u$、$v$、$w \in \Omega$，则，

$$\mathsf{d}(u,v)^z \leq (1+\epsilon)^{z-1} \cdot (\mathsf{d}(u,v))^z + \left(\frac{1+\epsilon}{\epsilon}\right)^{z-1} (\mathsf{d}(v,w))^z.$$

为了证明质心集定理，我们给出质心集的一个显式构造。回顾第一典范性质，$V$中彼此接近的点表现得与欧几里得情形相同。为了处理**靠近**最近中心的点这一情形，我们将一个点集 $\mathbb{C}_{\text{net}}$ 加入质心集。对于距离较远的点情形，我们进一步将其分为两个子情形。在第一个子情形中，我们处理这样一类点：它们到最近中心的最短路径仅相距**较少**或$O(1/\epsilon)$跳数。为了处理这一子情形，我们将一个点集 $\mathbb{C}_{\text{support}}$ 加入质心集。最后一个子情形涉及**较长的路径**，在这里我们利用平面生成器性质。具体而言，在这种情形下，我们根据由分解子图的底层平面伸展子图所引导的图递归分解来构造质心点。这一子情形类似于 [9, 18]中的排除子式图度量的质心集构造。

接下来，我们将详细介绍我们的构造。不过，在描述我们的方法之前，我们需要先定义一些基本要素，主要是为了理解由平面扩展器引导的图分解。

### 3.2 最短路径分隔集与递归分解

**命题 5** ([11, 31])**.** 给定一个顶点带有非负权重的平面图 $H = (V, E)$，存在一个最短路径集合 $\mathcal{P} = \{P_1, P_2, \ldots, P_b\}$，其满足 $b = \mathcal{O}(1)$，使得权重

属于 $H \setminus \bigcup_{P_i \in \mathcal{P}} V(P_i)$ 中任意一个连通分量的顶点集的权重，至多为 $V(H)$权重的一半。

令 $X \subseteq V(G)$ 为一个顶点集，并令 $I_X \colon V(G) \to \{0,1\}$ 为 $X$ 的指示函数。我们按如下方式得到一个递归分解 $\mathcal{T} = (\mathcal{R},\ E(\mathcal{T}))$，其中树中的每个节点$R_i \in \mathcal{R}$ 都对应于 $V(G)$ 的一个子集，该子集称为一个区域。

- 与 $\mathcal{T}$ 的根节点对应的区域 $R_1$ 等于 $V(G)$。
- 考虑 $\mathcal{T}$ 的一个内部节点，该节点对应于一个区域 $R_i \subseteq V(G)$，且满足 $|R_i \cap X| > 2$。接着，令 $G_i = G[R_i]$，并令 $H_i$ 为 $G_i$ 的一个平面 $\alpha$-伸展子图。然后，我们对 $H_i$ 应用命题 5，得到一组 $\mathcal{O}(1)$ 分隔路径 $\mathcal{P}_i$，使得每个 $P^{i,\ j} \in \mathcal{P}_i$ 都是 $H_i$ 中关于权重函数 $I_X$ 的一条最短路径（即，我们希望以平衡的方式对 $X$ 的顶点进行分解）。随后，在 $\mathcal{T}$ 中成为 $R_i$ 的子区域的区域如下。（示意见图3）

  1. $H_i$在 $R_i \setminus \bigcup_{P_j^i \in \mathcal{P}_i} V$（$P_j^i$）的诱导子图中，各连通分量对应的顶点子集称为 $R_i$的分量子节点，而

  2. $\mathcal{P}_i$ 中的每条路径都被拆分为极大子路径，每条子路径最多包含 $X$中的两个点。[2] 随后，将每条此类子路径的顶点集作为$R_i$的不同子节点加入。这些子节点称为子路径子节点。

  因此，区域 $R_i$ 等于与其分量子节点和子路径子节点对应的所有区域的并集。注意，只有与分隔路径上连续的子路径子节点对应的区域才可能在 $X$的某个顶点处相交；而与 $R_i$的分量子节点对应的区域则构成$R_i \setminus \bigcup_{P^{i,\ j} \in \mathcal{P}_i} V$（$P_j^i$）的一个划分。

- 当一个区域包含的 $X$顶点不超过 2 个时，递归停止，该区域称为叶区域/节点。由于每一层的权重至少缩小为原来的 $1/2$，因此 $\mathcal{T}$的高度至多为 $\mathcal{O}$（$\log|X|$）。

固定一个顶点 $s \in V(G)$，并令 $R_1, R_2, \ldots, R_t$ 为 $\mathcal{T}$ 中的一条根叶路径，使得 $s$ 包含在每个 $R_i$ 中，其中 $R_1$ 是根区域，$R_t$ 是叶区域（回忆一下，$t = \mathcal{O}(\log|X|)$）。现在考虑任意一个 $p \in X \subseteq V(G)$。令 $R_i$ 为包含 $G$ 中 $p$ 与 $s$之间每条最短路径 $\pi_G(p,s)$ 上所有顶点的最低区域（即 $i$最大的区域）。此外，假设 $R_i$不等于 $R_t$。那么，根据构造，存在 $R_i$的两个子区域 $R_{i+1}, R'_{i+1}$，使得 $p \in R'_{i+1}$ 且 $s \in R_{i+1}$。令 $u$ 为沿着 $\pi_G(p,s)$（从 $p$ 到 $s$）最后一个位于 $R'_{i+1}$中的顶点，并令 $v$ 为沿着 $\pi_G(p,s)$ 紧接在 $u$ 之后的顶点。注意，$u, v$ 属于 $R_i$，而 $R_i$ 在 $H_i$ 中诱导出一个连通分量，因此在 $G_i$ 中也是如此。令 $\pi_{H_i}(u,v)$ 为 $H_i$ 中 $u$ 与 $v$ 之间的一条最短路径。由于 $u \in R'_{i+1}$，且 $v \notin R'_{i+1}$，因此 $\pi_{H_i}(u,v)$ 与路径分隔符 $P^{i,j} \in \mathcal{P}_i$ 相交于顶点 $x$，使得 $x \in V(P^{i,j}) \cap V(\pi_{H_i}(u,v))$。在这种情况下，我们称 $P^{i,j}$ 将 $p$ 与 $s$ 分隔开，而 $x$ 是位于 $u$ 与 $v$ 之间的一个分隔顶点（见图 4）。下面我们总结几个性质。

**断言 2**。令 $P^{i,\ j} \in \mathcal{P}_i$ 为与区域 $R_i$ 相对应的分隔路径，其中 $R_i$ 是包含最短路径 $\pi$（$s, p$）的最低区域，且 $P^{i,\ j}$分隔 $p$和 $s$，而 $x$是位于 $u$和 $v$之间的分隔顶点。那么，

[2]更准确地说，如果一条路径 $P$包含 $X$中的至少一个顶点，则将其视为一个实区间 $[0,1]$，其断点为 $0 \le x_1 \le x_2 \le \ldots \le x_m \le 1$，其中 $x_i$是 $P \cap X$中的顶点。于是，子路径为$[0, x_1]$，$[x_1, x_2]$，.....，$[x_m, 1]$。如果 $P$不包含 $X$中的顶点，则令 $P$本身作为子路径。

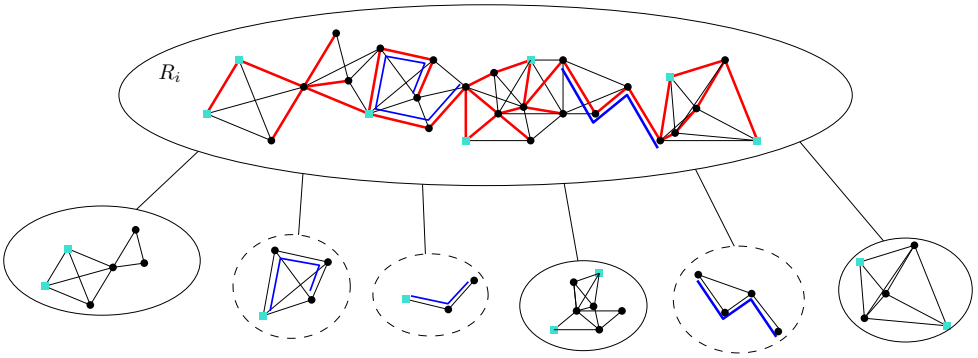


图3：将区域 $R_i$ 分解为多个子区域。在 $R_i$内部，我们展示了导出子图 $G_i = G[R_i]$ （其不是平面的）及其对应的平面伸展子图 $H_i$。 $H_i$ 的边以红色表示，而 $E(G_i) \setminus E(H_i)$中的边以黑色表示。 $H_i$ 的两条最短路径分隔器以蓝色表示，它们为 $X$ 的顶点构成一个平衡分隔器 $\mathcal{P}_i$ （这些顶点显示为浅蓝色正方形）。然后，我们为 $H \setminus V(\mathcal{P}_i)$的每个连通分量添加一个子区域（实线椭圆内的子区域）。我们还为 $\mathcal{P}_i$ 中的路径添加了对应的子区域（虚线椭圆）。其中一条路径由于 $X$中的一个顶点而被分成两段。

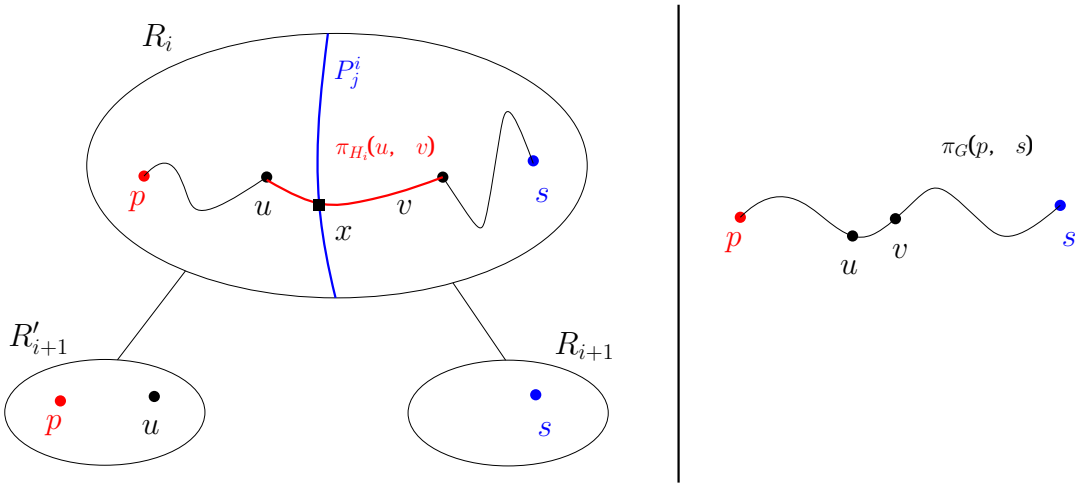


图4：右侧展示了一条最短路径 $\pi_G(p,\ s)$。左侧展示了一条路径 $P^{i,\ j}$，它将 $p$ 与 $s$ 分隔开； $x$ 是位于 $u$ 和 $v$ 之间的一个分隔顶点。请注意， $R_i$ 内的路径 $(P^{i,\ j})$ 是平面伸展子图 $H_i$ 中的最短路径。还请注意，根据定义， $u \in R'_{i+1}$ 和 $s \in R_{i+1}$，但 $v$ 可能不属于 $R_{i+1}$。

1. $\mathrm{d}_{G_i}(p,s) = \mathrm{d}_G(p,s)$，2. $\mathrm{d}_{G_i}(u,v) \le \mathrm{d}_{H_i}(u,v) \le \alpha \cdot \mathrm{d}_{G_i}(u,v) = \alpha \cdot \mathrm{d}_G(u,v)$，3. $\mathrm{d}_{H_i}(u,x) + \mathrm{d}_{H_i}(v,x) = \mathrm{d}_{H_i}(u,v) \le \alpha \cdot \mathrm{d}_{G_i}(u,v) = \alpha \cdot \mathrm{d}_G(u,v)$
4. 对任意 $q > 0$，若 $|\pi_G(p,s)| \ge q \cdot \frac{\alpha \cdot c_2'}{c_1' \cdot \epsilon}$，则 $\mathrm{d}_{H_i}(u,v) \le \epsilon/q \cdot \mathrm{d}_G(p,s)$。

—

证明。第一条性质由假设得出：$R_i$ 包含最短路径 $\pi_G(p,s)$ 上的所有顶点，因此 $G_i = G[R_i]$ 也包含最短路径 $\pi_G(p,s)$。现在考虑第二条性质。注意，$u$ 和 $v$ 是 $\pi_G(p,s)$ 上相邻的顶点，而这是一条位于 $G$ 和 $G_i$中的最短路径。因此，在 $H_i$中存在一条最短路径 $\pi_{H_i}(u,v)$；$H_i$是 $G_i$的一个 $\alpha$-伸展子图，其权重至多为 $\alpha$ 倍的 $\mathrm{d}_{G_i}(u,v)$。对于第三个不等式，回想一下，$x$ 是路径 $\pi_{H_i}(u,v)$ 上的一个顶点，然后我们使用第二项。

由于 $\pi_G(p, s)$ 至少包含 $q \cdot \frac{\alpha c_2'}{\epsilon c_1'}$ 个顶点，$G$ 的有界距离性质意味着，$\mathrm{d}_G(p, s) \ge c_1' \cdot q \frac{\alpha c_2'}{\epsilon c_1'} = q \cdot \frac{\alpha c_2'}{\epsilon}$，而 $\mathrm{d}_G(u, v) \le c_2'$。于是可得，$\mathrm{d}_{H_i}(u, v) \le \alpha \cdot \mathrm{d}_G(u, v) \le \alpha c_2' \le \frac{\epsilon}{q} \cdot \mathrm{d}_G(p, s)$。 □

请注意，上述断言中的最后一项表明，如果 $|\pi_G(p, s)|$ 足够大，则伸展子图 $H_i$ 中 $u$ 与 $v$ 之间的距离可以忽略不计。

### 3.3 质心集的构造

我们将构造三个中心点集合 $\mathbb{C}_{\text{landmark}}$，$\mathbb{C}_{\text{support}}$、和 $\mathbb{C}_{\text{net}}$，并将质心集定义为这三个集合的并集，即 $\mathbb{C} := \mathbb{C}_{\text{landmark}} \cup \mathbb{C}_{\text{support}} \cup \mathbb{C}_{\text{net}}$。这三个集合对应于我们之前讨论的三种情形。

**近邻点情形。** 首先，我们构造集合 $\mathbb{C}_{\text{net}}$。令 $p_1, p_2, \ldots, p_{n'}$ 为满足对所有 $p = p_i$，$\mathrm{d}_G(p, \mathcal{A}) < 1$ 的点，其中 $1 \le i \le n'$。令 $B_i = B(p_i, (10z/\epsilon) \cdot \mathrm{d}_G(p_i, \mathcal{A}))$ 为距 $p_i$ 不超过 $10z/\epsilon \cdot \mathrm{d}_G(p_i, \mathcal{A})$ 的点集。然后，我们通过令 $\mu = \Theta(\epsilon^3/z^3) \cdot \mathrm{d}_G(p_i, \mathcal{A})$，使用命题 3 计算集合 $B_i' \subseteq B_i$，并将每个这样的 $B_i'$ 中的所有点都加入 $\mathbb{C}_{\text{net}}$。

**短路径情形。** 接下来，我们构造集合 $\mathbb{C}_{\text{support}}$。考虑通过使用精度值 $\mu = \epsilon^2/z^2$ 应用命题4 得到的有界度支撑图 $H'' = (V'', E'')$。再考虑映射 $f\colon V(G) \to V(H'')$。对于每个点 $p \in X$，将 $v \in V(H'')$ 中距 $f(p)$ 在 $H''$ 中不超过 $\ell = \frac{\gamma \cdot z\alpha \cdot c_2'}{c_1' \cdot \epsilon}$ 跳的任意顶点加入 $\mathbb{C}_{\text{support}}$；也就是说，在 $H''$ 中存在一条连接 $v$ 和 $f(p)$ 且包含不超过 $\ell$ 条边的路径。这里，$\gamma \ge 1$ 是一个足够大的常数（取 $\gamma = 1600$ 即可）。

**长路径情形。** 首先，我们定义一个重要点集合，称为地标，它将帮助我们把 $V$ 中的点划分为等价类。具体而言，每个等价类都包含 $V$ 中的一个点子集，这些点相对于固定地标序列的距离向量近似相等。此外，我们将构造的质心集恰好从每个这样的等价类中选取一个代表点，从而得到一个规模较小的集合。为此，我们随后将定义到地标的取整距离 的概念，并用它来构造 $\mathbb{C}_{\text{landmark}}$，使得对于每个可能的距离元组，$\mathbb{C}_{\text{landmark}}$ 都包含一个具有该取整距离元组的点。

地标。令 $0 \le \mu \le \frac{1}{2}$ 足够小——稍后我们将把 $\mu$ 定义为 $\epsilon$ 和 $z$ 的函数。在以下讨论中，固定 $\mathcal{T}$ 中的一条根叶路径 $R_1, R_2, \ldots, R_t$，其中 $R_1$ 是根区域，而 $R_t$ 是叶区域。不失一般性，假设 $R_t \cap X \neq \emptyset$ —— 否则，我们就在包含 $X$ 中至少一个顶点的最低区域处剪去这条根叶路径。注意，$t = \mathcal{O}(\log |X|)$。我们回顾一下第 3.2 节中的一些记号。令 $G_i = G[R_i]$ 为与区域 $R_i$ 对应的导出子图，并令 $H_i$ 为 $G_i$ 的一个常数伸展生成器。对于 $1 \le i \le t-1$，令 $\mathcal{P}_i$ 为通过应用命题 5 来获得递归分解中 $R_i$ 的子区域时得到的最短路径分隔器集合，其中 $\mathcal{P}_i = \{P^{i,1}, P^{i,2}, \ldots, P^{i,b}\}$，且 $b = \mathcal{O}(1)$。令 $\mathcal{P} = \bigcup_{i=1}^{t-1} \mathcal{P}_i$。最后，我们定义一些记号简写：令 $d_i(\cdot,\cdot) = d_{G_i}(\cdot,\cdot)$，且 $\mathsf{d}(\cdot,\cdot) = d_G(\cdot,\cdot)$。对于一条路径 $P^{i,j}$ 和一个点 $q_1 \in X$，我们考虑两种定义舍入距离元组的方式，它们在不同情形下会派上用场。

**第一舍入距离元组**。考虑一个非叶区域 $R_i$ 和一条最短路径 $P^{i,j} \in \mathcal{P}_i$。对于每一对顶点 $q_1, q_2 \in X$，令 $D(i, q_1, q_2) := \mathsf{d}_i(q_1, q_2) + \mathsf{d}(q_2, \mathcal{A})$。在当前关于第一舍入距离元组的讨论结束之前，我们固定 $i$ 和点 $q_1, q_2$，并简记 $D := D(i, q_1, q_2)$。

令 $Q^{i,j}(q_1, q_2)$ 表示路径 $P^{i,j}$ 上满足 $\mathsf{d}_i(v, q_1) \le D/\mu^2$ 的顶点集合 $v$。任意将 $P^{i,j}$ 从一个端点指向另一个端点，令 $u$ 和 $v$ 分别表示属于 $Q^{i,j}$ 的第一个和最后一个顶点。注意，

$$\begin{aligned}
\mathsf{d}_{P^{i,j}}(u,v) &= \mathsf{d}_{H_i}(u,v) && \text{(Since } P^{i,j} \text{ is a shortest path in } H_i\text{)} \\
&\le \mathsf{d}_{H_i}(u,q_1) + \mathsf{d}_{H_i}(q_1,v) \\
&\le \alpha \cdot (\mathsf{d}_i(u,q_1) + \mathsf{d}_i(q_1,v)) \\
&\le \frac{2\alpha D}{\mu^2}.
\end{aligned}$$

现在，我们构造一个 $(\mu^2 \cdot D)$-网 $\mathcal{L}(i, j, q_1, q_2)$，它是 $Q^{i,j}(q_1, q_2)$ 的网。形式上，我们将 $u$ 和 $v$ 之间的子路径视为实区间 $[0, \beta]$，其中 $\beta \le \frac{2\alpha D}{\mu^2}$。我们每隔 $\mu^2 D$ 个距离单位做一个“标记”。现在回到这条由离散边序列构成的子路径。如果一个标记与路径上的某个顶点重合，就将该顶点加入集合 $\mathcal{L}(i, j, q^{i,j})$。否则，如果一个标记落在路径上两个相邻顶点之间，就将这两个顶点都加入集合 $\mathcal{L}(i, j, q_1, q_2)$。

**观察 1.** • $|\mathcal{L}(i, j, q_1, q_2)| = \mathcal{O}\left(\mu^{-4}\right)$（因为 $\alpha$ 是一个绝对常数），并且

- 对于任意满足 $d_{G_i}(q_1, x) \le D/\mu^2$ 的 $x \in V(P^{i,j})$，存在一个 $l \in \mathcal{L}(i, j, q_1, q_2)$，使得 $d_{G_i}(x, l) \le d_{H_i}(x, l) = d_{P^{i,j}}(x, l) \le \mu^2 D$。

令 $R'_t = \{s \in R_t : |\pi_G(x, s)| > \ell$ 对所有 $x \in X\}$ 均成立。也就是说，$R'_t$ 是 $R'_t$ 中满足以下条件的点的子集：沿着 $G$ 中的最短路径，$s$ 到 $X$ 中的每个点的跳数距离都严格大于 $\ell$。

对于路径 $P^{i,j} \in \mathcal{P}_i$，对于每个 $q_1, q_2 \in X$ 以及任意点 $s \in R'_t \subseteq R_t$，我们考虑以下舍入距离元组 $\tilde{\mathsf{d}}^1_{i,j}(s, (q_1, q_2))$，它定义为按固定顺序列出以下各项所形成的元组：

- For each $l \in \mathcal{L}(i, j, q_1, q_2)$: $\tilde{d}^1_{i,j}(s, l) = \min\left\{\left\lceil \frac{\mathsf{d}_i(s,l)}{\mu^2 D} \right\rceil, \frac{3D}{\mu^2}\right\}$.
- $\tilde{d}^1_{i,j}(s, q_1) = \left\{\left\lceil \frac{\mathsf{d}_i(s,q_1)}{\mu D} \right\rceil, \frac{3D}{\mu}\right\}$.

- For each $x \in X \cap R_t$: $\tilde{d}^1_{i,j}(s,x) = \left\{ \left\lceil \frac{\mathsf{d}_i(x,s)}{\mu \mathsf{d}(x,\mathcal{A})} \right\rfloor, \frac{1}{\mu}\mathsf{d}(x,\mathcal{A}) \right\}$.

这里，我们使用记号 $\lceil {p \atop q} \rfloor$ 表示最接近 $\frac{p}{q}$ 的 $q$ 的整数倍。

对于叶区域 $R_t$，不存在对应的最短路径分隔器。因此，我们首先考虑第一舍入距离元组的另一种定义。令 $\mathcal{L}(t) = R_t \cap X$。对于每个 $p \in \mathcal{L}(t)$，令$\tilde{d}_t(s, x) = \left\{ \left\lceil \frac{\mathsf{d}(x, s)}{\mu \mathsf{d}(x, \mathcal{A})} \right\rfloor, \frac{1}{\mu}\mathsf{d}(x, \mathcal{A}) \right\}$。令 $\tilde{\mathsf{d}}^1_t(s, t)$ 表示该元组。

**第二舍入距离元组**。再次固定一条路径 $P^{i,j} \in \mathcal{P}_i$、一个点 $q_1 \in X$ 以及一对 $q_3$，$q_4 \in R_i \cup \{\perp\}$，其中 $\perp$ 表示未定义。此外，我们假设 $q_3$ 和 $q_4$ 不会同时未定义。于是，对于一个点 $s \in R'_t \subseteq R_t$，我们定义第二舍入距离元组 $\tilde{d}^2_{i,j}(s, (q_1, q_3, q_4))$，定义如下。

- $\tilde{d}^2_{i,j}(s, (q_1, q_3, q_4))$ =当且仅当

$$\frac{1}{\mu} \cdot (\mathsf{d}_i(q_1, q_4) + \mathsf{d}(q_4, \mathcal{A})) < \mathsf{d}_i(q_1, s) < \mu \cdot (\mathsf{d}_i(q_1, q_3) + \mathsf{d}(q_3, \mathcal{A}))$$

这里，如果 $q_4 = \perp$，则定义中省略第一个不等式；如果 $q_3 = \perp$，则定义中省略第二个不等式。注意，由于我们的假设，$q_3$ 和 $q_4$ 中至少有一个不是未定义的，因此该定义不可能空泛地成立。

- $\tilde{d}^2_{i,j}(s, (q_1, q_3, q_4))$ =否则为 false；也就是说，当 $q_4 \neq \perp$ 且第一个不等式不成立时，或者当 $q_3 \neq \perp$ 且第二个不等式不成立时。

**构造** $\mathbb{C}_{\mathbf{landmark}}$。对于每条根叶路径 $R_1, \ldots, R_t$，对于每个 $P^{i,j} \in \mathcal{P}$，并且对于两种距离舍入方式中的每一种（即 $\tilde{d}^1_{i,j}(\cdot,\cdot)$ 或 $\tilde{d}^2_{i,j}(\cdot,\cdot)$），我们考虑相应舍入元组中的各项可能取到的所有值。我们还考虑舍入距离元组 $\tilde{\mathsf{d}}^1_t(\cdot)$。对于每种可能的舍入类型选择及各项取值，我们通过按固定顺序拼接所有如此构造的元组，得到一个较大的舍入距离元组 D。也就是说，D 由相应的舍入距离元组构成（即，$\tilde{d}^1_{\cdot}(\cdot)$ 或 $\tilde{d}^2_{\cdot}(\cdot)$ 中各项可能取到的势数值或真/假值）。如果存在一个点 $s \in R'_t \subseteq R_t$，能够实现所有相应的各项取值（即舍入距离值和/或真/假值），如 D 所指定的那样，则将其加入 $\mathbb{C}_{\text{landmark}}$。

## 3.4 近似解的构造

现在我们证明，对于任意解 $\mathcal{S}$，$\mathcal{S}$ 中的每个中心都可以由$\mathbb{C}$中的某个质心近似。考虑任意中心 $s \in \mathcal{S}$，并令 $X_s \subseteq X$ 为这样一个点子集：其中每个点在 $S$ 中的最近中心是$s$（任意打破平局）。首先，我们展示如何找到 $\tilde{s} \in \mathbb{C}$，使得对于任意满足代价$(p, s) \leq \left(\frac{10z}{\epsilon}\right)^z \cdot$代价$(p, \mathcal{A})$的点 $p \in X_s$，代价$(p, \tilde{s}) \leq$代价$(p, s) + 3\epsilon$（代价$(p, s) +$代价$(p, \mathcal{A})$。

现在，我们展示如何为任意 $s \in \mathcal{S}$ 找到一个替换中心，并将其记为 $\rho(s)$。我们考虑不同的情形来找到 $\rho(s)$。以下情形，即替换规则，按如下方式应用。我们按照替换规则呈现的顺序依次处理。对于某条替换规则，我们遍历所有中心 $s \in \mathcal{S}$，对于这些中心，使用先前规则尚未找到替换点，并检查当前规则是否适用于 $s$；如果适用，则按照该规则的规定将 $\rho(s)$定义为相应的替换点。因此，在遍历结束时，剩下的是这样一些中心 $s \in \mathcal{S}$，先前的规则均不适用于它们。然后，我们继续处理下一条规则，并以类似方式进行。

**使用** $\mathbb{C}_{\mathbf{net}}$ **进行替代**。

**情形 1.** 假设存在一个点 $p \in X_s$，使得 $\mathsf{d}_G(p, \mathcal{A}) < 1$。
设 $X'_s := \{q \in X_s : \mathsf{d}_G(q, \mathcal{A}) < 1\}$。注意到 $p \in X'_s$，这意味着 $X'_s \neq \emptyset$。令$p_i := \arg\min_{\{q \in X'_s\}} \mathsf{d}_G(q, \mathcal{A}) + \mathsf{d}_G(q, s)$。注意到 $s \in B_i = B(p_i, (10z/\epsilon) \cdot \mathsf{d}_G(p_i, \mathcal{A}))$。令 $\tilde{s}$ 为 $(\epsilon^3/z^3) \cdot \mathsf{d}_G(p_i, \mathcal{A})$-网 $B'_i$ 中距离 $s$ 最近的点，该网是为 $B_i$构造的。注意到这样的 $\tilde{s}$ 已被添加到 $\mathbb{C}_{\text{net}}$。在这种情况下，我们将 $\tilde{s} \in \mathbb{C}_{\text{net}} \subseteq \mathbb{C}$ 添加到 $\mathcal{S}$，并令 $\rho(s) = \tilde{s} \in \mathbb{C}_{\text{net}}$。将 $\mathcal{S}_{\text{net}}$ 定义为 $\mathcal{S}$ 的子集，其中包含所有以这种方式找到替代点的中心 $s$。

**情形 2.** 考虑以下情形。

1. 假设对于所有点 $p \in X_s$，均有 $\mathsf{d}_G(p, \mathcal{A}) \geq 1$，因此，在前一种情形中，我们没有找到 $\rho(s)$，

2. $f(s) \in V(H'')$ 被添加到了 $\mathbb{C}_{\text{support}} \subseteq \mathbb{C}$，并且

3. 存在一个点 $q \in X_{s_q}$，满足 $s \neq s_q \in \mathcal{S}_{\text{net}}$，且 $\tilde{s}_q = \rho(s_q) \in \mathbb{C}_{\text{net}}$ 具有以下性质：(1) $\mathsf{d}_G(q, s_q) \leq \epsilon/z$，(2) $\mathsf{d}_G(q, \tilde{s}_q) > \mathsf{d}_G(q, f(s))$。

在这种情形下，我们令 $\rho(s) = \tilde{s}_q$，如上文所定义（如果 $q$ 有多个选项，从而 $s_q$ 也有多个选项，我们可以任意选择）。我们将 $\mathcal{S}_{\text{net-sub}}$ 定义为 $\mathcal{S}$ 的子集，其中包含所有通过这种情形找到其替代点的中心 $s$。注意，对于 $\mathcal{S}_{\text{net}}$ 中的中心 $s$ 以及 $\mathcal{S}_{\text{net-sub}}$，$\rho(s)$ 属于 $\mathbb{C}_{\text{net}}$；但区别在于，对于中心 $s \in \mathcal{S}_{\text{net-sub}}$，我们利用 $\mathcal{S}$ 中的一个中心，该中心已被找到并用作另一个中心 $s_q \in \mathcal{S}_{\text{net}}$ 的替代点。

**使用替代点 $\mathbb{C}_{\text{support}}$。** 考虑前一种情况中的前两个条件成立，但不存在满足第三个条件的点 $p$ 的情形。也就是说，假设 $s$ 满足以下条件：

1. 假设对于所有点 $q \in X_s$，均有 $\mathsf{d}_G(q, \mathcal{A}) \geq 1$，

2. $f(s) \in V(H'')$ 已添加到 $\mathbb{C}_{\text{support}} \subseteq \mathbb{C}$，并且

3. 不存在这样的点 $p \in X_{s_p}$，使得 $s \neq s_p \in \mathcal{S}$，且 $s'_p = \rho(s_p) \in \mathbb{C}_{\text{net}}$ 满足以下性质：(1) $\mathsf{d}_G(p, s_p) \leq \epsilon/z$，以及 (2) $\mathsf{d}_G(p, s'_p) > \mathsf{d}_G(p, f(s))$.

然后，我们将 $f(s)$ 添加到 $\mathcal{S}$，并设置 $\rho(s) = f(s)$。令 $\mathcal{S}_{\text{support}} \subseteq \mathcal{S}$ 表示满足$\rho(s) \in \mathbb{C}_{\text{support}}$的中心集合 $s$。

我们令 $\mathcal{S}_{\text{landmark}} := \mathcal{S} \setminus (\mathcal{S}_{\text{net}} \cup \mathcal{S}_{\text{net-sub}} \cup \mathcal{S}_{\text{support}})$ 表示这样一个中心子集：其替代点尚未在前一段所述的 $\mathbb{C}_{\text{net}}$ 或 $\mathbb{C}_{\text{support}}$ 中找到。这里，该记号预示着，对于这类中心 $s \in \mathcal{S}_{\text{landmark}}$，我们将利用 $\mathbb{C}_{\text{landmark}}$ 找到 $\rho(s)$。在此之前，我们先注意到以下观察。

**观察 2。** 对于任意 $s \in \mathcal{S}_{landmark}$ 以及任意点 $p \in X$，$|\pi_G(p, s)| > \ell$。

证明。反设对于某些 $s \in \mathcal{S}_{\text{landmark}}$ 和 $p \in X$ $p$，$u_1, \ldots, u_t$，$s$ 是一条路径 $\pi_G(p, s)$。那么，路径 $f(p)$ $f(u_1) \ldots, f(u_t)$ $f(s)$ 位于 $H''$ 中。因此，$f(s)$ 至多距离 $f(p)$ $\ell$ 跳，因而根据我们的构造，$f(s)$ 应已被加入 $\mathbb{C}$。但这构成矛盾，因此 $|\pi_G(p, s)| > \ell$。 □

**使用 $\mathbb{C}_{\text{landmark}}$。** 现在，我们说明如何使用 $\mathbb{C}_{\text{landmark}}$为 $s \in \mathcal{S}_{\text{landmark}}$寻找一个替代点。为此，我们利用之前构造的递归分解树 $\mathcal{T}$和集合 $\mathbb{C}_{\text{landmark}}$。在下面的讨论中，固定中心 $s \in \mathcal{S}_{\text{landmark}}$，并令 $R_t$ 为包含 $s$的 $\mathcal{T}$的叶节点。令 $\Pi = (R_1, R_2, \ldots\ldots, R_t)$。注意，观察 2 意味着$s \in R'_t \subseteq R_t$。

考虑一个非叶区域 $R_i$和一条路径 $P^{i,j} \in \mathcal{P}_i$。同样，我们使用简写 $\mathsf{d}_i = \mathsf{d}_{G_i}$和 $\mathsf{d} = \mathsf{d}_G$。回想一下，$G_i = G[R_i]$。令 $q_1^i = \arg\min_{p \in R_i} \mathsf{d}_i(p, s)$。现在，我们考虑两种情况。

1. 存在某个 $q_2^i \in R_i$，使得 $\mu \mathrm{d}_i(q_1^i, s) \leq \mathrm{d}_i(q_1^i, q_2^i) + \mathrm{d}(q_2^i, \mathcal{A}) \leq \frac{1}{\mu}\mathrm{d}_i(q_2^i, s)$。在这种情况下，我们选择舍入 $\tilde{\mathrm{d}}_{i,j}^1(s, (q_1^i, q_2^i))$。

2. 否则，我们按如下方式进行。（i）如果存在一个点 $q \in R_i$，使得 (i) $\mathrm{d}_i(q_1^i, q) + \mathrm{d}(q, \mathcal{A}) > \frac{1}{\mu}\mathrm{d}_i(q_1^i, s)$。那么，令 $q_3^i = \arg\min_{q \in R_i} \mathrm{d}_i(q_1^i, q) + \mathrm{d}(q, \mathcal{A})$。如果不存在这样的 $q \in R_i$，则令 $q_3^i = \perp$，即未定义。（ii）如果存在一个点 $q \in R_i$，使得 (i) $\mathrm{d}_i(q_1^i, q) + \mathrm{d}(q, \mathcal{A}) < \mu \mathrm{d}_i(q_1^i, s)$。那么，令 $q_4^i = \arg\max_{q \in R_i} \mathrm{d}_i(q_1^i, q) + \mathrm{d}(q, \mathcal{A})$。如果不存在这样的 $q \in R_i$，则令 $q_4^i = \perp$，即未定义。注意，$q_3^i$ 和 $q_4^i$ 中至少有一个不是未定义的。在这种情况下，我们选择舍入 $\tilde{\mathrm{d}}_{i,j}^2(s, (q_1^i, q_3^i, q_4^i))$。

通过拼接所有路径 $P^{i,j} \in \mathcal{P}$ 对应的舍入距离元组 $\tilde{d}_{i,j}^1(s, \cdot)$ 或 $\tilde{d}_{i,j}^2(s, \cdot)$（如上文所定义），我们得到舍入距离元组 $\tilde{\mathrm{D}}(s)$。由于 $s \in R_t'$ and 具有舍入距离元组 $\mathrm{D}(s)$，我们可以得出结论：$\mathbb{C}_{\text{landmark}} \cap R_t'$ 必定包含一个点，比如 $\tilde{s}$，它也具有舍入距离元组 $\mathrm{D}(s)$。我们令 $\rho(s) = \tilde{s}$。

至此，集合 $\mathcal{S} \subseteq \mathbb{C}$的构造完成。在下一小节中，我们将分析该解的性质。

### 3.5 误差分析

本节将说明如何界定任意相关点 $p \in X_s$ 的误差。总体目标是证明，对于任意满足 $\mathrm{d}_G(p, \mathcal{S}) \leq \frac{\gamma' z}{\epsilon} \cdot \mathrm{d}_G(p, \mathcal{A})$ 或 $\mathrm{d}_G(p, \tilde{\mathcal{S}}) \leq \frac{\gamma' z}{\epsilon} \cdot \mathrm{d}_G(p, \mathcal{A})$ 的点 $p \in X$，都有 $|\mathrm{cost}(p, \mathcal{S}) - \mathrm{cost}(p, \mathcal{S})| \leq \mathcal{O}(\epsilon) \cdot (\mathrm{cost}(p, \mathcal{S}) + \mathrm{cost}(p, \mathcal{A}))$。由于集合 $\mathcal{S}$ 是通过仔细考察不同情形构造而成的，这一论断的证明也基于穷尽式案例分析，并组织为若干引理。在正式陈述并证明这些引理之前，我们先概述论证的整体思路。

**正向**。对于点 $p \in X$，其中 $s$ 是其在 $\mathcal{S}$ 中的最近中心，我们根据 $s$ 是否属于 $\mathcal{S}_{\text{net}}$、$\mathcal{S}_{\text{support}}$ 或 $\mathcal{S}_{\text{landmark}}$ 来考虑不同情形。这分别对应于断言 3、断言 4、断言 5 和引理 2。在此，我们证明，对于某个 $\mu \geq 0$，要么 $\mathrm{d}_G(p, s) > \frac{\mu z}{\epsilon} \cdot \mathrm{d}_G(p, \mathcal{A})$，要么 $\mathrm{d}_G(p, \rho(s)) \leq (1 + \epsilon/z) \cdot \mathrm{d}_G(p, s) + \mathcal{O}(\epsilon/z) \cdot \overline{\mathrm{d}}_G(p, \mathcal{A})$。这构成了证明的正向部分。

反向。对于点 $p \in X$，其中 $s \in \mathcal{S}$ 和 $\tilde{s} \in \mathcal{S}$ 分别是其在两个集合中的最近中心。如果 $\tilde{s} = \rho(s)$，那么可以相对直接地证明：若 $\mathrm{d}_G(p, \tilde{s}) \leq \frac{\mu z}{\epsilon} \cdot \mathrm{d}_G(p, \mathcal{A})$，则对于某个 $\mu \geq 0$，有 $\mathrm{d}_G(p, s) \leq (1 + \epsilon/z) \cdot \mathrm{d}_G(p, \tilde{s}) + \mathcal{O}(\epsilon/z) \cdot \mathrm{d}_G(p, \mathcal{A})$。在另一种情形下，可能有 $\tilde{s} = \rho(s')$，其中 $s' \neq s$。也就是说，$\tilde{s} \in \mathcal{S}$ 被找到作为中心 $s'$ 的替代点，而该中心并不是点 $p$ 在 $\mathcal{S}$ 中的最近中心。这里需要注意的微妙可能性是 $\mathrm{d}_G(p, \tilde{s}) \ll \mathrm{d}_G(p, s)$，在这种情况下，我们无法指望证明所需的界限。为论证这种情况不会发生，我们仔细检查用于找到 $\rho(s)$ 和 $\rho(s')$ 的不同情形；它们分别是 $s$ 和 $s'$ 的替代点。反向部分由断言 6 、断言 7 和引理 2 构成。

最后，我们结合两个方向，并证明引理 4 中解的误差满足所需的界。

#### 正向

*断言 3. 考虑一个点 $p \in X_s$，满足 $s \in \mathcal{S}_{net}$，且 $d_G(p, s) \leq \frac{10z}{\epsilon} d_G(p, \mathcal{A})$。那么，对于 $\rho(s) = \tilde{s} \in \mathbb{C}_{net}$，有 $|d_G(p, \tilde{s}) - d_G(p, s)| \leq (\epsilon/z) \cdot (d_G(p, s) + d_G(p, \mathcal{A}))$。*

证明。由于 $s \in \mathcal{S}_{\text{net}}$，集合 $X'_s = \{q \in X_s : \mathsf{d}_G(q, \mathcal{A}) < 1\}$ 非空。然后，我们选择 $p_i = \arg\min_{q \in X'_s} \mathsf{d}_G(q, \mathcal{A}) + \mathsf{d}_G(q, s)$。注意到 $s \in B_i = B(p_i, (10z/\epsilon) \cdot \mathsf{d}_G(p_i, \mathcal{A}))$，并将一个从 $s$ 最近的点 $\tilde{s}$ 添加到 $\mathbb{C}_{\text{net}}$ 的 $(\epsilon^3/z^3) \cdot \mathsf{d}_G(p_i, \mathcal{A})$-网中。现在，$\mathsf{d}_G(s, \tilde{s}) \leq (\epsilon^3/z^3) \cdot \mathsf{d}_G(p_i, \mathcal{A}) < (\epsilon/z)$。

现在有两种可能。如果 $\mathsf{d}_G(p, \mathcal{A}) \geq 1$，那么 $|\mathsf{d}_G(p, \tilde{s}) - \mathsf{d}_G(p, s)| \leq \epsilon/z \leq (\epsilon/z) \cdot \mathsf{d}_G(p, \mathcal{A})$。

否则，如果 $\mathsf{d}_G(p, \mathcal{A}) < 1$，那么 $p \in X'_s$。因此，$\mathsf{d}_G(p_i, s) + \mathsf{d}_G(p_i, \mathcal{A}) \leq \mathsf{d}_G(p, s) + \mathsf{d}_G(p, \mathcal{A})$。因此，$|\mathsf{d}_G(p, \tilde{s}) - \mathsf{d}_G(p, s)| \leq \frac{\epsilon}{z} \cdot \mathsf{d}_G(p_i, \mathcal{A}) \leq \frac{\epsilon}{z}(\mathsf{d}_G(p, s) + \mathsf{d}_G(p, \mathcal{A}))$。 □

*断言 4。考虑某个点 $p \in X_s$ for some $s \in \mathcal{S}_{\text{net-sub}}$，使得 $d_G(p, s) \leq \frac{10z}{\epsilon} \cdot d_G(p, \mathcal{A})$。那么，$\tilde{d}_G(p, \mathcal{S}) \leq d_G(p, s) + \mathcal{O}(\epsilon/z) \cdot d_G(p, \mathcal{A})$。*

证明。首先，我们断言 $\mathsf{d}_G(p, \mathcal{A}) \geq 1$。反设 $\mathsf{d}_G(p, \mathcal{A}) < 1$。那么，我们将能够使用情形 1 为 $s$ in $\mathbb{C}_{\text{net}}$ 找到一个替代点，即 $s \in \mathcal{S}_{\text{net}}$。然而，由于$s \in \mathcal{S}_{\text{net-sub}}$，情形 1 不适用，这就产生了矛盾。

由于 $s \in \mathcal{S}_{\text{net-sub}}$，存在一个点 $f(s) \in \mathbb{C}_{\text{support}}$，使得 $\mathsf{d}_G(s, f(s)) = \mathcal{O}(\epsilon^2/z^2)$。此外，对于某个 $s_q \in \mathcal{S}_{\text{net}}$，存在一个点 $q \in X_{s_q}$，使得 $\tilde{s}_q = \rho(s_q)$ 满足以下性质：(1) $\mathsf{d}_G(q, s_q) \leq \epsilon/z$；(2) $\mathsf{d}_G(q, \tilde{s}_q) > \mathsf{d}_G(q, f(s))$。于是，

$$
\begin{aligned}
\mathsf{d}_G(p, \tilde{s}_q) &\leq \mathsf{d}_G(p, s) + \mathsf{d}_G(s, f(s)) + \mathsf{d}_G(f(s), q) + \mathsf{d}_G(q, \tilde{s}_q) \\
&\leq \mathsf{d}_G(p, s) + \mathcal{O}(\epsilon^2/z^2) + 2 \cdot \mathsf{d}_G(q, \tilde{s}_q) && (\because\ \mathsf{d}_G(q, f(s)) \leq \mathsf{d}_G(q, \tilde{s}_q)) \\
&\leq \mathsf{d}_G(p, s) + \mathcal{O}(\epsilon^2/z^2) + 2 \cdot (\mathsf{d}_G(q, s_q) + \mathsf{d}_G(s_q, \tilde{s}_q)) \\
&\leq \mathsf{d}_G(p, s) + \mathcal{O}(\epsilon^2/z^2) + 2 \cdot \epsilon/z && (\because\ d(q, s_q) \leq \epsilon/z \text{ and } \mathsf{d}_G(s_q, s'_q) \leq \epsilon^3/z^3) \\
&\leq \mathsf{d}_G(p, s) + \mathcal{O}(\epsilon/z) \cdot \mathsf{d}_G(p, \mathcal{A}) && (\because\ \mathsf{d}_G(p, \mathcal{A}) \geq 1.)
\end{aligned}
$$

□

*断言 5。考虑某个 $s \in \mathcal{S}_{\text{support}}$ 的一个点 $p \in X_s$，使得 $d_G(p, s) \leq \frac{10z}{\epsilon} \cdot d_G(p, \mathcal{A})$。那么，$\tilde{d}_G(p, \mathcal{S}) \leq d_G(p, s) + \mathcal{O}(\epsilon/z) \cdot d_G(p, \mathcal{A})$。*

证明。仿照断言 4 的证明，我们首先注意到 $\mathsf{d}_G(p, \mathcal{A}) \geq 1$，否则 $s$ 会属于 $\mathcal{S}_{\text{net}}$，这与假设矛盾。由于 $s \in \mathcal{S}_{\text{support}}$, $f(s)$ 已被添加到 $\mathbb{C}_{\text{support}}$，且我们令 $\rho(s) = f(s)$。因此，$G$ 在 $p$ 与 $f(s)$ 之间存在一条长度恰为 $|\pi_G(p, s)| + 1$ 的路径，其权重为 $\mathsf{d}_G(p, s) + \mathcal{O}(\epsilon^2/z^2)$；具体而言，该路径由边 $\{s, f(s)\}$ 与路径 $\pi_G(p, s)$ 拼接而成。因此，

$$\mathsf{d}_G(p, f(s)) \leq \mathsf{d}_G(p, s) + \mathcal{O}(\epsilon^2/z^2) \leq \mathsf{d}_G(p, s) + \mathcal{O}(\epsilon^2/z^2) \cdot \mathsf{d}_G(p, \mathcal{A})$$

最后一个不等式成立，这是因为 $\mathsf{d}_G(p, \mathcal{A}) \geq 1$。 □

现在，考虑一个点 $p \in X_s$，满足 $\mathsf{d}_G(p, s) \leq \frac{10z}{\epsilon} \mathsf{d}_G(p, \mathcal{A})$。现在，若断言 3、断言 4 和断言 5 均不适用，则可得 $s \notin (\mathcal{S}_{\text{net}} \cup \mathcal{S}_{\text{net-sub}} \cup \mathcal{S}_{\text{support}})$。也就是说，$s \in \mathcal{S}_{\text{landmark}}$。由观察 2 可知，对于所有点 $q \in X$，路径 $\pi_G(q, s)$ 的跳数长度，即 $|\pi_G(q, s)|$，严格大于 $\ell$。

在下面的引理中，我们证明，从集合 $\mathbb{C}_{\text{landmark}}$ 中为 $s$ 找到的替代点 $\rho(s)$ 到 $p$ 的距离与 $s$ 的距离近似相同。事实上，下面的引理在以下两个方面更强。首先，它不要求 $s$ 是 $s$ 中距离 $p$ 最近的中心——我们证明，只要相应的距离受 $\mathcal{O}(\epsilon/z) \cdot \mathsf{d}_G(p, \mathcal{A})$ 限制，该不等式就对 任意 $s'$ 及其替代点 $\tilde{s}$ 成立。其次，该引理证明了两个方向的不等式。这两个性质随后都会很有用。

**引理 2**。 考虑一个中心 $s \in \mathcal{S}'$，并令 $\tilde{s} = \rho(s) \in \mathcal{S}$ 为从 $\mathbb{C}_{landmark}$ 中找到的其替代点。那么，存在一个常数 $\gamma$，使得对于任意点 $p \in X$，以下结论成立：

- *Either* $\mathsf{d}_G(p,s) \geq \frac{\gamma z \mathsf{d}_G(p,\mathcal{A})}{\epsilon}$, *or* $\mathsf{d}_G(p,\tilde{s}) \leq (1+\frac{\epsilon}{z})\mathsf{d}_G(p,s) + \frac{\epsilon}{z} \cdot \mathsf{d}_G(p,\mathcal{A})$, *and*
- *Either* $\mathsf{d}_G(p,\tilde{s}) \geq \frac{\gamma z \mathsf{d}_G(p,\mathcal{A})}{\epsilon}$, *or* $\mathsf{d}_G(p,s) \leq (1+\frac{\epsilon}{z})\mathsf{d}_G(p,\tilde{s}) + \frac{\epsilon}{z} \cdot \mathsf{d}_G(p,\mathcal{A})$.

证明。 令 $s_1 \in \{s, \tilde{s}\}$，并将 $s_2$ 作为另一种选择。令 $R_t$ 为包含 $s$ 和$\tilde{s}$的 $\mathcal{T}$的叶节点，并注意，$\mathcal{T}$中根叶路径 $\Pi = R_1, \ldots, R_t$上的每个节点都同时包含 $s$和 $s'$。观察可知 $s, \tilde{s} \in R'_t$。回顾一下，$\mu$ 是 $\mathbb{C}_{\text{landmark}}$的构造中定义的参数。之后我们将 $\mu$ 的值设为 $\frac{\gamma z}{\epsilon}$。在余下的讨论中，我们固定点 $p \in X$，并考察不同情形。

叶节点情形： $p \in R_t \cap X$。由于 $p \in \mathcal{L}(t) = R_t \cap X$，元组 $\tilde{d}^1_t(\cdot)$中的舍入意味着：要么 $\mathsf{d}_G(p,s) > \frac{1}{\mu} \cdot \mathsf{d}_G(p,\mathcal{A})$，在这种情况下，$\mathsf{d}_G(p,\tilde{s}) > \frac{1}{\mu} \cdot \mathsf{d}_G(p,\mathcal{A})$，因为 $s$和 $\tilde{s}$具有相同的舍入距离元组；要么 $|\mathsf{d}_G(p,s) - \mathsf{d}_G(p,\tilde{s})| \leq \mu \mathsf{d}_G(p,\mathcal{A})$。

**内部节点情形：** $p \notin R_t$。在此情形中，我们考虑区域 $R_i$ 沿着 $\Pi$，使得某条路径 $P^{i,j} \in \mathcal{P}_i$ 将 $p$ 与 $s_1$分隔开，并且 $x \in P^{i,j}$ 是一个位于 $u$ 和$v$之间的分隔顶点（定义见第 3.2 节）。在证明的其余部分，我们使用简写 $\mathsf{d}(\cdot,\cdot) := \mathsf{d}_G(\cdot,\cdot)$，以及 $\mathsf{d}_i(\cdot,\cdot) = \mathsf{d}_{G_i}(\cdot,\cdot)$。令 $q^i_1 = \arg\min_{q \in R_i} \mathsf{d}_i(q,s)$。下面的不等式列出了这些定义以及断言 2 的一些推论。

$$\mathsf{d}_i(p,s_1) = \mathsf{d}(p,s_1) = \mathsf{d}(p,u) + \mathsf{d}(u,v) + \mathsf{d}(v,s_1) \tag{1}$$
$$\mathsf{d}_i(u,v) \leq \mathsf{d}_{H_i}(u,v) \leq \alpha \cdot \mathsf{d}_i(u,v) = \alpha \cdot \mathsf{d}(u,v) \tag{2}$$
$$\mathsf{d}_{H_i}(u,x) + \mathsf{d}_{H_i}(x,v) = \mathsf{d}_{H_i}(u,v) \leq \alpha \cdot \mathsf{d}_i(u,v) \leq \alpha \cdot \mathsf{d}(u,v) \tag{3}$$
$$\mathsf{d}_{H_i}(u,v) \leq \mu \cdot \mathsf{d}_G(p,s) \tag{4}$$

在此情形中，我们根据为 $P^{i,j}$选择的舍入元组类型，考虑不同情形；在使用舍入距离元组为 $s$ 寻找替代点时亦如此。

**情形 1**。假设存在一个 $q \in R_i$ 满足

$$\mu \cdot \mathsf{d}_i(q^i_1, s) \leq \mathsf{d}_i(q^i_1, q) + \mathsf{d}(q, \mathcal{A}) \leq \frac{1}{\mu} \cdot \mathsf{d}_i(q^i_1, s) \tag{5}$$

注意，情形假设意味着，在定义舍入距离元组 $\mathsf{D}(s)$ 时，对于某个满足 (5) 的 $q^i_2$，我们选择了相对于路径 $P^{i,j}$的第一个舍入元组 $\tilde{d}^1_{i,j}(s, (q^i_1, q^i_2))$。为分析情形 1 及其子情形，我们使用 $D := D(i, q^i_1, q^i_2) = \mathsf{d}_i(q^i_1, q^i_2) + \mathsf{d}(q^i_2, \mathcal{A})$

**情况 1A**。 $\mathsf{d}_i(x, q^i_1) \leq \frac{D}{\mu^2}$。 由于 $x \in P^{i,j}$ 且 $\mathsf{d}_i(x, q^i_1) \leq \frac{D}{\mu^2}$，这意味着 $x \in Q^{i,j}(q^i_1, q^i_2)$。因此，利用观察 1，存在某个 $l \in \mathcal{L}(i,j,q^i_1,q^i_2)$，使得 $\mathsf{d}_i(x,l) \leq \mu^2 D$。现在考虑，

$$\begin{aligned} \mathsf{d}_i(s,l) &\leq \mathsf{d}_i(l,x) + \mathsf{d}_i(x,q^i_1) + \mathsf{d}_i(q^i_1,s) \\ &\leq \mu^2 D + \frac{D}{\mu^2} + \frac{D}{\mu} \qquad (\mathsf{d}_i(q^i_1,s) \leq \tfrac{D^2}{\mu} \text{ by case assumption}) \\ \implies \mathsf{d}_i(s,l) &\leq \frac{3D^2}{\mu} \end{aligned} \tag{6}$$

因此，$\tilde{d}^1_{i,j}(l, s) = \tilde{d}^1_{i,j}(l, \tilde{s}) = \left\{ \left\lceil \mathsf{d}_i(s,l) / \mu^2 D \right\rceil \right\}_{\cdot,\cdot}$，这意味着 $|\mathsf{d}_i(s_1, l) - \mathsf{d}_i(s_2, l)| \leq \mu^2 D$。现在，

我们考虑:

$$\begin{aligned}
\mathsf{d}_i(p,s_2) &\le \mathsf{d}_i(p,l)+\mathsf{d}_i(l,s_2)\\
&\le \mathsf{d}_i(p,l)+\mathsf{d}_i(l,s_1)+\mu^2 D && \text{(From above)}\\
&\le \mathsf{d}_i(p,x)+\mathsf{d}_i(x,l)+\mathsf{d}_i(x,l)+\mathsf{d}_i(x,s_1)+\mu^2 D\\
&\le \mathsf{d}_i(p,x)+\mathsf{d}_i(x,s_1)+3\mu^2 D && \text{(Since } \mathsf{d}_i(x,l)\le\mu^2 D)\\
&\le \mathsf{d}_i(p,u)+\mathsf{d}_i(u,x)+\mathsf{d}_i(x,v)+\mathsf{d}_i(v,s_1)+3\mu^2 D\\
&\le \mathsf{d}(p,u)+\mathsf{d}(v,s_1)+\mathsf{d}_{H_i}(u,x)+\mathsf{d}_{H_i}(x,v)+3\mu^2 D && \text{(Using (1))}\\
&\le \mathsf{d}(p,s_1)+\mu\cdot\mathsf{d}(p,s_1)+3\mu^2 D && \text{(Using (4))}\\
&\le (1+\mu)\cdot\mathsf{d}(p,s_1)+3\mu\mathsf{d}_i(p,s) && \text{(Using case assumption: (5))}
\end{aligned}$$

代入 $s_1=s$ 和 $s_2=\tilde{s}$，得到:

$$\begin{aligned}
\mathsf{d}(p,\tilde{s})\le \mathsf{d}_i(p,\tilde{s}) &\le (1+\mu)\cdot\mathsf{d}(p,s)+3\mu\cdot\mathsf{d}_i(p,s)\\
&\qquad\text{(First inequality follows since } G_i \text{ is a subgraph of } G)\\
&\le (1+4\mu)\cdot\mathsf{d}(p,s) \qquad (7)
\end{aligned}$$

代入 $s_1=s$ 和 $s_1=\tilde{s}$，得到

$$\begin{aligned}
\mathsf{d}_i(p,s) &\le (1+\epsilon)\mathsf{d}(p,\tilde{s})+3\mu\mathsf{d}_i(p,s)\\
(1-3\mu)\cdot\mathsf{d}(p,s)\le(1-3\mu)\cdot\mathsf{d}_i(p,s) &\le (1+\mu)\mathsf{d}(p,\tilde{s})\\
&\qquad\text{(First inequality follows since } G_i \text{ is a subgraph of } G)\\
\implies \mathsf{d}(p,s) &\le \frac{1+\mu}{1-3\mu}\cdot\mathsf{d}(p,\tilde{s})\le(1+13\mu)\cdot\mathsf{d}(p,\tilde{s}) \qquad (8)
\end{aligned}$$

**情况 1B.** $\mathsf{d}_i(x,q_1^i)>\frac{D}{\mu^{2\circ}}$ 首先，由 (5) 可知 $\mathsf{d}_i(q_1^i,s)\le\frac{1}{\mu}D_\circ$ 因此，根据舍入距离元组 $\tilde{d}^1_{i,j}(s,(q_1^i,q_2^i))$ 的定义，可得以下不等式

$$\mathsf{d}_i(s,q_1^i)-\mathsf{d}_i(\tilde{s},q_1^i)|\le\mu D \qquad (9)$$

因此，考虑:

$$\begin{aligned}
\mathsf{d}_i(s,\tilde{s})\le\mathsf{d}_i(s,q_1^i)+\mathsf{d}_i(q_1^i,\tilde{s}) &\le 2\mathsf{d}_i(q_1^i,s)+\mu D\\
&\le \frac{2}{\mu}D+\mu D\le\frac{3}{\mu}D\\
&\le 3\mu\cdot\mathsf{d}_i(x,q_1^i) && \text{(From assumption of Case 1B)}\\
&\le 3\mu\cdot(\mathsf{d}_i(x,s_1)+\mathsf{d}_i(s_1,s_2)+d_i(s_2,q_1^i))
\end{aligned}$$

重新整理最后一个不等式，可得:

$$\begin{aligned}
\mathsf{d}_i(s,\tilde{s}) &\le \frac{3\mu}{(1-3\mu)}\cdot(\mathsf{d}_i(x,s_1)+\mathsf{d}_i(s_2,q_1^i))\\
&\le 9\mu\cdot(\mathsf{d}_i(x,s_1)+\mathsf{d}_i(s,q_1^i)+\mu D) && \text{(Using (9))}\\
&\le 9\mu\cdot(\mathsf{d}_i(x,v)+\mathsf{d}_i(v,s_1)+\mathsf{d}_i(p,s)+\mathsf{d}_i(p,s))\\
&\qquad(\mu D\le\mathsf{d}_i(q_1^i,s)\text{ by (5) and }\mathsf{d}_i(q_1^i,s)\le\mathsf{d}_i(p,s)\text{ by the choice of }q_1^i)\\
&\le 9\mu\cdot(\mathsf{d}_{H_i}(x,v)+\mathsf{d}_i(v,s_1)+2\mathsf{d}_i(p,s))\\
&\le 9\mu\cdot(\mathsf{d}_{H_i}(u,v)+\mathsf{d}_i(p,s_1)+2\mathsf{d}_i(p,s))
\end{aligned}$$

其中，最后一个不等式是由以下事实得出的： $u,\ldots,x,\ldots,v$ 是 $H_i$中的一条最短路径，而$p,\ldots,v,\ldots,s_1$ 既是 $G$ 中的一条最短路径，也是 $G_i$中的一条最短路径。因此，

$$\mathsf{d}_i(s,\tilde{s}) \le 9\mu\cdot(\mu\cdot\mathsf{d}(p,s_1)+3\mathsf{d}_i(p,s)) \le 27\mu\cdot(\mathsf{d}(p,s_1)+\mathsf{d}_i(p,s))$$

因此，

$$\mathsf{d}_i(p,s_2) \le \mathsf{d}_i(p,s_1)+27\mu\cdot(\mathsf{d}(p,s_1)+\mathsf{d}_i(p,s)) \tag{10}$$

将 $s_1=s$ 和 $s_2=\tilde{s}$ 代入 (10)，得到：

$$\begin{aligned}\mathsf{d}(p,\tilde{s}) &\le \mathsf{d}_i(p,s)+27\mu(\mathsf{d}_G(p,s_1)+\mathsf{d}_i(p,s))\\ &= \mathsf{d}(p,s)+54\mu\cdot\mathsf{d}(p,s)=(1+54\mu)\mathsf{d}(p,s)\end{aligned} \tag{11}$$

这里用到 $\mathsf{d}_i(p,s_1)=\mathsf{d}(p,s_1)=\mathsf{d}(p,s)$。然后，将 $s_1=\tilde{s}$ 和 $s_2=s$ 代入 (10)，得到：

$$\begin{aligned}&\mathsf{d}_i(p,s)\le \mathsf{d}_i(p,\tilde{s})+27\mu\cdot(\mathsf{d}(p,\tilde{s})+\mathsf{d}_i(p,s))\\ \Longrightarrow &(1-27\mu)\mathsf{d}(p,s)\le(1-27\mu)\mathsf{d}_i(p,s)\le \mathsf{d}(p,\tilde{s})+27\mu\cdot\mathsf{d}(p,\tilde{s})\\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad(\text{Since } \mathsf{d}_i(p,s_1)=\mathsf{d}(p,s_1)=\mathsf{d}(p,\tilde{s}))\\ \Longrightarrow &\mathsf{d}(p,s)\le\frac{1+27\mu}{1-27\mu}\cdot\mathsf{d}(p,\tilde{s})\le(1+1539\mu)\mathsf{d}(p,\tilde{s})\end{aligned} \tag{12}$$

**情形 2**。假设不存在满足 (5) 的点 $q\in R_i$。那么，对于任意点 $q$（特别是对于 $q=p$），以下不等式中恰有一个成立：

$$\mathsf{d}(q,\mathcal{A})+\mathsf{d}_i(q,q_1^i)>\frac{1}{\mu}\mathsf{d}_i(q_1^i,s) \tag{13}$$

$$\textbf{OR } \mathsf{d}(q,\mathcal{A})+\mathsf{d}_i(q,q_1^i)<\mu\mathsf{d}_i(q_1^i,s) \tag{14}$$

在这种情况下，为 $s$ 寻找替代点时，我们在舍入元组 $\mathsf{D}(s)$ 中选择第二个舍入距离元组 $\tilde{d}^2_{i,j}(s,(q_1^i,q_3^i,q_4^i))\tilde{}$，它具有以下性质。若存在满足 (13) 的点 $q\in R_i$，则选择左侧取值最小的那个；否则选择 $q_3^i=\perp$。类似地，若存在满足 (14) 的点 $q\in R_i$，则选择左侧取值最大的那个；否则选择 $q_4^i=\perp$。注意，由于情形假设，点 $q_3^i$ 和 $q_4^i$ 中至少有一个不等于 $\perp$。因此，以下不等式（或多个不等式）成立：

$$\frac{1}{\mu}\cdot\big(\mathsf{d}_i(q_1^i,q_4^i)+\mathsf{d}(q_4^i,\mathcal{A})\big)\le\mathsf{d}_i(q_1^i,s)\le\mu\cdot\big(\mathsf{d}_i(q_1^i,q_3^i)+\mathsf{d}(q_3^i,\mathcal{A})\big) \tag{15}$$

其中，如果相应的点 $q_3^i$ 或 $q_4^i$ 是 $\perp$，我们就从要求中删去相应的不等式。因此，$\tilde{}\mathsf{d}^2_{i,j}(s,(q_1^i,q_3^i,q_4^i))=$ 为真。因此，$\tilde{}\mathsf{d}^2_{i,j}(\tilde{s},(q_1^i,q_3^i,q_4^i))=$ 为真，这意味着：

$$\frac{1}{\mu}\cdot\big(\mathsf{d}_i(q_1^i,q_4^i)+\mathsf{d}(q_4^i,\mathcal{A})\big)\le\mathsf{d}_i(q_1^i,\tilde{s})\le\mu\cdot\big(\mathsf{d}_i(q_1^i,q_3^i)+\mathsf{d}(q_3^i,\mathcal{A})\big) \tag{16}$$

同样，需注意删去适当的不等式。

现在我们考虑 根据 $p$ 满足 (13) 和 (14) 中的哪一个不等式，可分为两种情况 .

**情况 2A**。 假设 $\mathsf{d}(p,\ \mathcal{A})\ +\mathsf{d}_i(p,\ q_1^i) > \frac{1}{\mu}\mathsf{d}_i(q_1^i,\ s)$。那么，根据 $q_3^i$ 的选取，我们有：

$$\begin{aligned}
\mathsf{d}_i(s,\tilde{s}) &\le \mathsf{d}_i(s,q_1^i)+\mathsf{d}_i(\tilde{s},q_1^i)\\
&\le 2\mu\cdot\big(\mathsf{d}_i(q_1^i,q_3^i)+\mathsf{d}(q_3^i,\mathcal{A})\big) && \text{(From (15) and (16))}\\
&\le 2\mu\cdot\big(\mathsf{d}_i(q_1^i,p)+\mathsf{d}(p,\mathcal{A})\big) && \text{(Since } q_3^i \text{ is a point minimizing the LHS of (13))}\\
&\le 2\mu\cdot(\mathsf{d}_i(p,s)+\mathsf{d}_i(q_i,s)+\mathsf{d}(p,\mathcal{A}))\\
&\le 2\mu\cdot(2\mathsf{d}_i(p,s)+\mathsf{d}(p,\mathcal{A})) && \text{(Since } q_1^i \text{ is a point minimizing } \mathsf{d}_i(\cdot,s))\\
&\le 4\mu(\mathsf{d}_i(p,s)+\mathsf{d}(p,\mathcal{A}))
\end{aligned}$$

因此，我们得到以下不等式。

$$\begin{aligned}
\mathsf{d}(p,s_1)\le \mathsf{d}(p,s_2)+\mathsf{d}(s_1,s_2)&\le \mathsf{d}(p,s_2)+\mathsf{d}_i(s_1,s_2)\\
&\mathsf{d}(p,s_2)+4\mu\cdot(\mathsf{d}_i(p,s)+\mathsf{d}(p,\mathcal{A}))
\end{aligned}\tag{17}$$

代入 $s_1=s$ 和 $s_2=\tilde{s}$，得到：

$$\begin{aligned}
&\mathsf{d}(p,s)\le \mathsf{d}(p,\tilde{s})+4\mu\cdot\mathsf{d}(p,s)+4\mu\cdot\mathsf{d}(p,\mathcal{A})\\
\Longrightarrow\ &(1-4\mu)\cdot\mathsf{d}(p,s)\le \mathsf{d}(p,\tilde{s})+4\mu\cdot\mathsf{d}(p,\mathcal{A})\\
\Longrightarrow\ &\mathsf{d}(p,s)\le (1+8\mu)\cdot\mathsf{d}(p,\tilde{s})+12\mu\cdot\mathsf{d}(p,\mathcal{A})
\end{aligned}\tag{18}$$

代入 $s_1=\tilde{s}$ 和 $s_2=s$，得到：

$$\mathsf{d}(p,\tilde{s})\le(1+4\mu)\cdot\mathsf{d}(p,s)+4\mu\cdot\mathsf{d}(p,\mathcal{A})\tag{19}$$

**情况 2B**。 假设 $\mathsf{d}(p,\mathcal{A})\ +\mathsf{d}_i(p,q_1^i)\ <\mu\mathsf{d}_i(q_1^i,s)$。那么，根据 $q_4^i$ 的选取，有 $\mathsf{d}_i(q_1^i,s_1)\ >\frac{1}{\mu}\cdot\big(\mathsf{d}_i(q_1^i,q_4^i)\ +\mathsf{d}(q_4^i,\mathcal{A})\big)$，因为 $s$ 和 $\tilde{s}$ 都满足此不等式。因此，

$$\begin{aligned}
\mathsf{d}(p,s_1)=\mathsf{d}_i(p,s_1)&\ge \mathsf{d}_i(q_1^i,s_1)-\mathsf{d}_i(p,q_1^i)\\
&\ge \frac{1}{\mu}\cdot\big(\mathsf{d}_i(q_1^i,q_4^i)+\mathsf{d}(q_4^i,\mathcal{A})\big)-\mathsf{d}_i(p,q_1^i) && \text{(From above)}\\
&\ge \frac{1}{\mu}\cdot\big(\mathsf{d}_i(q_1^i,p)+\mathsf{d}(p,\mathcal{A})\big)-\mathsf{d}_i(p,q_1^i) && \text{(By the choice of } q_4^i)\\
&\ge \frac{1}{\mu}\cdot\mathsf{d}(p,\mathcal{A})
\end{aligned}$$

因此，上述不等式表明，在情况 2B 中，$\mathsf{d}(p,s)$ 和 $\mathsf{d}(p,\tilde{s})$ 都大于 $\frac{1}{\mu}\mathsf{d}(p,\mathcal{A})$。

因此，结合所有情形，我们得到：对于某些常数 $1\le c$、$c'\le 1539$，要么 $\mathsf{d}(p,s_1)\ >\frac{1}{\mu}\mathsf{d}(p,\mathcal{A})$，要么 $\mathsf{d}(p,s_2)\ \le(1+c\mu)\cdot\mathsf{d}(p,s_1)\ +c'\mu\mathsf{d}(p,s_2)$。令 $\gamma=\max\{c,c'\}=1539$。因此，通过设置 $\mu=\gamma\cdot\frac{\epsilon}{z}$，我们得到：要么 $\mathsf{d}(p,s_1)\ >\frac{\gamma z}{\epsilon}\mathsf{d}(p,\mathcal{A})$，要么 $\mathsf{d}(p,s_2)\ \le(1+\frac{\epsilon}{z})\mathsf{d}(p,s_1)\ +\frac{\epsilon}{z}\mathsf{d}(p,\mathcal{A})$。□

**反向**。

现在，我们旨在证明“反向”。具体而言，我们关注点 $p\in X$，其在 $\tilde{\mathcal{S}}$ 中的最近中心为 $\tilde{s}_1\in\tilde{\mathcal{S}}$。接着证明，若 $\mathsf{d}_G(p,\ \tilde{s}_1)\ \le\frac{10z}{\epsilon}\cdot\mathsf{d}_G(p,\ \mathcal{A})$，则 $\mathsf{d}_G(p,\ \mathcal{S})\ \le(1+\mathcal{O}(\epsilon/z))\ \cdot\mathsf{d}_G(p,\ \tilde{s}_1)\ +\mathcal{O}(\epsilon/z)\ \cdot\mathsf{d}_G(p,\ \mathcal{A})$。与之前一样，此证明根据不同情形组织为以下若干断言。在此之前，我们需要下面这个引理，其证明可直接由 [18] 的引理 18 得出。

引理*3.* 任取满足以下条件的 $s \in \mathcal{S}_{net}$：$d_G(q,\mathcal{A}) < 1\}$。令 $q' = arg\,min^{\{q\in X'_s\}} d_G(q,\mathcal{A}) + d_G(q,s)$。不失一般性，我们可以假设，对于所有满足 $d_G(p,\mathcal{A}) < 1$ 的点 $p$，均有 $(d_G(p,\mathcal{A}) + d_G(p,s)) > \frac{\epsilon^2}{8z^2} \cdot (d_G(q',\mathcal{A}) + d_G(q',s))$。

因此，我们假设引理3所保证的性质成立，并着手证明以下断言。

~断言 *6.* 设 $p \in X$ 为满足 *(1)* $d_G(p,\mathcal{A}) < 1$ 的一个点，且 *(2)* $\tilde{s}_1$ 是 $\mathcal{S}$ 中离$p$最近的中心，并且 $d_G(p,\tilde{s}_1) \le \frac{10z}{\epsilon} \cdot d_G(p,\mathcal{A})$。那么，$d_G(p,\mathcal{S}) \le (1+\mathcal{O}(\epsilon/z)) \cdot d_G(p,\tilde{s}_1) + \mathcal{O}(\epsilon/z) \cdot d_G(p,\mathcal{A})$。

证明。令 $s$ 为 $\mathcal{S}$ 中离 $p$最近的中心。注意，由于 $\mathsf{d}_G(p,\mathcal{A}) < 1$, $s \in \mathcal{S}_{\text{net}}$，且$\rho(s) = \tilde{s} \in \mathbb{C}_{\text{net}}$。此外，根据引理的条件，$\mathsf{d}_G(p,\tilde{s}_1) \le \mathsf{d}_G(p,\tilde{s})$。我们分三种情况讨论。

首先，假设存在某个 $s_1 \in \mathcal{S}_{\text{net}}$，使得 $\rho(s_1) = \tilde{s}_1$。[3] 那么，集合 $X'_{s_1} = \{q \in X_{s_1} : \mathsf{d}_G(q,\mathcal{A}) < 1\}$ 非空。令 $q' = \arg\min^{q\in X'_{s_1}} \mathsf{d}_G(q,\mathcal{A}) + \mathsf{d}_G(q,s_1)$。由此可得，$\mathsf{d}_G(s_1,\tilde{s}_1) \le \frac{\epsilon^3}{z^3} \cdot \mathsf{d}_G(q',\mathcal{A})$。然后，根据引理3,

$$\begin{aligned}
\mathsf{d}_G(p,s_1) &\le \mathsf{d}_G(p,\tilde{s}_1) + \mathsf{d}_G(s_1,\tilde{s}_1) \\
&\le \mathsf{d}_G(p,\tilde{s}_1) + \frac{\epsilon^3}{z^3} \cdot \mathsf{d}_G(q',\mathcal{A}) \\
&\le \mathsf{d}_G(p,\tilde{s}_1) + \frac{\epsilon^3}{z^3} \cdot (\mathsf{d}_G(q',\mathcal{A}) + \mathsf{d}_G(q',s_1)) \\
&\le \mathsf{d}_G(p,\tilde{s}_1) + \frac{\epsilon^3}{z^3} \cdot \frac{8z^2}{\epsilon^2} \cdot (\mathsf{d}_G(p,\mathcal{A}) + \mathsf{d}_G(p,s_1)) \\
&= \mathsf{d}_G(p,\tilde{s}_1) + \mathcal{O}(\epsilon/z) \cdot (\mathsf{d}_G(p,\mathcal{A}) + \mathsf{d}_G(p,s_1))
\end{aligned}$$

因此，

$$\mathsf{d}_G(p,s) \le \mathsf{d}_G(p,s_1) \le (1+\mathcal{O}(\epsilon/z)) \cdot \mathsf{d}_G(p,\tilde{s}_1) + \mathcal{O}(\epsilon/z) \cdot \mathsf{d}_G(p,\mathcal{A}).$$

在第二种情况下，存在某个 $s_1 \in \mathcal{S}_{\text{support}}$，使得 $\rho(s_1) = f(s_1) = \tilde{s}_1$。

$$\mathsf{d}_G(p,\tilde{s}_1) \ge \mathsf{d}_G(p,s_1) - \mathcal{O}(\epsilon^2/z^2) \ge \mathsf{d}_G(p,s) - \mathcal{O}(\epsilon^2/z^2)$$

现在，由于 $s_1$ 被 $\tilde{s}_1 \in \mathcal{S}_{\text{support}}$ 替代，尽管 $\mathsf{d}_G(p,\tilde{s}) > \mathsf{d}_G(p,\tilde{s}_1)$，根据我们的替换方案，必有 $\mathsf{d}_G(p,s) > \epsilon/z$。因此

$$\begin{aligned}
\mathsf{d}_G(p,s) \le \mathsf{d}_G(p,s_1) &\le \mathsf{d}_G(p,\tilde{s}_1) + \mathsf{d}_G(s_1,\tilde{s}_1) \\
&\le \mathsf{d}_G(p,\tilde{s}_1) + \mathcal{O}(\epsilon^2/z^2) \\
&\le \mathsf{d}_G(p,\tilde{s}_1) + \mathcal{O}(\epsilon/z) \cdot \mathsf{d}_G(p,s)
\end{aligned}$$

因此，$\mathsf{d}_G(p,s) \le (1+\mathcal{O}(\epsilon/z)) \cdot \mathsf{d}_G(p,\tilde{s}_1) \le (1+\mathcal{O}(\epsilon/z)) \cdot \mathsf{d}_G(p,\tilde{s}_1) + (\epsilon/z) \cdot \mathsf{d}_G(p,\mathcal{A})$。 □

~**断言 7**。考虑一个点 $p \in X$，其满足 $\mathsf{d}_G(p,\mathcal{A}) \ge 1$，且 $\tilde{s}_1$ 是从 $\mathcal{S}$ 中到 $p$ 的最近中心，满足 $\rho(s_1) = \tilde{s}_1$，其中 $s_1 \in \mathcal{S}_{net} \cup \mathcal{S}_{support}$。那么，$\mathsf{d}_G(p,\mathcal{S}) \le \mathsf{d}_G(p,\tilde{s}_1) + \frac{\epsilon}{z} \cdot \mathsf{d}_G(p,\mathcal{A})$。

证明。我们考虑两种情况。如果 $s_1 \in \mathcal{S}_{\text{net}}$，那么 $\mathsf{d}_G(s_1,\tilde{s}_1) \le \frac{\epsilon^3}{z^3} \le \frac{\epsilon}{z} \cdot \mathsf{d}_G(p,\mathcal{A})$。或者，如果 $s_1 \in \mathcal{S}_{\text{support}}$，那么 $\mathsf{d}_G(s_1,\tilde{s}_1) \le \frac{\epsilon^2}{z^2} \le \frac{\epsilon^2}{z^2} \cdot \mathsf{d}_G(p,\mathcal{A})$。这里，我们在每个不等式中都用到了 $\mathsf{d}_G(p,\mathcal{A}) \ge 1$ 这一事实。于是，结论成立，因为：

$$\mathsf{d}_G(p,\mathcal{S}) \le \mathsf{d}_G(p,s_1) \le \mathsf{d}_G(p,\tilde{s}_1) + \mathsf{d}_G(s_1,\tilde{s}_1) \le \mathsf{d}_G(p,\tilde{s}_1) + \frac{\epsilon}{z} \cdot \mathsf{d}_G(p,\mathcal{A}).$$

□

[3]注意，这里可能还存在某个 $s' \in \mathcal{S}_{\text{net-sub}}$，使得 $\rho(s') = \tilde{s}_1$。但是，在本分析中，只要 $s_1 \in \mathcal{S}_{\text{net}}$ 存在就足够了。

**综上所述**。 首先，我们证明以下引理中的误差界。

**引理 4**。 对于任意 $\mathcal{S} \in \mathbb{C}^k$，存在一个 $\tilde{\mathcal{S}} \in \mathbb{C}^k$，具有以下性质。对于任意点 $p \in X$，如果 代价$(p, \mathcal{S}) \leq \left(\frac{10z}{\epsilon}\right)^z \mathsf{d}(p, \mathcal{A})$，或 代价$(p, \tilde{\mathcal{S}}) \leq \left(\frac{10z}{\epsilon}\right)^z \mathsf{d}(p, \mathcal{A})$，则

$$|\mathsf{cost}(p, \mathcal{S}) - \mathsf{cost}(p, \tilde{\mathcal{S}})| \leq \frac{\epsilon}{z \log(z/\epsilon)} \cdot (\mathsf{cost}(p, \mathcal{S}) + \mathsf{cost}(p, \mathcal{A})) .$$

证明。 令 $\mu \geq 0$ 为充分大的常数。通过结合断言 3、断言 4、断言 5 和引理 2，并对 $\epsilon$ 进行适当的重新缩放，我们得出：对于每个满足 $\mathsf{d}_G(p, \mathcal{S}) \leq \frac{9\mu z}{\epsilon} \cdot \mathsf{d}_G(p, \mathcal{A})$ 的点 $p \in X$，以下结论成立。

$$\mathsf{d}_G(p, \tilde{\mathcal{S}}) \leq (1 + \epsilon/z) \cdot \mathsf{d}_G(p, \mathcal{S}) + (\epsilon/z) \cdot \mathsf{d}_G(p, \mathcal{A}) \tag{20}$$

$$\implies \mathsf{cost}(p, \tilde{\mathcal{S}}) \leq (1 + \epsilon) \cdot \mathsf{cost}(p, \mathcal{S}) + \epsilon \cdot \mathsf{cost}(p, \mathcal{A}) \tag{21}$$

这里，我们在最后一步使用了引理 1。

类似地，通过结合断言 6、断言 7 和引理 2，并对 $\epsilon$ 进行适当的重新缩放，我们得出：对于每个满足 $\mathsf{d}_G(p, \tilde{\mathcal{S}}) \leq \frac{9\mu z}{\epsilon} \mathsf{d}_G(p, \mathcal{A})$ 的点 $p \in X$，以下结论成立。

$$\mathsf{d}_G(p, \mathcal{S}) \leq (1 + \epsilon/z) \cdot \mathsf{d}_G(p, \tilde{\mathcal{S}}) + (\epsilon/z) \cdot \mathsf{d}_G(p, \mathcal{A}) \tag{22}$$

$$\implies \mathsf{cost}(p, \tilde{\mathcal{S}}) \leq (1 + \epsilon) \cdot \mathsf{cost}(p, \mathcal{S}) + \epsilon \cdot \mathsf{cost}(p, \mathcal{A}) \tag{23}$$

在最后一步中，我们再次使用引理 1。

现在，利用 (20)，我们可以推断，对于任意 $\lambda \leq 9\mu$，

$$\mathsf{d}_G(p, \mathcal{S}) \leq \frac{\lambda z}{\epsilon} \cdot \mathsf{d}_G(p, \mathcal{A}) \implies \mathsf{d}_G(p, \tilde{\mathcal{S}}) \leq \frac{3\lambda z}{\epsilon} \cdot \mathsf{d}_G(p, \mathcal{A}) \tag{24}$$

类似地，利用 (22)，我们可以推断，对于任意 $\lambda \leq 9\mu$，

$$\mathsf{d}_G(p, \tilde{\mathcal{S}}) \leq \frac{\lambda z}{\epsilon} \cdot \mathsf{d}_G(p, \mathcal{A}) \implies \mathsf{d}_G(p, \mathcal{S}) \leq \frac{3\lambda z}{\epsilon} \mathsf{d}_G(p, \mathcal{A}) \tag{25}$$

因此，通过对 $\lambda \leftarrow \mu$ 应用 (24) 和 (25)，我们可以得出结论：要么 $\mathsf{d}_G(p, \mathcal{S})$ 和 $\mathsf{d}_G(p, \tilde{\mathcal{S}})$ 都大于 $\frac{3\mu z}{\epsilon} \cdot \mathsf{d}_G(p, \mathcal{A})$，此时结论成立；否则，两者都小于 $\frac{9\mu z}{\epsilon} \cdot \mathsf{d}_G(p, \mathcal{A})$。然后，利用 (21) 和 (23)，我们得到，对于任意这样的点 $p \in X$，

$$\begin{aligned}
&|\mathsf{cost}(p, \mathcal{S}) - \mathsf{cost}(p, \tilde{\mathcal{S}})| \leq 2\epsilon \cdot \left(\mathsf{cost}(p, \mathcal{S}) + \mathsf{cost}(p, \tilde{\mathcal{S}}) + \mathsf{cost}(p, \mathcal{A})\right) \\
\implies &(1 - 2\epsilon) \cdot |\mathsf{cost}(p, \mathcal{S}) - \mathsf{cost}(p, \tilde{\mathcal{S}})| \leq 4\epsilon \cdot (\mathsf{cost}(p, \mathcal{S}) + \mathsf{cost}(p, \mathcal{A})) \\
&\qquad\qquad (\text{Using } \mathsf{cost}(p, \tilde{\mathcal{S}}) \leq \mathsf{cost}(p, \mathcal{S}) + |\mathsf{cost}(p, \tilde{\mathcal{S}}) - \mathsf{cost}(p, \mathcal{S})|) \\
\implies &|\mathsf{cost}(p, \mathcal{S}) - \mathsf{cost}(p, \tilde{\mathcal{S}})| \leq 12\epsilon \left(\mathsf{cost}(p, \mathcal{S}) + \mathsf{cost}(p, \mathcal{A})\right)
\end{aligned} \tag{26}$$

因此，通过最后进行一次重新缩放 $\epsilon \leftarrow \frac{\epsilon}{cz \log(z/\epsilon)}$，其中 $c$ 为充分大的常数，我们即可得到该引理。 □

现在，我们来界定质心集的大小。

**引理 5.** $|\mathbb{C}| = \exp\left(\mathcal{O}\left(\log^2 |X| + z^{16}\epsilon^{-8}(\log(z/\epsilon))^8 \log |X|\right)\right)$.

证明。 回顾一下，$\mathbb{C} := \mathbb{C}_{\text{landmark}} \cup \mathbb{C}_{\text{support}} \cup \mathbb{C}_{\text{net}}$。首先，根据命题4，$\mathbb{C}_{\text{support}}$ 的大小至多为（$\Delta$（$z^2/\epsilon^2$））$^{\ell}|X|$。由于 $\ell = \Theta$（$z/\epsilon$），这等于（$z/\epsilon$）$^{\mathcal{O}(z/\epsilon)} \cdot |X|$。接下来，我们

界定 $\mathbb{C}_{\text{net}}$。利用命题 3，可以将每个 $B_i'$（它是 $B_i$ 的一个（$\epsilon^3/z^3$）$\mathsf{d}_G$（$p_i$，$\mathcal{A}$）-网）的大小上界为$\mathcal{O}$（$z^8\epsilon^{-8}$）。因此，$\mathbb{C}_{\text{net}}$ 的大小为 $\mathcal{O}$（$|X|z^8\epsilon^{-8}$）。

现在，我们对 $\mathbb{C}_{\text{landmark}}$ 进行界定。注意，$\mathbb{C}_{\text{landmark}}$ 中每个舍入距离元组和每个叶子至多包含一个点。注意，叶子的数量至多为 $\mathcal{O}$（$|X|$），因此根叶路径的数量也相同。现在，固定一条根叶路径 $\Pi = (R_1, R_2, \ldots, R_t)$，以及与之对应的、大小为 $\mathcal{O}$（$\log |X|$）的最短路径分隔器集合 $\mathcal{P}$。对于每个 $P^{i,j} \in \mathcal{P}$，我们可以选择第一舍入距离元组或第二舍入距离元组。对于一条路径 $P^{i,j}$，若选择第一舍入距离元组，首先需要选取点 $q_1$ 和 $q_2$，这带来 $|X|^2$ 种选择。接着，路径上至多有 $\mathcal{O}$（$(z/\mu)^4$）个地标点，且每个距离至多有 $\mathcal{O}$（$z^4/\epsilon^4$）种取值。另一方面，若选择第二舍入距离元组，则需要选取 $q_1, q_3$ 和 $q_4$，这带来 $|X|^3$ 种选择。因此，对于特定的 $P^{i,j}$，选择数量的上界为 $|X|^{\mathcal{O}(1)} \cdot (z/\epsilon)^{\mathcal{O}(z^4/\epsilon^4)}$。因此，舍入距离元组总数的上界为 $|X|^{\mathcal{O}(\log |X|)} \cdot (z/\epsilon)^{\mathcal{O}(z^4 \log |X|/\epsilon^4)}$。这也是 $\mathbb{C}_{\text{landmark}}$ 大小的上界。

因此，$\mathbb{C}$ 的大小上界为 $\exp\left(\mathcal{O}(\log^2 |X| + z^8\epsilon^{-8} \log |X| \log(z/\epsilon))\right)$。最后，我们将 $\epsilon$ 的重新缩放计入，缩放因子为 $\epsilon/(cz\log(z/\epsilon))$，由此可得 $|\mathbb{C}|$ 的实际上界为：

$$\exp\left(\mathcal{O}\left(\log^2 |X| + z^{16}\epsilon^{-8}(\log(z/\epsilon))^8 \log |X|\right)\right).$$

□

# 4 几何相交图的应用

本节将定理 3 应用于各种几何相交图，以获得大小不依赖于 $n$的核心集。首先，我们考虑欧几里得加权 UDG 度量和 $\ell_\infty$-加权 USG 的情形，并解释它们为何满足两项规范性质。随后，我们考虑具有一般 $\ell_p$ 范数权重的 UDG 和 USG（第 4.3 节）。最后，我们考虑具有有界度的无权 UDG 的特殊情形（第 4.4 节）；这里我们仅讨论论证中所需的修改，以说明相应度量也满足这些性质。

## 4.1 欧几里得加权单位圆盘图。

很容易验证，欧几里得加权单位圆盘图满足局部欧几里得性质。事实上，考虑一个实现单位圆盘图 $G$ 的嵌入 $\lambda: V(G) \to \mathbb{R}^2$。当且仅当对应两点之间的欧几里得距离至多为 2 时，这两个顶点之间才有一条边。此外，这条边的权重恰好等于这两点之间的欧几里得距离。因此，$c_1 = c_2 = 2$ 和 $c_3 = c_4 = 1$ 满足该性质。

至于平面伸展子图性质，我们将用到以下已知结果。

**命题 6。** *[Li、Calinescu 和 Wan [37]]* 对于任意单位圆盘图 $G$，存在一个平面伸展子图 $H$，使得对于任意 $u, v \in V(H)$，$\mathsf{d}_G(u, v) \le \mathsf{d}_H(u, v) \le 2.42 \cdot \mathsf{d}_G(u, v)$。

因此，定理 3 推出了关于欧几里得加权单位圆盘图的以下定理。

**定理4。** 考虑由任意欧几里得加权单位圆盘图$G = (V, E)$所诱导的度量空间$(V, \mathsf{d}_G)$，一个包含 $n$ 个互不相同点的集合 $P \subseteq V$，以及两个正整数 $k$ 和 $z$。则存在一种多项式时间算法，以至少 $1 - \delta$ 的概率构造出一个用于在 $P$上进行$(k, z)$-聚类的核心集，其大小为 $\mathcal{O}(\epsilon^{-\beta} k \log^2 k \log^3(1/\delta))$，其中 $z$ 为一个常数，$\beta = \mathcal{O}(z \log z)$，且 $\delta < 1/4$。

## 4.2 $\ell_\infty$-加权单位正方形图

**预备知识**。考虑平面上坐标为$(a,b)$的点 $q$。对于一个 $p \geq 1$，令 $||(a,b)||_p := (|a|^p + |b|^p)^{1/p}$，其中 $p \leq \infty$；并令 $||(a,b)||_\infty := \lim_{p\to\infty} ||(a,b)||_p = \max\{|a|,|b|\}$。不同范数之间存在如下关系。

**观察 3.** 对于任意 $(a,b) \in \mathbb{R}^2$ 以及任意 $1 \leq p \leq q \leq \infty$，均有 $||(a,b)||_p \leq ||(a,b)||_q \leq \sqrt{2}\cdot||(a,b)||_p$。

以 $p$ 为中心的单位正方形记作 $S(p)$，它是边长为 2、以 $p$ 为中心且包含其内部的轴平行正方形。[4]另一种说法是，$S(p)$ 是满足 $||p-q||_\infty \leq 1$ 的点 $q$ 的集合，其中 $||\cdot||_\infty$ 表示 $\ell_\infty$ 范数，其定义为 $||(a,b)||_\infty = \max\{|a|,|b|\}$。

点集为 $P \subset \mathbb{R}^2$ 的单位正方形图（USG）是以 $P$ 中每个点为中心的单位正方形所构成的交图。注意，当且仅当单位正方形 $S(p)$ 与 $S(q)$ 相交时，$p,q \in P$, $\{p,q\}$ 是 USG 中的一条边。边 $\{p,q\}$ 的权重定义为 $||p-q||_\infty$。注意，边的权重介于 0 和 2 之间。

典型几何性质。首先，我们论证 $\ell_\infty$-加权 USG 满足局部欧几里得性质。考虑一个带有嵌入 $\lambda$ 的 USG $G$： $V(G) \to \mathbb{R}^2$。根据观察 3 可知，若对于某个 $u$， $v \in V(G)$，若 $||\lambda(u) - \lambda(v)||_2 = |\lambda(u)\lambda(v)| > 2$，则 $||\lambda(u) - \lambda(v)||_\infty > 2$，这意味着 $uv \notin E(G)$。另一方面，若 $||\lambda(u) - \lambda(v)||_2 = |\lambda(u)\lambda(v)| \leq \sqrt{2}$，则$||\lambda(u) - \lambda(v)||_\infty \leq 2$，这意味着 $uv \in E(G)$。此外，此类边的权重为 $w(uv) = ||\lambda(u) - \lambda(v)||_\infty$，这意味着 $|\lambda(u)\lambda(v)| \leq w(u,v) \leq \sqrt{2}\cdot|\lambda(u)\lambda(v)|$。因此，$\ell_\infty$-加权 USG 在$c_1 = \sqrt{2}, c_2 = 2$, $c_3 = 1, c_4 = \sqrt{2}$ 的条件下满足局部欧几里得性质。

回顾一下，文献 [37] 已证明欧几里得边权单位圆盘图存在常数平面伸展子图。然而，此前尚无人知道 $L_\infty$ 边权单位正方形图是否存在类似的伸展子图。我们通过证明如下定理所述的伸展子图确实存在，弥补了这一空白。定理 5 本身就很有趣；我们预计，与单位圆盘图的平面伸展子图类似，单位正方形图的平面伸展子图也会在聚类之外找到应用。为了保持论文行文流畅，我们将定理 5 的证明推迟到下一节，即第 5 节。

**定理 5.** 设 $P$ 是满足一般位置假设的点集，并令 $G = (P,E)$ 为与 $P$ 相关联的单位正方形图，其中边 $uv$ 的权重等于 $||u-v||_\infty$。那么，存在 $G$ 的一个平面子图 $H$，使得对于任意两个点 $a$, $b \in P$，$\mathsf{d}_G(a,b) \leq \mathsf{d}_H(a,b) \leq 3\mathsf{d}_G(a,b)$。

因此，单位正方形图具有伸展因子为 $\alpha = 3$ 的平面生成器性质。然后，通过应用定理 1，我们得到以下定理。

**定理 6**。考虑由任意 $L_\infty$ 带权单位正方形图$G = (V,E)$诱导的度量空间 $(V,\mathsf{d}_G)$ 、一个包含 $n$ 个不同点的集合 $P \subseteq V$，以及两个正整数 $k$ 和 $z$。则存在一个多项式时间算法，以至少 $1-\delta$ 的概率为 $(k,z)$-聚类在 $P$ 上构造一个大小为 $\mathcal{O}(\epsilon^{-\beta}k\log^2 k\log^3(1/\delta))$ 的核心集，其中 $z$ 为常数，$\beta = \mathcal{O}(z\log z)$，且 $\delta < 1/4$。

## 4.3 具有其他范数的单位圆盘图和单位正方形图

令 $G = (V, E)$ 为一个与平面中的点集 $V$ 对应的单位圆盘图。回顾一下，对于 $u$， $v \in V$，当且仅当 $||u-v||_2 \leq 2$ 时，存在一条边 $\{u,v\} \in E(G)$。在欧几里得边权单位圆盘图中，我们将边的权重定义为

[4]今后，当我们提到正方形时，总是指包括边界上的点及其内部的点。

该边的权重同样为 $||u-v||_2$。但是，对于任意 $p \geq 1$，也可以将权重定义为点之间的$\ell_p$ 距离。为了证明 $\ell_p$-加权单位圆盘图满足局部欧几里得 性质，我们可以采用与 $\ell_\infty$-加权 USG 情形类似的论证。细节略。

**平面伸展子图**。考虑将命题 6 应用于 $G'$ 后得到的伸展子图 $H'$，其中$G'$ 是对应于点集 $V$ 的欧几里得（即 $L_2$）加权单位圆盘图。由此可知，对于任意 $u, v \in V$，$\mathrm{d}_{G'}(u, v) \leq \mathrm{d}_{H'}(u, v) \leq 2.42 \cdot \mathrm{d}_{G'}(u, v)$。特别地，对于任意边 $u, v \in E(G)$，$||u-v||_2 \leq \mathrm{d}_{H'}(u, v) \leq 2.42 \cdot ||u-v||_2$。

首先，考虑 $G$ 中的权重由点之间的 $L_p$ 范数距离给出的情形，其中 $1 \leq p < 2$。由观察 3 可知，$\mathrm{d}_G(u, v) = ||u-v||_p \leq ||u-v||_2 \leq \mathrm{d}_{H'}(u, v) \leq 2.42 \cdot ||u-v||_2 \leq 2.42 \cdot \sqrt{2} \cdot ||u-v||_p$。由此可知，对于任意 $u', v' \in G$，$\mathrm{d}_G(u', v') \leq \mathrm{d}_{H'}(u', v') \leq 3.42 \cdot \mathrm{d}_G(u', v')$。

否则，$G$ 中的权重由 $L_p$ 范数距离给出，其中 $2 < p \leq \infty$。接着，令$H$ 为通过将伸展子图 $H'$ 中每条边的权重乘以因子$\sqrt{2}$得到的加权图。注意，对于任意边 $\{u, v\} \in E(G)$，观察 3 蕴含 $||u-v||_p \leq \sqrt{2} \cdot ||u-v||_2 \leq \sqrt{2} \cdot ||u-v||_p$。这意味着，对于 $\{u, v\} \in E(G)$，$\mathrm{d}_G(u, v) = ||u-v||_p \leq \sqrt{2}||u-v||_2 \leq \sqrt{2} \cdot \mathrm{d}_{H'}(u, v) = \mathrm{d}_H(u, v) \leq 2.42 \cdot \sqrt{2} \cdot \sqrt{2} \cdot ||u-v||_p$。这意味着，对于任意$u', v' \in V$，有 $\mathrm{d}_G(u, v) \leq \mathrm{d}_H(u', v') \leq 4.84 \cdot \mathrm{d}_G(u, v)$。

无论哪种情况，我们都可以得出结论：命题 6 可用于证明存在一个 $\alpha$-伸展平面伸展子图，其中 $\alpha \leq 4.84$。

根据上述讨论可知，$\ell_p$-加权单位圆盘图也满足应用我们框架所需的两个性质。最后，稍作修改前面的论证，也容易证明 $\ell_p$-加权 USG 满足典型几何性质。因此，我们得出以下定理。

**定理 7.** 考虑由任意 $\ell_p$-加权单位圆盘图（分别为 $\ell_p$-加权 *USG*）$G = (V, E)$ 所诱导的度量空间 $(V, \mathrm{d}_G)$，其中 $1 \leq p \leq \infty$，集合 $P \subseteq V$ 包含 $n$ 个不同的点，且有两个正整数$k$ 和 $z$。那么，存在一个多项式时间算法，以至少 $1-\delta$的概率，为 $P$上的$(k, z)$-聚类构造一个大小为 $\mathcal{O}(\epsilon^{-\beta} k \log^2 k \log^3(1/\delta))$的核心集，其中 $z$和 $\beta$为常数，且 $\delta < 1/4$。

## 4.4 有界度的无权单位圆盘图

这里，我们考虑由无权 UDG 诱导的度量。也就是说，如果 $G = (V, E)$ 是一个单位圆盘图，那么对于任意两个顶点 $u, v \in V(G)$，距离 $\mathrm{d}_G(u, v)$ 由最短路径 $\pi_G(u, v)$ 的跳数长度 $|\pi_G(u, v)|$ 给出。请注意，在 $G$ 的嵌入中，彼此非常接近的两个点，根据跳数度量，它们之间的距离仍为 1。因此，无权单位圆盘图不满足 Locally Euclidean 性质。不过，如下文所述，它们确实具有常数伸展平面生成器。下面我们将展示，当最大度上界为Δ时，依赖局部欧几里得性质的论证可以针对无权单位圆盘图进行修改。因此，我们所构造的核心集大小将取决于 Δ。

令 $G = (V, E)$ 为无权单位圆盘图。首先，我们注意到，任意一对顶点之间的距离由它们之间最短路径的跳数长度给出，因此始终为非负整数。此外，$\mathrm{d}_G(u, v) = 0$ 当且仅当 $u = v$；否则 $\mathrm{d}_G(u, v) \geq 1$。

通过考察质心集定理的证明，我们观察到以下事实。局部欧几里得性质用于构造集合 $\mathbb{C}_{\text{net}}$ 和 $\mathbb{C}_{\text{support}}$，而平面伸展子图 性质用于构造集合 $\mathbb{C}_{\text{landmark}}$。最后，质心集 $\mathbb{C}$ 被定义为三个集合 $\mathbb{C}_{\text{net}}$、$\mathbb{C}_{\text{support}}$ 和 $\mathbb{C}_{\text{landmark}}$ 的并集。下面，我们将针对无权单位圆盘图调整 $\mathbb{C}$ 的构造。

首先，我们利用以下结果。

**命题 7** ([7])**：** 对于任意无权单位圆盘图 $G'$，都存在一个平面伸展子图 $H'$，使其能够近似 $G'$ 中的跳数距离。也就是说，对于任意 $u$、$v \in V(G')$，$\mathsf{d}_{G'}(u,v) \leq \mathsf{d}_{H'}(u,v) \leq \alpha \cdot \mathsf{d}_{G'}(u,v)$，其中 $\alpha \geq 1$ 为某个绝对常数。

注意，伸展因子 $\alpha$ 是一个绝对常数，并且该结果 不 要求 UDG 的最大度有界。在此基础上，令 $\ell := \alpha/\epsilon$。我们完全按照原始构造，利用递归分解和平面伸展子图来构造集合 $\mathbb{C}_{\text{landmark}}$。

我们定义 $\mathbb{C}_{\text{net}} := \emptyset$，并如下构造 $\mathbb{C}_{\text{support}}$。对于每个点 $p \in X$，我们将 $G$ 中距离 $p$ 不超过 $\ell$ 跳的 所有 顶点加入集合 $\mathbb{C}_{\text{support}}$。或者，也可以将其视为构造有界度支撑图的命题4的另一种方案，即支撑图 $H$ 等于 $G$，且 $f: V(G) \to V(H)$ 为恒等映射。唯一需要注意的是，$H$ 的度由 $\Delta$ 所界定，而不是原始定义所要求的 $\mathcal{O}(1/\mu^2)$。最后，令 $\mathbb{C} = \mathbb{C}_{\text{landmark}} \cup \mathbb{C}_{\text{support}}$。我们有如下观察。

**断言 8.** $|\mathbb{C}| = \exp\left(\mathcal{O}\left(\log^2|X| + z^{16}\epsilon^{-8}(\log(z/\epsilon))^8 \log|X| + \log\Delta\right)\right)$

证明. 由于 $\mathbb{C}_{\text{landmark}}$ 的构造方式与之前完全相同，我们可以采用与引理 5 的证明中相同的上界。

也就是说，$|\mathbb{C}_{\text{landmark}}| = \exp\left(\log^2|X| + z^{16}\epsilon^{-8}(\log(z/\epsilon))^8 \log|X|\right)$。

现在我们分别给出 $|\mathbb{C}_{\text{support}}|$ 的上界。考虑 $G$ 在平面中的一个实现。我们在平面上覆盖一个边长为 1 的网格。注意，位于同一网格单元中的任意两点之间必定有一条边。由于 $G$ 的最大度为 $\Delta$，因此每个单元至多包含 $V(G)$ 的 $\Delta$ 个顶点。令 $H$ 表示 ..... 此外，如果 $\{u,v\} \in E(G)$，那么分别包含 $u$ 和 $v$ 的网格单元中心之间的欧几里得距离至多为 3。

现在考虑一个点 $p \in X$，以及满足 $\mathsf{d}_G(p,q) \leq \ell$ 的任意顶点 $q$。由此可知，包含 $p$ 和 $q$ 的单元的中心之间的欧几里得距离至多为 $3\ell$。因此，$q$ 至多可以属于满足此条件的 $\pi(3\ell)^2 = \mathcal{O}(\ell^2)$ 个单元中的一个。由于每个单元至多包含 $\Delta$ 个来自 $V(G)$ 的点，并且 $\ell = \mathcal{O}(z/\epsilon)$，所以对于 $p \in X$ 中的每个点，我们至多向 $\mathbb{C}_{\text{support}}$ 添加 $\mathcal{O}(z^2\Delta/\epsilon^2)$ 个点，这意味着 $|\mathbb{C}_{\text{support}}| = \mathcal{O}(|X| \cdot z^2\Delta/\epsilon^2)$。 □

现在我们讨论构造 $\tilde{s}$ 以及进行误差分析所需的修改。考虑某个 $s \in \tilde{\mathcal{S}}$。如果 $s \in \mathbb{C}_{\text{support}}$，那么令 $\tilde{s} = s$，并将其添加到 $\mathcal{S}$ 中。显然，$\mathsf{d}_G(p,s) = \mathsf{d}_G(p,\tilde{s})$。否则，对于所有点 $p \in X$，都有 $\mathsf{d}_G(p,s) > \ell$。然后，我们使用递归分解和分类讨论，找到一个点 $\tilde{s} \in \mathbb{C}_{\text{landmark}}$，使其相对于地标集 $\mathcal{L}(\cdot)$ 的取整距离与原证明中的相同。接着，利用类似的分析，我们可以证明，对于任意 $p \in X_s$，都有 $\mathsf{cost}(p,\mathcal{S}) \leq (1+\epsilon) \cdot \mathsf{cost}(p,\mathcal{S}) + \epsilon \cdot \mathsf{cost}(p,\mathcal{A})$。我们省略细节，并给出以下定理。

**定理 8**。考虑由无权单位圆盘图 $G = (V,E)$ 诱导的度量空间 $(V, \mathsf{d}_G)$，该图的最大度为 $\Delta$；设 $P \subseteq V$ 是一个包含 $n$ 个互异点的集合，并设 $k$ 和 $z$ 为两个正整数。那么，存在一种多项式时间算法，能够以至少 $1-\delta$ 的概率构造出用于 $P$ 上$(k,z)$-聚类的核心集，其大小为 $\mathcal{O}(\epsilon^{-\beta}k\log^2 k\log^3(1/\delta) + \log(\Delta))$，其中 $z$ 和 $\beta$ 为常数，且 $\delta < 1/4$。

## 5 Con单位正方形图的常数伸展平面伸展子图 s

在本节中，我们证明定理 5。

预备知识。对于点 $p=(x_p,y_p)$ $q=(x_q,y_q)\in\mathbb{R}^2$，令 $\Delta_x(p,q)=|x_p-x_q|$ 和 $\Delta_y(p,q)=|y_p-y_q|$，并令 $D(p,q)=||p-q||_\infty=\max\{\Delta_x(p,q)\ \Delta_y(p,q)\}$，以及 $\delta(p,q)=\min\{\Delta_x(p,q)\ \Delta_y(p,q)\}$。对于点 $p$ 和 $q$，我们用 $pq$ （或 $qp$）表示连接 $a$ 和 $b$的（闭）线段。我们用 $|pq|$ 表示线段 $pq$ 的（欧几里得）长度。

下文中，除非另有明确说明，我们所说的正方形或矩形始终指轴平行正方形或轴平行矩形及其内部。我们将正方形 $S$ 的边长记为 $\ell(S)$。以点 $p=(x_p,y_p)$ 为中心的单位正方形为 $S(p)$：$=\{q=(x_q,y_q)\colon\ D(p,q)\le 1\}$。注意，$\ell(S(p))=2$。注意，对于位于 $S(p)$ 边界上的任意点 $q$，都有 $D(p,q)=1$。为方便起见，我们分别用 N、E、S、W 表示与基本方位对应的正方形四条边。对于一对点 $a$、$b$，若 $ab$ 不是轴平行的，我们用 $R(a,b)$ 表示一个轴平行矩形，其对角线的两个端点为 $a$ 和 $b$。

以下两个命题可通过初等几何和 $\ell_\infty$距离的性质证明。

**命题 8**。设 $a=(x_a,y_a)$ 和 $b=(x_b,y_b)$ 是 $\mathbb{R}^2$中的任意两个不同点。设 $S$ 是一个正方形，其边界上有 $a$ 和 $b$。那么，$\ell(S)\ge D(a,b)$。

**命题9**。设 $a$ 和 $b$ 是平面上的两个不同点，且 $ab$ 不平行于 $x$ 或$y$ 轴。那么，对于位于 $R(a,b)$上或内部的任意点 $c$，以下结论成立：$D(a,c)\le D(a,b)$ 且$D(b,c)\le D(a,b)$。

我们沿用 Binichon等人的构造 [8], ，并证明当 $\ell_\infty$ 德劳内三角剖分限制为 USG边时，它仍是相应 USG 的常数伸展生成器。

固定一个点集 $P$。我们采用一般位置假设，即 $P$ 中没有四个点位于某个轴平行正方形的边界上。此外，如有必要，通过对点集进行微小扰动，我们假设所有成对 $L_\infty$ 距离均不相同。在此假设下，$\ell_\infty$德劳内三角剖分是唯一且平面的。若 $pq$ 是德劳内三角剖分 $T$中的一条边，则称 $pq$为一条$DT$边。最后，设 $H$是 USG 在 $P$上的一个子图，定义如下：对于 $p,q\in P$,$pq\in E(H)$，当且仅当 $pq$ 是 DT 边且 $pq\in E(G)$，即 $pq$ 是 USG 边。此外，每条边 $pq\in E(H)$的长度设为 $D(pq)$；也就是说，若 $pq\in E(H)$，则 $\mathrm{d}_H(p,q)=\mathrm{d}_G(p,q)=D(p,q)$。

接下来，我们将证明以下关键引理。

**引理 6.** 设 $a=(x_a,y_a)$ 和 $b=(x_b,y_b)$ 是 $P$ 中的两个点，且满足 $ab\in E(G)$。那么，存在一条路径 $\pi_H(a,b)=(a=v_0,v_1,\dots,v_t=b)$ 位于 $H$ 中，并满足以下 性质。

- $\mathrm{d}_H(a,b)\le\sum_{i=1}^{t}\mathrm{d}_H(v_{i-1},v_i)\le 2D(a,b)+\delta(a,b)\le 3\cdot\mathrm{d}_G(a,b)$，且
- 对于每个 $1\le i\le t$, $D(v_{i-1},v_i)=\mathrm{d}_H(v_{i-1},v_i)\le D(ab)$。

接下来，我们固定一对点 $a$，$b\in P$，使得 $ab$ 不是一条边，并使用点的德劳内三角剖分，从 [8] 开始进行几何构造。我们将更仔细地分析这个点集，以便随后利用这些性质来证明引理 6。

我们按照 [8] 的做法，并令 $a=(0,0)$ 和 $b=(w,h)$ 为一对点，使得 $ab$ 不是 DT 边。这里，我们假设 $0\le h\le w$ 和 $0<w$。因此，注意到 $D(a,b)=w\ge h$。令 $T_1,T_2,\dots,T_k$ 为 $ab$ 从 $a$ 朝 $b$ 移动时所相交的三角形序列。假设矩形 $R(a,b)$ 内除 $a$ 和 $b$ 外不含 $P$ 中的其他点。令 $h_0=l_0=a$。对于每个三角形 $T_i$, $1\le i\le k$，线段 $ab$ 与 $T_i$ 的边相交两次。令 $h_i$ 和 $l_i$ 为沿 $a$ 朝 $b$ 移动时最后遇到的 $T_i$ 的边的端点，且 $h_i$ 位于 $ab$ 上方，而 $l_i$ 位于其下方

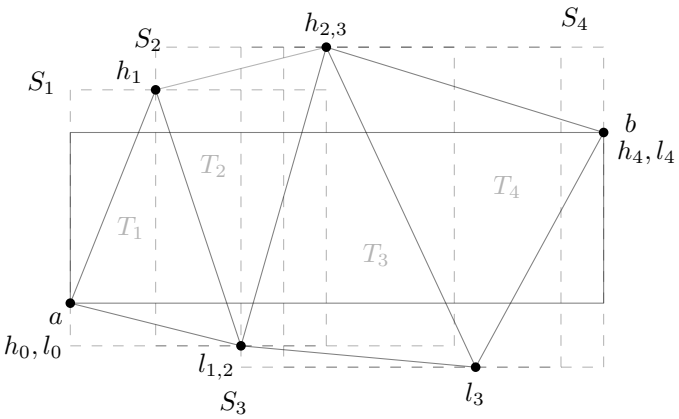


图5：构造中使用的各种点、三角形和正方形。

$\overline{ab}$（这里，上方和下方是根据经过 $a$ 和 $b$ 的直线所定义的两个半平面来确定的）。注意，对于 $1 < i < k$，要么 $h_{i-1} = h_i$ 且 $T_i = \Delta(h_i, l_i, l_{i-1})$，要么 $l_{i-1} = l_i$ 且 $T_i = \Delta(h_{i-1}, h_i, l_i)$。回忆一下，$T_1 = \Delta(a, h_1, l_1)$，且 $T_k = \Delta(h_k, l_k, b)$。对于 $1 \le i \le k$，令 $S_i$ 为一个空正方形，其边界上有 $T_i$ 的顶点（回忆一下，空正方形假设是由相应的边为 DT 边这一事实推得的）。最后，若点 $u = (x_u, y_u)$ 满足 $0 \le x_u \le w$ 和 $y_u > h$（分别为 $y_u < 0$），则称它相对于 $R(a, b)$ 为高点（分别为低点）。Bonichon 等人 [8] 证明了以下引理。

**引理 7** (引理 7 见 [8])。 如果 $ab$ 不是 *DT* 边，且矩形 $R(a, b)$ 中除了 $a$ 和 $b$之外不含$P$ 的任何点，则以下性质成立。

1. $a$ 位于 $S_1$的 *W* 侧，
2. $b$ 位于 $S_k$的东侧，
3. 点 $h_1, \ldots, h_{k-1}$ 相对于 $R(a, b)$为高点，点 $l_1, \ldots, l_{k-1}$ 为低点，并且
4. 对任意 $1 < i < k$，

   - 或者 $T_i = \Delta(h_{i-1}, h_i, l_{i-1} = l_i)$，且点 $h_{i-1}$、$h_i$ 和 $l_{i-1} = l_i$ 按顺时针顺序位于 $S_i$ 的各边上，任意两点不在同一边上，并且 $(h_{i-1}, h_i)$ 是 $S_i$ 中的一条 *WN*、*WE* 或 *NE* 边；或者
   - $T_i = \Delta(h_{i-1} = h_i, l_{i-1}, l_i)$，且点 $h_{i-1} = h_i$、$l_i$ 和 $l_{i-1}$ 按顺时针顺序位于 $S_i$ 的各边上，任意两点不在同一边上，并且 $(l_{i-1}, l_i)$ 是 $S_i$中的一条 *WS*、*WE* 或 *SE* 边。

利用前面的引理，我们证明以下两个性质。

**Cl命题 9.** 对于任意 $1 \le i \le k$，$\max\{D(h_i, h_{i-1})\ D(l_i, l_{i-1})\ D(h_i, l_i)\} < w = D(a, b\quad)$.

证明。 首先，考虑三角形 $T_1 = \Delta(a, h_1, l_1)$。假设存在 $T_1$ 的某条边 $pq$，使得$D(pq) \ge w$。那么，命题 8 表明 $\ell(S_1) \ge D(pq) \ge w$。现在，根据引理 7 的性质 1 和 3， $a$ 位于 $S_1$ 的 W 侧，且相对于 $R(a, b)$， $h_1$ （分别为 $l_1$）较高（分别为较低）。因此，点 $b$ 包含在 $S_1$ 中。然而，这与定义矛盾，因为 $S_1$ 由三条 DT 边定义，因此为空。对于三角形 $T_k = \Delta(h_k, l_k, b)$ 的情形，证明类似。

现在考虑一个带有 $T_i$ $1 < i < k$ 的三角形，并假设 $T_i = \Delta(h_{i-1},\ h_i,\ l_{i-1} = l_i)$（另一种情况类似）。注意， $T_i$ 的顶点均不是 $a$ 或 $b$。再次假设对于某个

$T_i$ 的边 $pq$，有 hat $D(pq) \geq w$。再次根据命题 8，可得 $\ell(S_i) \geq D(pq) \geq w$。回顾引理 7 的性质 3，$h_i$（分别为 $l_i$）相对于 $R(a, b)$ 是高的（分别为低的），即 $h_i$ 和 $l_i$ 的 $x$-坐标位于区间 $[0, w]$ 内。由于 $h_i$ 和 $l_i$ 位于边长至少为 $w$ 的正方形边界上，这意味着 $S_i$ 必须包含 $a$ 或 $b$，这与 $S_i$ 的空正方形性质矛盾。 □

**断言 10**。对于任意 $1 \leq i \leq k$、$D(a, h_i) < D(a, b)$，以及 $D(a, l_i) < D(a, b)$。

证明。固定某个 $1 \leq i \leq k$，首先反设 $D(a, h_i) > D(a, b) = w$。现在，$D(a, h_i) = \max\{|x_{h_i}|, |y_{h_i}|\}$。然而，根据引理 7，我们知道 $h_i$ 是高的，即 $0 \leq x_{h_i} \leq w$，并且 $y_{h_i} > h$；同时，$l_i$ 是低的，即 $y_{l_i} < 0$。因此，$D(a, h_i) = y_{h_i} > w$。现在，考虑 $D(h_i, l_i) \geq y_{h_i} - y_{l_i} > w$，因为 $y_{l_i} < 0$。然而，这与断言 9 矛盾，断言 9 表明 $D(h_i, l_i) < w$。

现在，假设 $D(a, l_i) > D(a, b) = w$。现在，$D(a, l_i) = \max\{|x_{l_i}|, |y_{l_i}|\} = |y_{l_i}|$。这意味着 $y_{l_i} < -w$。现在，考虑 $D(h_i, l_i) \geq y_{h_i} - y_{l_i} > h + w > w$，这与断言 9 矛盾。 □

现在我们引入更多定义和术语。若三角形 $T_i$ 的顶点 $c$ 位于 $S_i$的东侧，则称其为东侧的 $S_i$。若 $T$ 中的一条边所对应线段的斜率在 $[-1, 1]$范围内，则称该边为缓斜的；否则称其为陡峭的。由引理 7以及一般位置假设可知，若 $T$ 中的一条边 $l_j h_j$ 是缓斜的，则 $l_j$ 或 $h_j$ 必为 $S_j$的东侧顶点。

**定义 3.** 假设$(a, b)$ 不是 *DT* 边，且矩形 $R(a, b)$ 中除 $a$ 和 $b$外不含 $P$中的任何点。若正方形 $S_j$，$1 \leq j \leq k$满足边 $l_j h_j$ 为缓斜边，则称其为归纳的 。东侧点 $c = h_j$ 或 $c = l_j$ 是 $S_j$的归纳点 。

现在，我们有以下由 Bonichon 等人 [8]给出的引理。

**引理 8 (**引理 9 出自 [8]**).** 假设 $ab$ 不是 *DT* 边，且矩形 $R(a, b)$ 中除 $a$和 $b$外不含 $P$中的任何点。假设点 $c = h_i$或 $c = l_i$的坐标满足 $w - x_c < |h - y_c|$。

*1.* 如果 $c = h_i$，从而 $w - x_{h_i} < y_{h_i} - h$，那么存在一个指标 $j$，使得 $i < j \leq k$，并且路径 $h_i$、$h_{i+1}$、……、$h_j$ 中的所有边分别都是相应正方形中的东北边，且 $w - x_{h_j} \geq y_{h_j} - h \geq 0$。*2.* 如果 $c = l_i$，从而 $w - x_{l_i} < h - y_{l_i}$，那么存在一个指标 $j$，使得 $i < j \leq k$，并且路径 $l_i$、$l_{i+1}$、……、$l_j$ 中的所有边分别都是相应正方形中的东北边，且 $w - x_{l_j} \geq y_{l_j} - h \geq 0$。

引理 *9*（引理 *8* 的部分表述，来自 [8]）。假设 *(a, b)* 不是 *DT* 边，并且矩形 *R(a, b)* 不包含 $P$ 中除 $a$ 和 $b$ 之外的任何点。如果 $S_j$ 是序列 $S_1$、$S_2$、……、$S_{k-1}$ 中第一个归纳正方形（如果存在）。如果 $h_j$ 是 $S_j$ 的归纳点，那么 $d_H(a, h_j) + (y_j - h) \leq 2h x_{h_j}$。否则，如果 $l_j$ 是 $S_j$ 的归纳点，那么 $d_H(a, l_j) - y_{l_j} \leq 2x_{l_j}$。

现在，我们可以证明引理 6 了。

引理 *6* 的证明。我们对 $D(a, b)$ 使用归纳法来证明该引理。注意，根据我们的假设，距离都是唯一的，因此我们只需考虑满足 $D(a, b) \leq 2$ 的点对 $a, b \in P$。

基础情形。考虑具有最小 $D(a, b)$ 值的点 $a, b \in P$。假设 $a = (0, 0)$，且 $b = (w, h)$，其中 $0 \leq h \leq w$ 且 $0 < w = D(a, b)$。由于 $a$ 和 $b$ 是最近点（按 $\ell_\infty$ 距离计算），以 $a$ 为西南顶点且内部不包含 $P$ 中其他点的最大正方形，其边界上必定有 $b$（根据命题9）。因此，$ab$ 是一条 DT 边。此外，由于 $D(a, b) \leq 2$，$ab \in E(G)$，这意味着 $ab \in E(H)$。最后，我们注意到 $\mathrm{d}_H(a, b) = D(a, b) \leq w + h \leq 3w + h$。

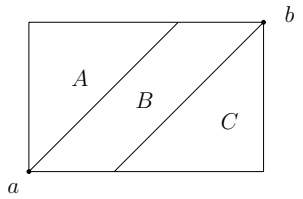


图6：引理 6 的证明中情形 1 所用的不同区域。

**归纳步骤**。考虑点 $a$ 和 $b$，并假设对于所有 $\ell_\infty$ 距离小于 $D(a,b) \le 2$ 的点对，该结论均成立。如果 $ab \in E(H)$，则按照前一种情况处理即可得到结果。因此，假设 $ab$ 不是 $T$ 中的一条边。

同样，我们假设 $a = (0, 0)$ 和 $b = (w, h)$，其中 $0 \le h \le w$ 且 $0 < w = D(a, b)$（注意，这一假设不失一般性，因为我们可以将平面平移并旋转整数倍的 $90°$，而且可以观察到，这两种操作都不会影响 $\ell_\infty$ 距离）。

情形 1a。假设 $P$ 中至少有一个点位于矩形 $R(a,b)$ 内。若存在一个点 $c \neq a, b$，使得 $0 \le y_c \le x_c$, $0 < x_c$, $0 \le h - y_c \le w - x_c$，且 $0 < w - x_c$。那么，$D(a,c) = x_c < w = D(a,b)$，并且 $D(b,c) = w - x_c < w = D(a,b)$ 图6中的区域B。因此，我们使用归纳假设，得到 $\mathsf{d}_H(a,c) \le 3x_c + y_c$，且 $\mathsf{d}_H(c,b) \le 3(w - x_c) + (h - y_c)$。此外，路径 $\pi_H(a,c)$（分别为 $\pi_H(c,b)$）中每条边的长度至多为 $D(a,c) < w$（分别为 $D(c,b) < w$）。将这两条路径连接起来，我们得到一条长度至多为 $2(x_c + w - x_c) + (y_c + h - y_c) = 2w + h$ 的路径 $\pi_H(a,b)$，且边的长度至多为 $w$。

情形 1b。现在，假设区域B中没有点，但 $R(a,b)$ 包含 $P$中的一个点。在这种情况下，令 $S_a$（分别令 $S_b$）为一个正方形，其西南角（分别为东北角）是 $a$（分别是 $b$），且其内部不包含 $P$ 中的其他点。于是，根据假设，存在一点 $c \neq a$， $b$，使得 $c$ 位于正方形 $S_a$ 或正方形 $S_b$的边界上。不失一般性，假设前一种情况成立（另一种情况类似）。注意， $c$ 属于图6中的区域A，即 $y_c > x_c > 0$，并且 $ac$ 是一条 DT 边。注意，$x_c > 0$，因为根据一般位置假设， $a$ 和 $c$ 不可能位于同一个空正方形的同一侧。于是，$\mathsf{d}_H(a,c) = D(a,c) \le x_c + y_c$。根据命题9，我们还可得到 $D(a,c) \le D(a,b)$。此外， $D(b,c) = w - x_c < w = D(b,c)$。因此，我们可以对点对 $b, c$应用归纳假设，从而得出存在一条路径 $\pi_H(c,b)$，满足 $\mathsf{d}_H(b,c) \le 2(w - x_c) + (h - y_c)$，且路径上的每条边的长度至多为 $D(b,c) < D(a,b)$。然后，将 $ac$ 与$\pi_H(c,b)$连接起来，我们便得到一条长度至多为 $x_c + y_c + 2(w - x_c) + h - y_c \le 2w + h$的路径，且每条边的长度至多为 $D(a,b)$。

**情形 2.** 假设 $R(a,b)$ 中除 $a$ 和 $b$之外不包含任何 $P$ 点。

**情况 2A.** 如果没有正方形 $S_1, S_2, \ldots, S_{k-1}$ 是归纳的，那么我们使用引理 10 得到 $\mathsf{d}_H(a,b) \le 2w$，且对应的路径 $\pi_H(a,b)$ 包含边 $uv$，使得 $D(u,v) < w$。该引理使用归纳假设证明，我们将在完成当前证明后再给出其证明。

否则，令 $S_i$ 为第一个归纳正方形。现在我们根据 $S_i$的归纳点考虑两种情况。

情形 2b。假设 $h_i$ 是 $S_i$ 的归纳点。由引理 7， $x_{ih} > 0$；由引理 8，存在指标 $j$，使得 $i \le j \le k$，且 $\pi_H(h_i, h_j) = (h_i, h_{i+1}, \ldots, h_j)$ 是一条长度至多为 $(x_{h_j} - x_{h_i}) + (y_{h_i} - y_{h_j})$ 的路径。此外，由断言 9 可知，对于 $i \le i' \le j - 1$，有 $D(h_{i'}, h_{i'+1}) < D(a,b)$。引理 8 还意味着 $w - x_{h_j} \ge y_{h_j} - h \ge 0$，而由引理 7， $x_{h_j} \ge x_{h_i}$。现在，或者 (1) $h_j = b$，此时可得 $\mathsf{d}_H(h_j, b) = 0 \le 2(w - x_{h_j}) + (y_{h_j} - h)$。在这种情况下，我们令 $\pi_H(h_j, b)$ 为一条长度为 0 的路径。(2) 否则， $h_j \neq b$，这意味着 $x_{h_j} \le x_{h_{k-1}} < x_b = w$；严格不等号是由一般位置假设得出的。

因此，$D(h_j,b) = \max\{w - x_{h_j}, y_{h_j} - h\} = w - x_{h_j} < w = D(a,b)$，且 $\delta(h_j,b) = y_{h_j} - h$。因此，根据归纳假设，我们得到存在一条路径 $\pi_H(h_j,b)$，其长度至多为 $2(w - x_{h_j}) + (y_{h_j} - h)$，且该路径的每条边的长度至多为 $D(a,b) \le 2$。现在，令 $\pi_H(a,b)$ 为通过拼接路径 $\pi_H(a,h_i)$ $\pi_H(h_i, h_j)$ 以及上文定义的 $\pi_{h_j,b}$ 得到的路径。由此可得

$$\begin{aligned}\mathsf{d}_H(a,b) &\le \mathsf{d}_H(a,h_i) + \mathsf{d}_H(h_i,h_j) + \mathsf{d}_H(h_j,b)\\ &\le 2x_{h_i} - (y_{h_i} - h) + (x_{h_j} - x_{h_i}) + (y_{h_i} - y_{h_j}) + 2(w - x_{h_j}) + (y_{h_j} - h)\\ &\le 2w \qquad\qquad (\text{Since } x_{h_i} \le x_{h_j})\end{aligned}$$

情形 2c。假设 $l_i$ 是 $S_i$ 的归纳点。根据引理 7，$x_{l_i} > 0$；根据引理 8，存在一个 $j$，满足 $i \le j \le k$，使得 $\pi_H(l_i,l_j) = (l_i, l_{i+1}, \ldots, l_j)$ 是一条长度至多为 $(x_{l_j} - x_{l_i}) + (y_{l_j} - y_{l_i})$ 的路径。此外，由断言 9 可知，对于 $i \le i' \le j-1$，$D(l_{i'}, l_{i'+1}) < D(a,b)$。引理 7 还意味着 $x_{l_j} \ge x_{l_i}$。然后，我们像之前一样考虑两种情况，即 (1) $l_j = b$，或 (2) $w - x_{l_j} > 0$。无论哪种情况，$D(l_j,b) < D(a,b)$。接着，仿照前一种情况的推导，我们可以得到一条路径 $\pi_H(l_j,b)$，其长度至多为 $2(w - x_{l_j}) + (h - y_{l_j})$，且路径中的每条边的长度至多为 $D(a,b)$。同样，通过连接路径 $\pi_H(a,l_i)$ $\pi_H(l_i,l_j)$ 和 $\pi_H(l_j,b)$，我们得到一条路径 $\pi_H(a,b)$，其中每条边的长度至多为 $D(a,b) \le 2$。由此可知，

$$\begin{aligned}\mathsf{d}_H(a,b) &\le \mathsf{d}_H(a,l_i) + \mathsf{d}_H(l_i,l_j) + \mathsf{d}_H(l_j,b)\\ &\le 2x_{l_j} + y_{l_i} + (x_{l_j} - x_{l_i}) + (y_{l_j} - y_{l_i}) + 2(w - x_{l_j}) + (h - y_{l_j})\\ &\le 2w + h\end{aligned}$$

□

**引理 10**。设 $a$、$b \in P$ 为一对点，满足：*(i)* $D(a,b) \le 2$；*(ii)* 矩形 $R(a,b)$ 中除了 $a$ 和 $b$ 外不包含 $P$ 的任何点；*(iii)* 没有正方形 $S_1, S_2, \ldots, S_{k-1}$ 是归纳的。那么，存在一条路径 $\pi_H(a,b)$，其长度至多为 $2w$，且路径 $\pi_H(a,b)$ 上的每条边 $uv$ 都满足 $D(u,v) < w$。

证明。为了证明此引理，我们需要引入额外的术语。

**术语**。设 $\delta_0 = 0$，对于 $1 \le i \le k$，令 $\delta_i$ 表示点 $a$ 与正方形 $S_i$ 东侧之间的水平距离。若 $\mathsf{d}_H(a, h_i) + \mathsf{d}_H(a,l_i) + P_{S_i}(h_i,l_i) \le 4\delta_i$，其中 $P_{S_i}(h_i,l_i)$ 是沿正方形 $S_i$ 的边按顺时针方向从 $h_i$ 移动到 $l_i$ 时路径的长度，则称正方形 $S_i$ 具有势。

根据引理陈述中的假设，我们将证明正方形 $S_1$、$S_2$、……、$S_{k-1}$、$S_k$ 都具有势，并利用这一事实构造具有所需性质的路径 $\pi_H(a, b)$。

根据引理 7，$a$ 位于 $S_1$ 的 W 侧，且 $\delta_1$ 是正方形 $S_1$ 的边长。那么，$\mathsf{d}_H(a, h_1) + \mathsf{d}_H(a, l_1) + P_{S_1}(h_1, l_1)$ 不超过 $S_1$ 的周长，即 $4\delta_1$。

现在，归纳地假设我们已证明正方形 $S_i$ 具有势，即 $\mathsf{d}_H(a, h_i) + \mathsf{d}_H(a, l_i) + P_{S_i}(h_i, l_i) \le 4\delta_i$。注意，根据假设，$S_i$ 不是归纳的。再归纳地假设路径 $\pi_H(a, h_i)$、$\pi_H(a, l_i)$ 分别见证 $\mathsf{d}_H(a, h_i)$ 和 $\mathsf{d}_H(a, l_i)$。接着，我们证明正方形 $S_{i+1}$ 具有势，并构造相应路径 $\pi_H(a, h_{i+1})$ 和 $\pi_H(a, l_{i+1})$，分别见证 $\mathsf{d}_H(a, h_{i+1})$ 和 $\mathsf{d}_H(a, l_{i+1})$。

正方形 $S_i$ 和 $S_{i+1}$ 都包含点 $l_i$ 和 $h_i$。由于 $S_i$ 不是归纳的，边 $l_ih_i$ 必定是陡峭的，即 $\Delta_x(l_i, h_i) < \Delta_y(l_i, h_i)$。首先，我们考虑 $x_{l_i} < x_{h_i}$ 的情况；$x_{l_i} > x_{h_i}$ 的情况可以用类似的方法证明。

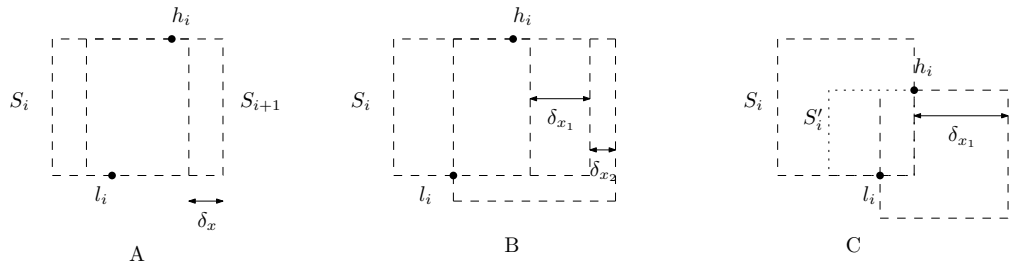


图7：引理 10 证明中三种情形的示意图

根据引理 7，$T_i = \Delta(h_{i-1},\ h_i,\ l_{i-1} = l_i)$，或者 $T_i = \Delta(h_{i-1} = h_i,\ l_{i-1},\ l_i)$；并且当沿顺时针方向从 $l_i$ 移动到 $h_i$时，$S_i$有一条边位于 $l_i$ 和 $h_i$ 所在的边之间。由引理 7和$x_{l_i} < x_{h_i}$，我们得出 $l_i$ 位于正方形 $S_i$的南侧，而 $h_i$ 位于其北侧或东侧。

如果 $h_i$ 位于 $S_i$ 的北侧，并且由于 $x_{l_i} < x_{h_i}$、$h_i$ 也必须位于 $S_{i+1}$ 的北侧，因此 $l_i$ 在 $S_{i+1}$ 边界上的位置有两种可能。如果 $l_i$ 位于 $S_{i+1}$ 的南侧，则如图7（A）所示，$S_{i+1}$ 通过长度为 $\delta_x = \delta_{i+1} - \delta_i$ 的水平平移得到。于是，

$$P_{S_{i+1}}(h_i, l_i) - P_{S_i}(h_i, l_i) = 2\delta_x = 2(\delta_{i+1} - \delta_i) \tag{27}$$

否则，$l_i$ 位于 $S_{i+1}$ 的 W 侧。令 $S'_i$ 为以 $l_i$ 为其西南角、通过对 $S_i$ 进行水平平移得到的正方形。令 $\delta_{x_1}$ 表示水平平移的长度，并令 $\delta_{x_2}$ 表示如图7（B）所示的 $S'_i$ 与 $S_{i+1}$ 的边长之差。于是，$\delta_{x_1} + \delta_{x_2} = \delta_{i+1} - \delta_i$，并且

$$P_{S_{i+1}}(h_i, l_i) - P_{S_i}(h_i, l_i) = 2\delta_{x_1} + 4\delta_{x_2} \le 4(\delta_{i+1} - \delta_i) \tag{28}$$

如果 $h_i$ 位于 $S_i$的东侧，那么令 $S'_i$ 为与 $S_i$共用其东南角、并以$h_i$ 为其东北角的正方形，如图7（C）所示。由于 $l_ih_i$ 是陡峭的，$l_i$ 位于 $S'_i$的南侧。然后，利用前一种情况的分析，我们得到 $P_{S_{i+1}}(h_i, l_i) - P_{S'_i}(h_i, l_i) \le 4(\delta_{i+1} - \delta_i)$；又由于 $P_{S'_i}(h_i, l_i) = P_{S_i}(h_i, l_i)$，因此我们得出结论，以下不等式在所有情况下均成立。

$$P_{S_{i+1}}(h_i, l_i) - P_{S_i}(h_i, l_i) \le 4(\delta_{i+1} - \delta_i) \tag{29}$$

现在，由于 $S_i$ 具有势，

$$\begin{aligned}
&\mathsf{d}_H(a, h_i) + \mathsf{d}_H(a, l_i) + P_{S_{i+1}}(h_i, l_i) \\
&= \mathsf{d}_H(a, h_i) + \mathsf{d}_H(a, l_i) + P_{S_i}(h_i, l_i) + (P_{S_{i+1}}(h_i, l_i) - P_{S_i}(h_i, l_i)) \\
&\le 4\delta_i + 4(\delta_{i+1} - \delta_i) = 4\delta_{i+1}
\end{aligned} \tag{30}$$

现在，假设 $T_{i+1} = \Delta(h_i,\ h_{i+1},\ l_i = l_{i+1})$。那么，$h_ih_{i+1}$是 DT 中的一条边。因此，$h_ih_{i+1}$是 $H$ 中的一条边。根据引理 7，$h_{i+1}$ 位于 $S_{i+1}$ 的边界上，从 $h_i$ 顺时针移向 $l_i$时，处于 $h_i$ 和 $l_i$ 之间。于是，根据三角不等式，$D(h_i,\ h_{i+1}) \le P_{S_{i+1}}(h_i,\ h_{i+1})$。接着，

$$\begin{aligned}
&\mathsf{d}_H(a, h_{i+1}) + \mathsf{d}_H(a, l_{i+1}) + P_{S_{i+1}}(h_{i+1}, l_{i+1}) \\
&\le \mathsf{d}_H(a, h_i) + D(h_i, h_{i+1}) + \mathsf{d}_H(a, l_i) + P_{S_{i+1}}(h_{i+1}, l_i) \\
&\le \mathsf{d}_H(a, h_i) + \mathsf{d}_H(a, l_i) + P_{S_{i+1}}(h_i, l_i) \\
&\le 4\delta_{i+1} \qquad \text{(From Equation (30))}
\end{aligned}$$

此外，根据断言 9，$D(h_i,\ h_{i+1}) < D(a,\ b)$。因此，将边 $h_ih_{i+1}$ 追加到归纳构造的路径 $\pi_H(a,\ h_i)$ 的末尾，可得到路径 $\pi_H(a,\ h_{i+1})$。由于 $l_i = l_{i+1}$，路径 $\pi_H(a,\ l_{i+1})$ 与 $\pi_H(a,\ l_i)$ 相同。当 $T_{i+1} = \Delta(h_i = h_{i+1},\ l_i,\ l_{i+1})$ 时，情形对称。因此，我们完成了归纳步骤。

因此，最后我们证明了正方形 $S_k$ 具有一个势，即 $\mathsf{d}_H(a, h_k) + \mathsf{d}_H(a, l_k) + P_{S_k}(h_k, l_k) \leq 4\delta_k = 4w$。回顾一下，$h_k = l_k = b$，这意味着 $P_{S_k}(h_k, l_k) = 0$。因此，归纳构造的路径 $\pi_H(a, b)$ 中至少有一条长度至多为 $2\delta_k = 2w$，且路径上的每条边$uv$ 都满足 $D(u, v) < w$。至此，引理得证。

□

我们以 本节主定理的证明作结，现将其重述如下 .

**定理 5.** 设 $P$ 是满足一般位置假设的点集，且 $G = (P, E)$ 是与 $P$相关联的单位正方形图，其中边 $uv$ 的权重等于$||u - v||_\infty$。那么，存在 $G$ 的一个平面子图 $H$，使得对于任意两个点 $a, b \in P$，$\mathsf{d}_G(a, b) \leq \mathsf{d}_H(a, b) \leq 3\mathsf{d}_G(a, b)$。

证明。考虑两个点 $a$ 和 $b \in P$，并令 $\pi_G(a, b) = (a = v_0, v_1, v_2, \ldots, v_t = b)$ 表示 $G$ 中的一条最短路径，其中对于 $0 \leq i < t$, $v_iv_{i+1} \in E(G)$，有 $\mathsf{d}_G(v_i, v_{i+1}) = D(v_i, v_{i+1})$。由于 $H$ 是德劳内三角剖分的子图，因此它是平面的；根据构造，它同时也是 $G$ 的子图。现在，对每条边 $v_iv_{i+1}$ 应用引理 6，我们知道在 $H$ 中存在一条路径 $\pi_H(v_i, v_{i+1})$，其长度至多为 $3D(v_i, v_{i+1}) = 3\mathsf{d}_G(v_i, v_{i+1})$。容易看出，路径 $\pi_H(a, b)$ 是通过连接所有此类路径，并在必要时对边进行捷径处理而得到的，它满足定理陈述中的条件。

□

## 6 结论与开放问题

我们针对 $k$-聚类问题，在多种几何图度量上获得了首个规模独立于 $n$ 的核心集，例如单位圆盘和正方形的加权相交图。单元圆盘图（或单位正方形图）可以包含任意大的团，即它们可以是（局部）稠密的。因此，据我们所知，这是首个针对稠密图族上的最短路径度量、且规模较小（即独立于 $n$）的核心集构造。由于这类度量固有的“混合”性质，我们的核心集构造必须谨慎处理该度量局部欧几里得而全局稀疏的特性。

我们认为，我们工作的贡献还具有概念层面的意义，因为我们借助 Cohen-Addad 等人 [19] 提出的通用框架，将获得小规模核心集所需的度量几何结构性质“抽象出来”。这些结构性质同样适用于 $\ell_p$-范数加权单位圆盘图和单位正方形图。此外，通过适当修改构造方法，我们还可以处理由有界度单位圆盘图诱导的跳数度量（即无权边）。因此，我们为所有这些图族中的 $k$-聚类问题都获得了小规模核心集。为了得到关于单位正方形图的结果，我们证明了这些图存在一个 3-伸展平面生成器；这一结果本身可能具有独立意义，并可进一步应用。

一个最自然的问题是，寻找更多满足本研究所确定的结构性质的几何交图族实例。$\mathbb{R}^2$ 中的圆盘图和 $\mathbb{R}^d$ 中的单位球图（对于常数 $d \geq 3$）是单位圆盘图的两种正交推广，因此可能是最显而易见的候选对象。然而，目前尚不清楚这些图族是否存在常数伸展平面伸展子图。作为一个中间步骤，研究亏格有界曲面 $\Sigma$ 上的单位圆盘图或许会很有意思。在这种情况下，更自然的问题或许是：这类图是否存在常数伸展伸展子图，且该子图还能嵌入 $\Sigma$ 上（这是对平面性的放宽）。它

或许可以将我们的框架扩展到这种放宽的设定，从而也为这类几何交图族构造更小的核心集。

## 参考文献

[1] Sara Ahmadian、Ashkan Norouzi-Fard、Ola Svensson 和 Justin Ward。通过原始对偶算法，为 k-均值和欧几里得 k 中位数提供更好的保证。载于 Chris Umans（编），第*58*届 *IEEE* 计算机科学基础年会，*FOCS 2017*，美国加利福尼亚州伯克利，*2017*年*10*月*15*日至*17*日，第61–72页。IEEE计算机学会，2017。[2] Sanjeev Arora、Prabhakar Raghavan 和 Satish Rao。欧几里得 k 中位数及相关问题的近似方案。载于第三十届 *ACM* 年度计算理论研讨会论文集，STOC '98，第106–113页，美国纽约，1998。ACM。网址：http://doi.acm.org/10.1145/276698.276718，doi:10.1145/276698.276718。[3]Daniel N. Baker、Vladimir Braverman、Lingxiao Huang、Shaofeng H.-C. Jiang、Robert Krauthgamer 和 Xuan Wu。有界树宽图上的聚类核心集。载于第*37*届国际机器学习会议论文集，*ICML 2020*，*2020*年*7*月*13*日至*18*日，线上会议，《机器学习研究论文集》第119卷，第569–579页。PMLR，2020。网址：http://proceedings.mlr.press/v119/baker20a.html。[4] Hari Balakrishnan、Christopher L Barrett、VS Anil Kumar、Madhav V Marathe 和 Shripad Thite。距离-2 匹配问题及其与自组织无线网络介质访问控制层容量的关系。*IEEE*通信领域精选期刊，22(6):1069–1079，2004。[5]Luca Becchetti、Marc Bury、Vincent Cohen-Addad、Fabrizio Grandoni 和 Chris Schwiegelshohn。k-均值的无关降维：超越子空间和 Johnson-Lindenstrauss 引理。载于第*51*届 *ACM SIGACT* 年度计算理论研讨会论文集，第1039–1050页，2019。[6] Amariah Becker、Philip N Klein 和 David Saulpic。有界高速公路维数下 k 中心、k 中位数和容量受限车辆路径问题的多项式时间近似方案。载于第*26*届欧洲算法年会（*ESA 2018*）。Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik，2018。[7] Ahmad Biniaz。单位圆盘图的平面跳数稀疏化器：更简单、更优。《计算几何》，89:101622，2020。doi:10.1016/j.comgeo.2020.101622。[8] Nicolas Bonichon、Cyril Gavoille、Nicolas Hanusse 和 Ljubomir Perkovic。l1 和 $l_\infty$ 德劳内三角剖分的紧致伸展因子。《计算几何》，48(3):237–250，2015。doi: 10.1016/j.comgeo.2014.10.005。[9] Vladimir Braverman、Shaofeng H-C Jiang、Robert Krauthgamer 和 Xuan Wu。排除子式图及更广泛图类上的聚类核心集。载于*2021*年 *ACM-SIAM* 离散算法研讨会论文集（*SODA*），第2679–2696页。SIAM，2021。[10] Jaroslaw Byrka、Thomas W. Pensyl、Bartosz Rybicki、Aravind Srinivasan 和 Khoa Trinh。预算约束优化中 k-中位数的改进近似与正相关。*ACM*算法汇刊，13(2):23:1–23:31，2017。

[11] Timothy M. Chan 和 Dimitrios Skrepetos。带权单位圆盘图中的近似最短路径和距离预言机。《计算几何学杂志》，10(2):3–20，2019。doi:10.20382/jocg. v10i2a2。[12] Moses Charikar、Sudipto Guha、 ́Eva Tardos 和 David B. Shmoys。用于 k 中位数问题的常数因子近似算法（扩展摘要）。载于第*31*届 *ACM* 年度计算理论研讨会论文集，第1–10页，1999。[13] Moses Charikar 和 Shi Li。用于 k 中位数问题的相关LP舍入方法。载于 Artur Czumaj、Kurt Mehlhorn、Andrew M. Pitts 和 Roger Wattenhofer（编），自动机、语言与编程——第*39*届国际计算机科学与应用数学研讨会（*ICALP 2012*），英国 *Warwick*，*2012*年*7*月*9*日至*13*日，论文集第一部分，《计算机科学讲义》第7391卷，第194–205页。Springer，2012。[14] Ke Chen。关于度量空间和欧几里得空间中 k 中位数及 k-均值聚类的核心集及其应用。《SIAM计算杂志》，39(3):923–947，2009。[15] Vincent Cohen-Addad、Andreas Emil Feldmann 和 David Saulpic。倍增度量中聚类的近线性时间近似方案。《ACM学报》，68(6):1–34，2021。[16] Vincent Cohen-Addad、Anupam Gupta、Amit Kumar、Euiwoong Lee 和 Jason Li。k 中位数和 k-均值的紧致固定参数可解近似。载于 Christel Baier、Ioannis Chatzigiannakis、Paola Flocchini 和 Stefano Leonardi（编），第*46*届自动机、语言与编程国际研讨会，*ICALP 2019*，希腊 *Patras*，*2019*年*7*月*9*日至*12*日，LIPIcs 第132卷，第42:1–42:14页。Schloss Dagstuhl - Leibniz-Zentrum f¨ur Informatik，2019。doi: 10.4230/LIPIcs.ICALP.2019.42。[17] Vincent Cohen-Addad、Philip N. Klein 和 Claire Mathieu。局部搜索为欧几里得度量和无小图度量中的 k-均值与 k 中位数提供近似方案。《SIAM计算杂志》，48(2):644–667，2019。doi:10.1137/17M112717X。[18]Vincent Cohen-Addad、Kasper Green Larsen、David Saulpic 和 Chris Schwiegelshohn。迈向 k 中位数和 k-均值核心集的最优下界。载于 Stefano Leonardi 和 Anupam Gupta（编），*STOC '22*：第*54*届 *ACM SIGACT* 年度计算理论研讨会，意大利罗马，*2022*年*6*月*20*日至*24*日，第1038–1051页。ACM，2022。doi: 10.1145/3519935.3519946。[19] Vincent Cohen-Addad、David Saulpic 和 Chris Schwiegelshohn。用于聚类的新核心集框架。载于 Samir Khuller 和 Virginia Vassilevska Williams（编），*STOC '21*：第*53*届 *ACM SIGACT* 年度计算理论研讨会，线上会议，意大利，*2021*年*6*月*21*日至*25*日，第169–182页。ACM，2021。doi:10.1145/3406325.3451022。[20] Dan Feldman 和 Michael Langberg。用于近似和数据聚类的统一框架。载于 Lance Fortnow 和 Salil P. Vadhan（编），第*43*届 *ACM* 计算理论研讨会论文集，*STOC 2011*，美国加利福尼亚州圣何塞，*2011*年*6*月*6*日至*8*日，第569–578页。ACM，2011。doi:10.1145/1993636.1993712。[21] Dan Feldman 和 Michael Langberg。用于近似和数据聚类的统一框架。载于第四十三届 *ACM* 年度计算理论研讨会论文集，第569–578页，2011。

[22]Dan Feldman、Melanie Schmidt 和 Christian Sohler。将大数据转化为微型数据：用于 k-均值、主成分分析和投影聚类的常数大小核心集。*SIAM*计算杂志，49(3):601–657，2020。[23] Gereon Frahling 和 Christian Sohler。动态几何数据流中的核心集。载于第三十七届年度 *ACM*计算理论研讨会论文集，第209–217页，2005年。[24] Zachary Friggstad、Mohsen Rezapour 和 Mohammad R. Salavatipour。局部搜索为倍增度量中的 k-均值提供多项式时间近似方案。*SIAM*计算杂志，48(2):452–480，2019年。doi: 10.1137/17M1127181。[25] Jie Gao 和 Li Zhang。单位圆盘图度量的良分离对分解及其应用。*SIAM*计算杂志，35(1):151–169，2005年。[26] Sariel Har-Peled 和 Akash Kushal。更小的 k-中位数和 k-均值聚类核心集。离散计算几何，37(1):3–19，2007年。[27] Sariel Har-Peled 和 Soham Mazumdar。关于 k-均值和 k-中位数聚类的核心集。载于 L´aszl´o Babai（编辑），第*36*届年度 *ACM*计算理论研讨会论文集，美国伊利诺伊州芝加哥，*2004*年*6*月*13*日至*16*日，第291–300页。ACM，2004年。[28] Lingxiao Huang、Shaofeng H-C Jiang、Jian Li 和 Xuan Wu。倍增度量中带离群点聚类的 ε-核心集。载于*2018*年 *IEEE*第*59*届计算机科学基础年度研讨会（*FOCS*），第814–825页。IEEE，2018年。[29] Lingxiao Huang 和 Nisheeth K. Vishnoi。欧几里得空间中的聚类核心集：重要性采样几乎是最优的。载于 Konstantin Makarychev、Yury Makarychev、Madhur Tulsiani、Gautam Kamath 和 Julia Chuzhoy（编辑），第*52*届年度 *ACM SIGACT*计算理论研讨会论文集，*STOC 2020*，美国伊利诺伊州芝加哥，*2020*年*6*月*22*日至*26*日，第1416–1429页。ACM，2020年。doi:10.1145/3357713.3384296。[30] Kamal Jain 和 Vijay V. Vazirani。使用原始-对偶框架和拉格朗日松弛求解度量设施选址和$k$-中位数问题的近似算法。*J. ACM*，48(2):274–296，2001年。[31] Ken-ichi Kawarabayashi、Christian Sommer 和 Mikkel Thorup。无向平面图中近似距离的更紧凑预言机。载于 Sanjeev Khanna（编辑），第二十四届年度 *ACM-SIAM*离散算法研讨会论文集，*SODA 2013*，美国路易斯安那州新奥尔良，*2013*年*1*月*6*日至*8*日，第550–563页。SIAM，2013年。doi:10.1137/1. 9781611973105.40。[32] Fabian Kuhn、Tim Nieberg、Thomas Moscibroda 和 Rogert Wattenhofer。自组织网络和传感器网络的局部近似方案。载于*2005*年移动计算基础联合研讨会论文集，第97–103页，2005年。[33] Amit Kumar、Yogish Sabharwal 和 Sandeep Sen。任意维度聚类问题的线性时间近似方案。*J. ACM*，57(2):5:1–5:32，2010年。[34] Michael Langberg 和 Leonard J Schulman。用于积分的通用 $\varepsilon$-近似器。载于第二十一届年度 *ACM-SIAM*离散算法研讨会论文集，第598–607页。SIAM，2010年。

[35] Emmanuelle Lebhar 和 Zvi Lotker。单位圆盘图与物理干扰模型：将各个部分结合起来。载于*2009*年*IEEE*并行与分布式处理国际研讨会，第1–8页。IEEE，2009年。[36] Shi Li 和 Ola Svensson。通过伪近似逼近 k-中位数。*SIAM*计算杂志，45(2):530–547，2016年。[37] Xiang-Yang Li、Gruia Calinescu 和 Peng-Jun Wan。为自组织无线网络分布式构造平面伸展子图并进行路由。载于第二十一届年度*IEEE*计算机与通信学会联合会议论文集，第3卷，第1268–1277页。IEEE，2002年。[38] Xiang-Yang Li、Wen-Zhan Song 和 Yu Wang。具有非均匀传输范围的自组织无线网络的高效拓扑控制。《无线网络》，11(3):255–264，2005年。[39] Konstantin Makarychev、Yury Makarychev 和 Ilya P. Razenshteyn。Johnson-Lindenstrauss变换在$k$-均值和$k$-中位数聚类中的性能。载于 Moses Charikar 和 Edith Cohen（编辑），第*51*届年度 *ACM SIGACT*计算理论研讨会论文集，*STOC 2019*，美国亚利桑那州菲尼克斯，*2019*年*6*月*23*日至*26*日，第1027–1038页。ACM，2019年。doi:10.1145/3313276.3316350。[40] Ramgopal R Mettu 和 C Greg Plaxton。近似聚类的最优时间界。《机器学习》，56(1):35–60，2004年。[41] Frank Schulz。传感器网络和自组织网络建模。载于《传感器与自组织网络算法》，第21–36页。Springer，2007年。[42] Amin Shahraki、Amir Taherkordi、Øystein Haugen 和 Frank Eliassen。无线传感器网络中的聚类目标：综述与研究方向分析。《计算机网络》，180:107376，2020年。[43] Christian Sohler 和 David P Woodruff。用于 k-中位数和子空间近似的强核心集：告别维度。载于*2018*年*IEEE*第*59*届计算机科学基础年度研讨会（*FOCS*），第802–813页。IEEE，2018年。