arXiv:2611.32524v1 [cs.DS] 24 Nov 2026

加权谱稀疏化草图维度的紧致超线性下界

Yanyu Chen* Yu Chen† Mingji Yang‡

摘要

我们证明，任何能够计算具有多项式有界权重的 n 个顶点加权图的 $(1\pm O(1))$-谱稀疏化器的线性草图，其维数都必须为 $\Omega(n^{6/5})$。此前，Chen、Khanna 和 Li（FOCS 2022）针对关联草图给出了 $\widetilde{O}(n^{6/5})$ 的草图维度上界和 $\Omega(n^{21/20-o(1)})$ 的下界，其中每次测量只考察与单个顶点关联的边。我们的下界移除了这一限制，并且与上界 $\widetilde{}$ 相比仅相差多对数因子。结合他们针对加权割稀疏化的 $O(n)$ 上界，我们的结果还给出了割稀疏化与谱稀疏化之间的紧致分离。

在我们的困难分布中，我们将顶点随机排列到一个圈上的各个块中，并构造一个完全图，使其边权重随循环块距离的增加而减小。我们将问题归约为证明：与按边权重缩放的其他列相比，连接相对块的边所对应的草图列具有较小的杠杆分数。对于关联草图，我们通过块循环上的路由分解得到这一界限；该分解在相距较远的块之间仅使用少量列。随后，我们额外加入一个团，将任意草图归约为关联草图，同时使期望维数与原始维数呈线性关系。

1 引言

图稀疏化旨在将一个图压缩为一个稀疏图，同时近似保留原图的某些性质。它是图算法中的一项基础工具，可应用于近似最小割和最大流 [She13；KLOS14；BK15；Pen16]，求解拉普拉斯系统 [ST14；JS25]，以及图聚类 [CSWZ16]。图稀疏化的两个重要概念是 谱稀疏化 [ST11]，，它近似保留图拉普拉斯矩阵的每个二次型；以及割稀疏化 [BK96]，，它近似保留每个割的权重。

我们首先介绍谱稀疏化和割稀疏化的定义。对于 $\boldsymbol{w}\in\mathbb{R}^{\binom{n}{2}}_{\geq 0}$，权重向量为 $\boldsymbol{w}$ 的加权图的拉普拉斯矩阵为

$$\mathbf{L}_{\boldsymbol{w}} \coloneqq \sum_{uv\in\binom{[n]}{2}} w_{uv}(\boldsymbol{e}_u-\boldsymbol{e}_v)(\boldsymbol{e}_u-\boldsymbol{e}_v)^\top.$$

^我们称对称矩阵 $\mathbf{L}$ 是 $L_{\boldsymbol{w}}$ 的 $(1\pm\varepsilon)$-**谱稀疏化器**，如果

$$(1-\varepsilon)\boldsymbol{x}^\top\mathbf{L}_{\boldsymbol{w}}\boldsymbol{x}\leq\boldsymbol{x}^\top\widehat{\mathbf{L}}\,\boldsymbol{x}\leq(1+\varepsilon)\boldsymbol{x}^\top\mathbf{L}_{\boldsymbol{w}}\boldsymbol{x}\qquad\text{for all }\boldsymbol{x}\in\mathbb{R}^n.$$

*新加坡国立大学。电子邮箱：yanyu.chen@u.nus.edu。
†新加坡国立大学。电子邮箱：yu.chen@nus.edu.sg .
‡中国人民大学。电子邮箱：kyleyoung@ruc.edu.cn。

割稀疏化只要求对 $\boldsymbol{x} \in \{0,1\}^n$满足这一不等式。因此，每个谱稀疏化器也是割稀疏化器。尽管谱稀疏化的要求更强，但对于常数近似误差，这两种概念都存在大小近线性的稀疏化器。Spielman 和 Srivastava [SS11]展示了如何在近线性时间内构造具有$(1\pm\varepsilon)$-谱稀疏化器，其包含 $O\left(n\log n/\varepsilon^2\right)$ 条边；Batson、Spielman 和 Srivastava [BSS12]证明了存在具有 $O\left(n/\varepsilon^2\right)$条边的此类稀疏化器。

在许多计算场景中，这两种稀疏化也具有相似的复杂度界，包括顺序构造 [BK15；SS11], 并行和分布式算法 [KX16],以及全动态算法 [ADKKP16]。这些相似的界表明，在许多场景中，割稀疏化与谱稀疏化的复杂度之间存在密切联系。

然而，我们证明，对于加权图，在线性草图模型中，谱稀疏化所需的测量数量比割稀疏化多一个多项式因子。具体而言，对于固定常数$\tilde{}$近似误差， $O(n)$ 测量足以实现割稀疏化， [CKL22], 而谱稀疏化则需要$\Omega\left(n^{6/5}\right)$ 个测量，即使权重比有多项式界。在该模型中，维数为 N 的随机线性草图对矩阵$\mathbf{A} \in \mathbb{R}^{N\times\binom{n}{2}}$ 进行采样，且与输入边权重向量 $\boldsymbol{w}$无关。随后，恢复算法根据$\mathbf{A}$和 N 测量$\mathbf{A}\boldsymbol{w}$计算稀疏化结果。

这一模型的一个动机来自流算法：图以边权更新序列的形式到达，目标是在单遍处理过程中使用有限空间计算稀疏化结果。线性草图非常适合转闸流，因为这类流允许边权增加和减少。对于每次更新，我们将权重更新向量乘以草图矩阵，并将结果加到存储的测量向量中，而无需存储图本身。反过来，在流长度和值域满足适当假设的情况下，通常的单遍转闸算法可以由具有相当空间复杂度的线性草图来模拟 [LNW14；AHLW16；KP20；JLY26]。这种联系使线性草图成为证明流空间下界的自然起点。

对于无权图，线性草图在割稀疏化和谱稀疏化中都能达到近线性维数。Ahn、Guha 和 McGregor [AGM12] 率先研究了通过线性草图实现图稀疏化，并给出了用于割稀疏化的草图。Kapralov、Lee、Musco、Musco 和 Sidford [KLMMS17] 随后利用维数为 $\widetilde{O}\left(n/\varepsilon^2\right)$ 的草图获得了谱稀疏化器，后续工作则改进了恢复时间 [KMMMNST20]。对于常数 ε，这些结果表明，$\tilde{}$ $\tilde{}$对于 $O(n)$ 个线性测量，足以在转闸流中恢复一个具有 $O(n)$ 条边的稀疏化器。

针对加权图，在仅插入流中进行谱稀疏化的近线性空间算法早已为人所知 [KL13], ，近期工作则改进了对数因子 [CWXZ26]。对于允许插入与删除加权边的流，只要每次删除指定的权重与相应插入时相同，按权重对边分组也可以将无权草图扩展到加权图 [CKL22]。然而，一般的转向更新指定的是边权重的变化，这可能导致一条边在不同分组之间移动。因此，更新分组草图需要知道边的当前权重，所以这种方法无法直接得到用于加权稀疏化的小空间转向算法。按权重分组在 $\boldsymbol{w}$ 中也是非线性的，因此无法直接导出用于加权割稀疏化或谱稀疏化的线性草图。

因此，问题在于：针对割稀疏化和谱稀疏化的近线性草图维度界，对于加权图是否仍然成立。Chen、Khanna 和 Li [CKL22] 研究了这一问题，并得到了加权割稀疏化草图维度上界 $\widetilde{O}\left(n/\varepsilon^3\right)$；该上界最近被 Lin、Mirrokni 和 Woodruff $\widetilde{O}\left(n/\varepsilon^2\right)$ LMW [改进为 26]。对于加权谱稀疏化， [CKL22] 给出了草图维度上界 $\widetilde{O}\left(n^{6/5}/\varepsilon^4\right)$，其对 $w_{\max}/w_{\min}$输入图中允许的最大边权重与最小非零值之比

的依赖为多对数依赖。他们的谱算法将顶点采样与对顶点诱导子图中有效电阻的分析相结合，从而恢复在谱意义下重要的边。

另一方面， [CKL22] 进一步证明了，对于通过关联草图进行的加权谱稀疏化，其维数下界为 $\Omega\left(n^{21/20-o(1)}\right)$ 。关联草图是一类自然但受限的草图，其中每个测量仅支撑在与某个顶点关联的边上。该下界在这一类草图中确立了分离性，但对于加权谱稀疏化仍留下了两个问题：任意线性草图是否需要超线性维数，以及上界中的指数 $6/5$ 是否是最优的。

1.1 我们的结果

我们通过证明任意线性草图的 $\Omega\left(n^{6/5}\right)$ 下界，解决了这两个问题。即使最大边权重与最小非零边权重之比为 $O(n^5)$，该下界仍然成立。

定理 1.1。 任何维数为 N 、以至少 0.99 的概率在每个具有 $w_{\max}/w_{\min} = O\left(n^5\right)$ 的 n 个顶点的图上计算出 (1 ± 0.01)-谱稀疏化器的线性草图，都必须满足 $N = \Omega\left(n^{6/5}\right)$。

在近似误差为常数且权重比有多项式界的情况下，定理1.1 在多对数因子范围内匹配了 [CKL22] 的上界。它提高了他们下界中的指数，并去除了对关联草图的限制。即使恢复算法可以输出任意满足谱近似保证的对称矩阵，且对输出没有稀疏性要求，这一结果仍然成立。

结合 $\widetilde{O}(n)$ 在 [CKL22], 中关于加权割稀疏化的上界，我们的结果清晰揭示了割稀疏化与谱稀疏化之间的差距。要求构造谱稀疏化器，会使必要的草图维度从近线性增长至$\Omega\left(n^{6/5}\right)$。而 $\widetilde{O}(n)$ 无权谱稀疏化的上界 [KLMMS17] 进一步表明，边权重对于这一下界至关重要。

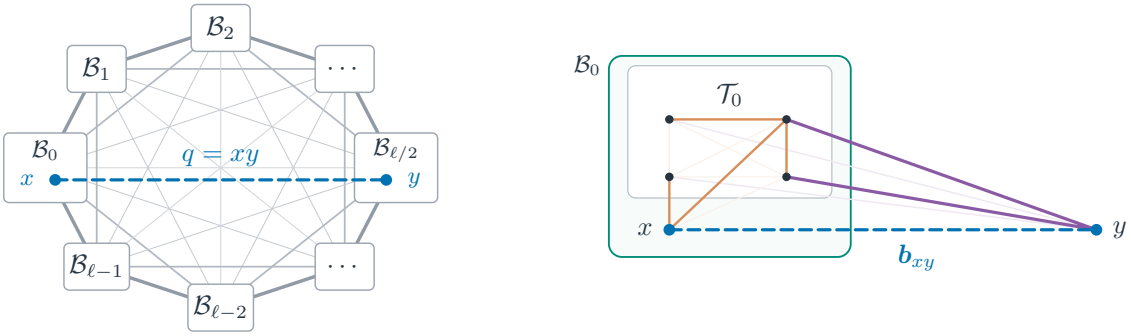
Chen、Shi 和 Trevisan [CST22], 通过比较边数受限时加权子图稀疏化器所能达到的近似误差，得出了割稀疏化与谱稀疏化之间另一种差距。我们的差距关注的是计算稀疏化器所需的线性测量数量。

与转向流的联系。我们的下界还揭示了一条为一般单遍转向流算法证明空间下界的途径。一系列研究表明，在对流作出各种假设的情况下，可以用线性草图模拟一般流算法 [LNW14; AHLW16; JLY26]。结合将实值草图下界转化为离散输入下界的技术 [GLWYZ25], ，这些模拟有望将我们的维数下界等结果推广为空间下界。这类转化取决于允许的流长度和坐标范围 [KP20]。如何将这些技术应用于我们的困难分布，并确定所得下界成立的具体流模型，仍是开放问题。

2 技术概述

我们采用一个标准框架来建立下界：构造两个可以通过谱稀疏化器加以区分的“困难”输入分布是和否，然后证明，当草图的行数过少时，通过线性草图得到的相应输出分布具有很小的全变差距离（TV distance）。我们首先证明关联草图的下界，其中每一行的支撑集都包含与某个单一顶点相 incident 的边。

困难分布。受 [CKL22], 中的构造启发，我们将顶点随机排列成大小相等的块 $\mathcal{B}_0$，……，$\mathcal{B}_{\ell-1}$ 并沿一个圈排列，如图 1a所示。该图是完全图，块内的边具有较大的基准权重，而不同块之间的权重则按其循环距离的四次方的倒数衰减。在 否情况下，我们将每个基准权重 w_e 乘以支撑集为 $[1/2, 3/2]$的独立噪声；在 是情况下，我们还将查询边 $q = xy$ 连接的 $\mathcal{B}_0$ 和 $\mathcal{B}_{\ell/2}$ 之间的边权增加一。一个在块内保持不变、并沿圈逐渐变化的势，可以揭示出一个足够精确的谱稀疏化器必须检测到的能量间隙。



(a) 环结构。　　(b) 第一表示。

图 1：**环与第一表示**。(a) 每个块都是一个团，且所有块对彼此相连；线条粗细表示基准权重的尺度。(b) 对于保留良好 x，我们用橙色块内列和紫色 $\mathcal{T}_0$–y 列表示虚线蓝色目标 $\boldsymbol{b}_{xy}$ （对应于根星项）。

从统计距离到杠杆分数。固定一个关联草图矩阵$\mathbf{B}$ 和顶点置换 π。查询增量通过其列 $\boldsymbol{b}_q$使草图发生偏移。直观地说，如果该列可以用其他列表示，且其系数相对于相应边权重较小，那么现有噪声中的微小偏移就能在草图中产生相同的变化，从而使查询增量难以检测。具体而言，通过界定这些偏移下的海林格距离，我们将全变差距离归约为 $\boldsymbol{b}_q$ 相对于$\mathbf{B}$ 的其他列、按基准权重加权的杠杆分数：

$$\mathrm{lev}_{\boldsymbol{b}_q}\Big(\big(\mathbf{B}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big) = \min\Bigg\{\sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}} \frac{f_e^2}{w_e^2} : \boldsymbol{b}_q = \sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}} f_e\boldsymbol{b}_e\Bigg\},$$

其中， $\mathcal{R}$ 是环顶点集， $\boldsymbol{w} = (w_e)$是基准权重向量。因此，只需使该杠杆分数在一处截断后关于 π 的期望较小即可。这里，截断允许表示偶尔失败。因此，根据上述杠杆分数表达式，我们的目标归结为：通过$\mathbf{B}$中的其他列对 $\boldsymbol{b}_q$ 进行分解，使其在相距较远的块之间使用较少的列，并控制系数，因为这些列的基准权重较小。

第一次分解。 我们首先尝试表示 $\boldsymbol{b}_{xy}$ 使用列表示位于单独预留集中的边 $\mathcal{T}_0 \subseteq \mathcal{B}_0$，以及从 x 和 y 连接到该预留集的边，如图 1b所示。完成此分解后，位于 $\mathcal{T}_0$ 以及从 x 连接到 $\mathcal{T}_0$ 的边仍位于块0中，只有从 y 连接到 $\mathcal{T}_0$ 的星跨越到相距甚远的块。

为获得这一表示，我们使用以下杠杆得分平均论证。矩阵各列的杠杆分数之和等于其秩。因此，如果列数远大于秩，则以均匀方式选取的一列以高概率可以用其他列表示，且系数范数较小（见引理3.1）。我们将这一事实应用于$\mathcal{T}_0 \cup \{x, y\}$上的团。关联性质确保其列仅支撑于分配给这些顶点的测量上

因此，随机标记将其期望秩限制为该顶点集大小乘以每个顶点的平均测量数量，而其列数则与该集合大小呈二次关系。因此，取一个足够大的保留集后，以高概率即可获得所需的表示，且系数范数较小。

更一般地，我们在每个块中指定一个保留集 $\mathcal{T}_i$，并对顶点v 和保留集 $\mathcal{T}_i$ 应用相同的论证，其中 $v, y \notin \mathcal{T}_i$。只要上述表示成立，我们就写作

$$\boldsymbol{b}_{vy} = \mathbf{B}_{v,\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{v,\mathcal{T}_i} + \mathbf{B}_{y,\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{y,\mathcal{T}_i} + \mathbf{B}_{\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{\mathcal{T}_i},$$

其中，这三个矩阵分别包含从 v 到保留集、从 y 到保留集以及保留集内部的列。我们将中间项称为根星项。

移动根星项。我们使用望远镜求和消去中间的根星项，将星从 y 与 $\mathcal{T}_0$ 之间的星，移至包含 y的块内。先考虑图 $\mathcal{T}_0$到 $\mathcal{T}_1$ 的转移，如图 2所示。我们引入一个辅助集 $\mathcal{S}_0 \subseteq \mathcal{B}_0$，它与 $\mathcal{T}_0 \cup \{x\}$不相交。利用相同的杠杆得分平均论证，我们将 x 的两个根星项之差表示为顶点 $v \in \mathcal{S}_0$对应差值的组合，且系数范数较小。对于每个这样的 v，将它的两个 $\boldsymbol{b}_{vy}$ 分解相减，即可消去该列。因此，紫色根星项之差可由橙色块内列以及 $\mathcal{S}_0$ 与 $\mathcal{T}_1$之间的深橙色跨块列表示。这就得到了一个跨块系数范数较小的转移，且没有表示列与 y相接。

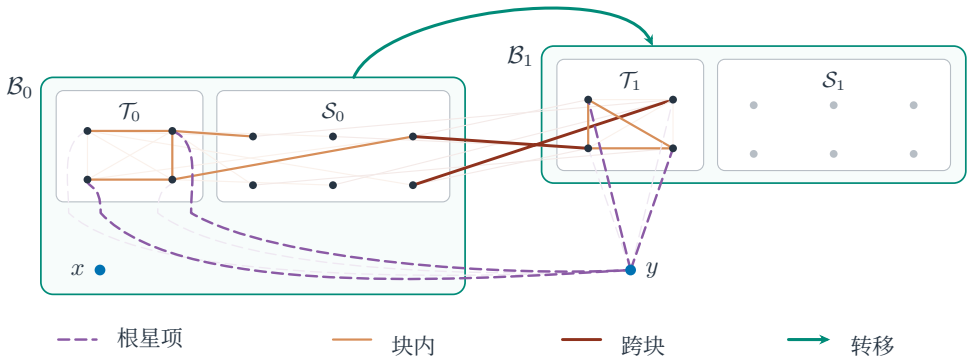


图 2：**从 $\mathcal{B}_0$ 到 $\mathcal{B}_1$的转移。**

这一构造可推广到从 $\mathcal{T}_i$ 到 $\mathcal{T}_j$ 的转移，其中在 $\mathcal{S}_i$中使用辅助顶点。连续的转移会逐项抵消，如图 3所示，直到剩余的根星项位于 y的块内。连同最初的块内项，它们构成了 $\boldsymbol{b}_{xy}$ 的表示，其跨块代价由所选转移决定。

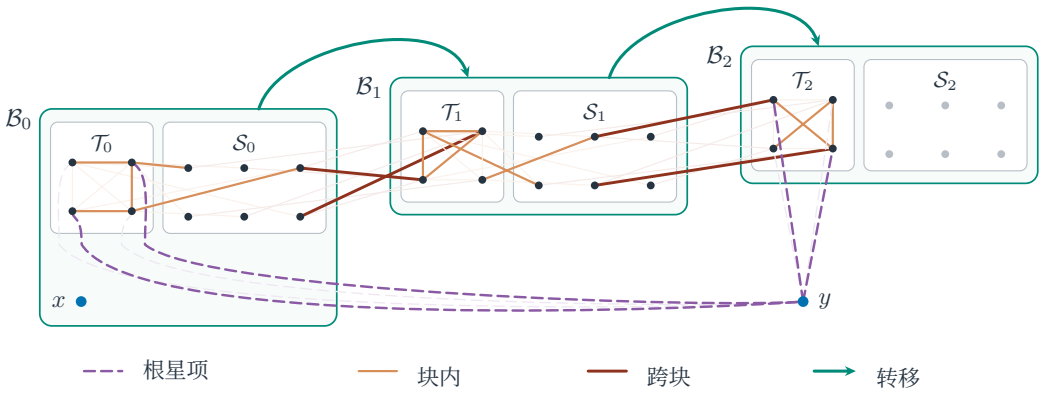


图 3：**Co ntinuing the transition to $\mathcal{B}_2$.** y 和 $\mathcal{B}_1$ 之间的根星项相互抵消。

在块循环上进行路由。 将成功的转移视为块之间的有向边，我们的目标是找到一条从块 0 到块 $\ell/2$ 的递增路径，使其总距离加权成本较低；见图 4。大多数块缺失的入向和出向转移很少，因此直观来看，当块被随机置换时，路径应该很少需要长距离跳跃。我们证明，尽可能多地经过这些块的路径具有所需的保证：其步长八次方之和的期望为 $O(\ell)$，与我们的边权缩放所产生的代价相匹配。这给出了关联草图$\Omega(n^{6/5})$ 下界所需的小杠杆分数界。

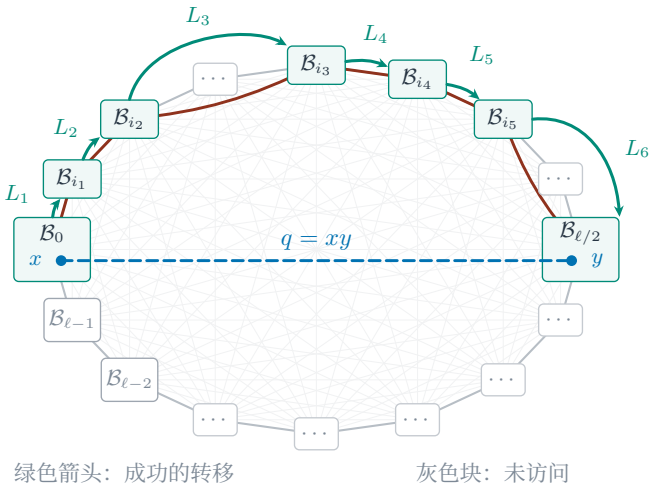


图4：**沿半循环路由。** 块循环上的一条示意路径。绿色箭头表示成功的转移，深橙色线条表示相应的跨块边。

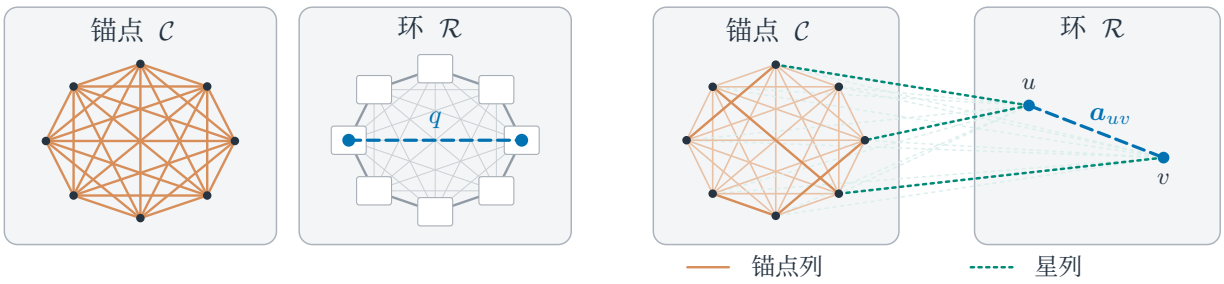
从任意草图到关联草图。 为处理任意草图矩阵$\mathbf{A}$，其维数为 N，我们添加一个带有重边权重、随机标记的锚点团 $\mathcal{C}$，并将其与环断开；见图5a。这保留了困难分布之间的谱区分。

W记 $\mathbf{A}_{\mathcal{C}}$ 为锚点列。我们的目标是分解大部分环列 a s

$$\boldsymbol{a}_{uv} = \mathbf{T}\boldsymbol{b}_{uv} + \mathbf{A}_{\mathcal{C}}\boldsymbol{c}_{uv},$$

其中 $\mathbf{T}$ 是一个公共变换，$\boldsymbol{b}_{uv}$ 仅支撑在分配给其端点的坐标上u, v，且 $\boldsymbol{c}_{uv}$ 的范数很小。因此，列 $\boldsymbol{b}_{uv}$ 构成一个关联草图 $\mathbf{B}$。

杠杆分数平均首先使用锚点列和两个端点星来表示大多数环列 $\mathbf{A}_{u,\ \mathcal{C}}$ 和 $\mathbf{A}_{v,\ \mathcal{C}}$，且系数范数较小；见图 5b。这些星仅用于分析，因为它们的边不属于输入图。



(a) 锚定图。　　（b）锚点/星表示。

图 5：**添加锚点。**（a）锚点团 $\mathcal{C}$ 与环 $\mathcal{R}$ 不连通，并且具有相同数量的顶点。（b）对于一条锚点良好环边 uv，长虚线蓝色列 $\boldsymbol{a}_{uv}$ 由橙色锚点列和短虚线绿色星列表示。

构造 T 时，对每个环顶点 v ，令 T_v 包含与 $\mathrm{A}_{v,\mathcal{C}}\mathrm{A}_{v}^{\top},_{\mathcal{C}} - \mathrm{A}_{\mathcal{C}}\mathrm{A}_{\mathcal{C}}^{\top}$的正特征值相对应的正交归一特征向量，并按其特征值的平方根缩放。将它们拼接为 $\mathrm{T} := \left[\mathrm{T}_v : v \in \mathcal{R}\right]$，并把与 T_v 相关的坐标分配给顶点 v。引理 3.2 中的杠杆分数刻画给出了所需的分解，其中$\boldsymbol{b}_{uv}$ 支撑在 u 和 v对应的坐标块上。利用西尔维斯特惯性定律和迹–杠杆界，我们证明 T 的列数期望值，因此也是 B 的期望维数，为 $O(N)$。

表示的提升。 由于**T** 对所有列都相同，因此， $\boldsymbol{b}_q$ 的一个小系数表示可以提升为 $\boldsymbol{a}_q$ 的一个表示，其中环系数保持不变，另加一个锚点残差。系数界和较大的锚点权重使残差的加权代价可以忽略不计。因此，关联矩阵论证得出：对于边权重有多项式上界的任意草图，下界为 $\Omega\left(n^{6/5}\right)$ 。

与 [CKL22]中的下界证明比较。 Chen、Khanna 和 Li [CKL22]采用随机标记的块环图、高斯权重和一条额外的跨越边。关键的是，在他们的块环图中，非相邻块之间没有边。在他们的证明中，他们将关联草图化归为一类结构更特殊的草图，称为带符号草图；对这类草图而言，相关的杠杆分数变为矩阵加权有效电阻。随后，他们利用扩张图分解的矩阵加权扩展以及顶点采样下扩张性保持的性质，对该电阻进行界定。

我们没有将关联草图化归为带符号草图，而是更直接地利用关联结构，通过杠杆分数平均和路由论证来界定一般的杠杆分数。此外，从路由论证的角度看，若将转移限制在相邻块之间，一旦转移失败，可供选择的方案就会很少。因此，我们将块环补全为一个团，使转移可以跨越非相邻块，并绕过失败的转移。直观而言，这使得构造成功的路由分解更容易；我们的分析表明，使用较长转移所产生的代价仍然可控。

我们还使用平移贝塔噪声，使每个实现的权重保持在正的多项式范围内。贝塔噪声在边界处的平滑衰减还给出了平移情形下的二次Hellinger距离界，从而使我们能够借助加权杠杆分数控制全变差距离。最后，锚点分解将我们的下界结果推广至任意草图。

论文结构。第3 节汇集了证明中使用的关于杠杆分数和统计距离的事实。第4 节构造了困难分布，并将不可区分性归约为杠杆分数界。第5 节通过分解与路由论证证明了关联草图的这一界。最后，第6节发展了锚点归约与提升论证，以证明任意线性草图的下界。

3 预备知识

基本记号。 对于任意正整数 n，我们记作 $[n] := \{1, 2, \ldots, n\}$。对于向量 $\boldsymbol{x} \in \mathbb{R}^m$，我们用 $\mathrm{diag}(\boldsymbol{x})$ 表示其第 i 个对角元素为 x_i 的 $m \times m$ 对角矩阵。对于矩阵 $\mathbf{M}$，我们用 $\|\mathbf{M}\|_2$ 表示其谱范数，用$\mathbf{M}^{\dagger}$ 表示其摩尔–彭若斯伪逆。对于具有 m 列且满足 $i \in [m]$ 的矩阵 $\mathbf{M}$ ，我们记 $\mathbf{M}_{-i}$ 为删除 $\mathbf{M}$ 的第 i 列后得到的矩阵。对于集合 S，我们用 $\binom{S}{2}$ 表示其中不同元素组成的无序对集合。对于图上的两个顶点 u, v ，我们用 uv 表示它们之间的无向边槽。

洛文纳序。 对于维数相同的实对称矩阵 $\mathbf{M}$,$\mathbf{N}$ ，若$\mathbf{N}-\mathbf{M}$ 为半正定（PSD），即对每个向量 $\boldsymbol{x}$，都有 $\boldsymbol{x}^\top(\mathbf{N}-\mathbf{M})\boldsymbol{x}\geq 0$，则记作 $\mathbf{M}\preceq\mathbf{N}$ 。对于任意维数兼容的矩阵$\mathbf{U}$ ，$\mathbf{M}\preceq\mathbf{N}$ 意味着$\mathbf{U}\mathbf{M}\mathbf{U}^\top\preceq\mathbf{U}\mathbf{N}\mathbf{U}^\top$。

3.1 杠杆分数

对于目标向量 $\boldsymbol{v}$ 以及与 $\boldsymbol{v}$维数相同的向量集合 $(\boldsymbol{u}_i)_{i=1}^m$ ，我们将 $\boldsymbol{v}$ 关于 $(\boldsymbol{u}_i)_{i=1}^m$ 的杠杆分数定义为

$$\operatorname{lev}_{\boldsymbol{v}}\big((\boldsymbol{u}_i)_{i=1}^m\big)\coloneqq\min\Big\{\sum_{i=1}^m c_i^2:\boldsymbol{v}=\sum_{i=1}^m c_i\boldsymbol{u}_i\Big\},\tag{1}$$

当 $\boldsymbol{v}\notin$值域$([\boldsymbol{u}_1\ \cdots\ \boldsymbol{u}_m])$时，将其解释为 $+\infty$。对于矩阵 $\mathbf{U}$，我们用 $\operatorname{lev}_{\boldsymbol{v}}(\mathbf{U})$ 表示 $\boldsymbol{v}$ 关于 $\mathbf{U}$ 的列的杠杆分数。对于 $\mathbf{U}=[\boldsymbol{u}_1\ \cdots\ \boldsymbol{u}_m]$，有 $\sum_{i=1}^m\operatorname{lev}_{\boldsymbol{u}_i}(\mathbf{U})=\operatorname{rank}(\mathbf{U})$，并且对于任意 $\boldsymbol{v}\in\operatorname{range}(\mathbf{U})$，还有 $\operatorname{lev}_{\boldsymbol{v}}(\mathbf{U})=\boldsymbol{v}^\top(\mathbf{U}\mathbf{U}^\top)^\dagger\boldsymbol{v}$ 。

引理 3.1。 设 $\mathbf{U}=[\boldsymbol{u}_1\ \cdots\ \boldsymbol{u}_m]$。如果从 $[m]$ 中均匀随机选取 I，则

$$\Pr\big\{\operatorname{lev}_{\boldsymbol{u}_I}(\mathbf{U}_{-I})>1\big\}\leq\frac{2\operatorname{rank}(\mathbf{U})}{m}.$$

证明。 我们首先将 $\operatorname{lev}_{\boldsymbol{u}_i}(\mathbf{U}_{-i})$ 与 $\operatorname{lev}_{\boldsymbol{u}_i}(\mathbf{U})$ 联系起来，其中 $i\in[m]$。假设 $\operatorname{lev}_{\boldsymbol{u}_i}(\mathbf{U}_{-i})$ 有限。如果将 $\boldsymbol{u}_i$ 在其自身表示中的系数固定为 a，则其余列必须表示 $(1-a)\boldsymbol{u}_i$。根据最小系数平方范数的齐次性，

$$\operatorname{lev}_{\boldsymbol{u}_i}(\mathbf{U})=\min_{a\in\mathbb{R}}\big\{a^2+(1-a)^2\operatorname{lev}_{\boldsymbol{u}_i}(\mathbf{U}_{-i})\big\}=\frac{\operatorname{lev}_{\boldsymbol{u}_i}(\mathbf{U}_{-i})}{1+\operatorname{lev}_{\boldsymbol{u}_i}(\mathbf{U}_{-i})},$$

当留一值为无穷大时，右侧按 1 解释。如果留一值为无穷大，则每个可行表示都必须使 $\boldsymbol{u}_i$ 系数为 1；否则，它将得到一个只使用其余列的 $\boldsymbol{u}_i$ 表示。令其他所有系数均为零，便可知普通杠杆分数等于 1，正如所声称的那样。

因此，$\operatorname{lev}_{\boldsymbol{u}_i}(\mathbf{U}_{-i})>1$ 等价于 $\operatorname{lev}_{\boldsymbol{u}_i}(\mathbf{U})>1/2$。将此与 $\sum_{i=1}^m\operatorname{lev}_{\boldsymbol{u}_i}(\mathbf{U})=\operatorname{rank}(\mathbf{U})$ 以及 I 的均匀性结合起来，即可证明该引理。 □

引理 3.2。 $\operatorname{lev}_{\boldsymbol{v}}(\mathbf{U})\leq 1$ 等价于 $\boldsymbol{v}\boldsymbol{v}^\top\preceq\mathbf{U}\mathbf{U}^\top$。

证明。 首先假设 $\operatorname{lev}_{\boldsymbol{v}}(\mathbf{U})\leq 1$。那么存在一个系数向量 $\boldsymbol{c}$，满足 $\|\boldsymbol{c}\|_2\leq 1$，使得 $\boldsymbol{v}=\mathbf{U}\boldsymbol{c}$。因此，$\boldsymbol{c}\boldsymbol{c}^\top\preceq\mathbf{I}$ 和 $\boldsymbol{v}\boldsymbol{v}^\top=\mathbf{U}\boldsymbol{c}\boldsymbol{c}^\top\mathbf{U}^\top\preceq\mathbf{U}\mathbf{U}^\top$。另一方面，假设 $\boldsymbol{v}\boldsymbol{v}^\top\preceq\mathbf{U}\mathbf{U}^\top$。这意味着

$$\big(\boldsymbol{v}^\top(\mathbf{U}\mathbf{U}^\top)^\dagger\boldsymbol{v}\big)^2=\boldsymbol{v}^\top(\mathbf{U}\mathbf{U}^\top)^\dagger\boldsymbol{v}\boldsymbol{v}^\top(\mathbf{U}\mathbf{U}^\top)^\dagger\boldsymbol{v}\leq\boldsymbol{v}^\top(\mathbf{U}\mathbf{U}^\top)^\dagger\mathbf{U}\mathbf{U}^\top(\mathbf{U}\mathbf{U}^\top)^\dagger\boldsymbol{v}=\boldsymbol{v}^\top(\mathbf{U}\mathbf{U}^\top)^\dagger\boldsymbol{v},$$

这给出 $\boldsymbol{v}^\top(\mathbf{U}\mathbf{U}^\top)^\dagger\boldsymbol{v}\leq 1$。关系 $\boldsymbol{v}\boldsymbol{v}^\top\preceq\mathbf{U}\mathbf{U}^\top$ 还意味着 $\boldsymbol{v}\in\mathbf{U}$的值域，因此 $\operatorname{lev}_{\boldsymbol{v}}(\mathbf{U})=\boldsymbol{v}^\top(\mathbf{U}\mathbf{U}^\top)^\dagger\boldsymbol{v}\leq 1$，证明完成。 □

3.2 总变差距离与Hellinger距离

对于定义在同一可测空间上的概率分布 P，Q，其全变差距离 （全变差距离）为

$$\operatorname{TV}(P,Q)\coloneqq\sup_A|P(A)-Q(A)|.$$

如果以概率 $1/2$ 从 P 中抽取样本，否则从 Q 中抽取样本，那么任何用于判断生成该样本的分布的随机化算法都满足

$$\Pr[\text{correct identification}] \le \frac{1}{2} + \frac{1}{2}\,\mathrm{TV}(P,Q). \tag{2}$$

对于随机变量 X，Y，我们用 $\mathrm{TV}(X,\ Y)$ 表示它们的分布之间的全变差距离，用 $\mathrm{TV}(X,\ Y \mid U)$ 表示给定 U 的观测值后，它们的条件分布之间的距离。

设 p，q 是 P，Q 关于共同测度 μ 的密度，这意味着对于每个可测集 A，都有 $P(A) = \int_A p\,\mathrm{d}\mu$，且 $Q(A) = \int_A q\mathrm{d}\mu$。$P$ 与 Q 之间的海林格距离 的平方定义为

$$H^2(P,Q) \coloneqq \frac{1}{2}\int \left(\sqrt{p} - \sqrt{q}\right)^2 \mathrm{d}\mu.$$

此定义与 μ 的选取无关，而且总可以取 $\mu = P + Q$。对于随机变量 X 和 Y，我们用 $H(X,\ Y)$ 表示它们的分布之间的海林格距离。对于概率分布 P_1，……，P_m，记号 $\bigotimes_{i=1}^m P_i$ 表示它们的乘积分布，即相互独立的随机变量 $X_i \sim P_i$ 的联合分布。

我们将使用海林格距离的以下标准性质。

引理 3.3。对于同一可测空间上的概率分布 P_i、Q_i，对每个 $i \in [m]$,

$$H^2\Big(\bigotimes_{i=1}^m P_i, \bigotimes_{i=1}^m Q_i\Big) \le \sum_{i=1}^m H^2(P_i,Q_i).$$

对于概率 分布 P、Q，在同一可测空间上，$\mathrm{TV}(P$、$Q) \le \sqrt{2}\,H(P$、$Q)$。 –

4 困难分布与杠杆分数的归约

4.1 困难分布

在本小节中，我们描述用于证明关联草图下界的困难分布。

给定一个整数 $b \ge 1$ 和一个偶数 $\ell \ge 2$，我们设定 $n \coloneqq \ell b$，并令该图由顶点集 $\mathcal{R}\coloneqq [n]$ 构成。我们从一个固定的、包含 ℓ 个循环排序的块的模板开始，这些块的索引为 0、……、$\ell - 1$，因此块 i 包含 b 个顶点槽位 $ib+1$、……、$(i+1)b$。此外，我们任意固定块 0 中的一个指定查询槽位，以及块 $\ell/2$ 中的另一个指定查询槽位。

令 π 为 $[n]$ 的均匀随机置换，并对每个 $t \in [n]$，将顶点 $\pi(t)$ 放置在槽位 t 中。因此，所得的顶点块为

$$\mathcal{B}_i \coloneqq \big\{\pi(ib+j) : j \in [b]\big\}, \qquad i = 0,\dots,\ell-1.$$

令 x 和 y 分别为占据块 0 和块 $\ell/2$ 中指定查询槽位的顶点，并定义查询边 $q \coloneqq xy$。因此，顶点划分和查询边都由 π 决定。

对于块索引 $0 \le i,\ j < \ell$，令 $L(i,\ j) \coloneqq \min\big\{|i-j|, \ell - |i-j|\big\}$ 表示循环块距离。我们将每条边的基准权重定义为

$$w_e \coloneqq \begin{cases} w_{\mathrm{int}}, & e \in \binom{\mathcal{B}_i}{2} \text{ for some } i, \\ w_{\mathrm{ext}}\,L(i,j)^{-4}, & e = uv, u \in \mathcal{B}_i, v \in \mathcal{B}_j, i \ne j, \end{cases} \tag{3}$$

其中 $w_{\mathrm{int}} := n^2$ 且 $w_{\mathrm{ext}} := \ell/\big(16b^2\big)$。我们将 $\boldsymbol{w} := (w_e)_{e\in\binom{\mathcal{R}}{2}}$ 记作基准权重向量，在给定 π的条件下，该向量是确定性的。

我们将 Z 设为 $1/2$ 加上一个服从 Beta(3，3) 分布的随机变量。等价地，Z 的密度为

$$f(z) := \mathbf{1}\Big\{\frac{1}{2} \le z \le \frac{3}{2}\Big\} \cdot 30\Big(z-\frac{1}{2}\Big)^2\Big(\frac{3}{2}-z\Big)^2. \tag{4}$$

$Z \in [1/2, 3/2]$ 和 $\mathbb{E}[Z] = 1$成立。然后，在给定 π的条件下，我们为每个 $e \in \binom{\mathcal{R}}{2}$独立采样 Z 的副本Z_e。独立于 π 和各个 Z_e，我们进一步以相同的概率 $1/2$随机选择是或否情形。输入图的扰动后边权重 设为

$$\widetilde{w}_e := \begin{cases} w_e Z_e, & \text{in the No case,} \\ w_e Z_e + \mathbf{1}\{e = q\}, & \text{in the Yes case.} \end{cases}$$

输入图是定义在 $\mathcal{R}$ 上的完全图，其权重向量为 $\widetilde{\boldsymbol{w}} := (\widetilde{w}_e)_{e\in\binom{\mathcal{R}}{2}}$。

根据边权重的设定，我们可以得到输入图正权重比的如下界。

引理 4.1。 在是 或否 情形下采样得到的每个图，其正权重比均为 $O\big(n^5\big)$。

证明。对于连接不同块 $\mathcal{B}_i$ 和 $\mathcal{B}_j$ 的边 e，循环块距离满足 $1 \le L(i,\ j) \le \ell/2$。因此，其基准权重满足

$$w_e = \frac{\ell}{16b^2}\, L(i,j)^{-4} \ge \frac{\ell}{16b^2}\left(\frac{\ell}{2}\right)^{-4} = \frac{1}{\ell^3 b^2} \ge \frac{1}{n^3}.$$

该下界对于块内的边同样成立，其基准权重为 n^2。此外，每个跨块基础权重至多为 $\ell/\big(16b^2\big) \le n^2$。因此，每条边都满足 $1/n^3 \le w_e \le n^2$。

在是或否情形中，每个基准权重都乘以一个 $[1/2, 3/2]$ 中的噪声，并且查询边权重还可能额外增加一。因此，

$$\frac{1}{2n^3} \le \frac{1}{2}w_e \le \widetilde{w}_e \le \frac{3}{2}w_e + 1 \le 2n^2.$$

因此，任一分布的每个实现的正权重比至多为 $2n^2/\big(1/\big(2n^3\big)\big) = 4n^5$，这证明了引理。

□

下一个引理表明，图的 (1 ± 0.01)-谱稀疏化器连同 π，可以帮助区分是分布和否分布。

引理 4.2。 假设某个线性草图在 (1 ± 0.01) 的谱稀疏化器上，对 否 或是 的支撑集中的每个输入，以至少 $\widehat{\ }0.99$ 的概率计算出结果。那么，给定该线性草图的恢复算法输出 $\mathbf{L}$ 和排列 π，可以在任一分布下以至少0.99 的成功概率，识别输入是从是 还是 否 中采样得到的。

证明。 我们构造一个测试势向量 $\boldsymbol{\phi} \in \mathbb{R}^n$，计算其在两个输入分布下的二次型 $\boldsymbol{\phi}^\top \mathbf{L}_{\widetilde{\boldsymbol{w}}} \boldsymbol{\phi}$，然后证明该线性草图的谱近似保证能够区分这两种情况。

令 $\phi_i := 2\min\{i, \ell - i\}/\ell$，于是 $|\phi_i - \phi_j| \le 2L(i,\ j)/\ell$。对于每个 $0 \le i < \ell$，令 $\phi(v) := \phi_i$ 对所有 $v \in \mathcal{B}_i$成立，并将所得向量记为 $\boldsymbol{\phi}$。首先考虑 No 情形。对于 $u \in \mathcal{B}_i$ 和

$v \in \mathcal{B}_j$ 其中 $i \neq j$，扰动后的权重满足 $\widetilde{w}_{uv} = w_{uv} Z_{uv} = Z_{uv} \cdot \ell/\big(16b^2\big) \cdot L(i,\ j)^{-4} \leq (3/2) \cdot \ell/\big(16b^2\big) \cdot L(i,\ j)^{-4}$。块内的边对二次型的贡献为零，因此

$$\begin{aligned} 0 \leq \boldsymbol{\phi}^\top \mathbf{L}_{\widetilde{\boldsymbol{w}}} \boldsymbol{\phi} &\leq \frac{3\ell}{32b^2} \sum_{0 \leq i < j < \ell} \sum_{u \in \mathcal{B}_i} \sum_{v \in \mathcal{B}_j} L(i,j)^{-4} \left(\phi_i - \phi_j\right)^2 \\ &\leq \frac{3\ell}{32b^2} \sum_{0 \leq i < j < \ell} b^2 L(i,j)^{-4} \frac{4L(i,j)^2}{\ell^2} \\ &= \frac{3}{8\ell} \sum_{0 \leq i < j < \ell} L(i,j)^{-2} \leq \frac{3}{8\ell} \sum_{d \geq 1} \ell \cdot \frac{1}{d^2} \leq \frac{3}{4}. \end{aligned}$$

在倒数第二步中，我们使用了以下事实：对于每个 $d \geq 1$，至多有 ℓ 对 $0 \leq i < j < \ell$ 满足 $L(i,\ j) = d$。在 Yes 情形下，查询边上额外增加的单位权重使得 $\boldsymbol{\phi}^\top \mathbf{L}_{\widetilde{\boldsymbol{w}}} \boldsymbol{\phi}$ 至少为 1，因为 $\phi(x) = 0$ 且 $\phi(y) = 1$。

现在我们考虑以下区分这两个分布的方法。给定 π 以及恢复算法的输出 $\widehat{\mathbf{L}}$，计算 $\boldsymbol{\phi}^\top \widehat{\mathbf{L}} \boldsymbol{\phi}$。如果 $\boldsymbol{\phi}^\top \widehat{\mathbf{L}} \boldsymbol{\phi} > 7/8$，则声明为 Yes，否则声明为 No。

如果恢复算法输出一个有效的 $\widehat{\mathbf{L}}$，在 No 情形下 $\boldsymbol{\phi}^\top \widehat{\mathbf{L}} \boldsymbol{\phi} \leq 1.01(3/4) < 7/8$，而在 Yes 情形下，$\boldsymbol{\phi}^\top \widehat{\mathbf{L}} \boldsymbol{\phi} \geq 0.99 > 7/8$。根据假设，对于 No 或 Yes 的支撑集中的每个输入，在草图矩阵和恢复算法的随机性上，恢复以至少 0.99 的概率成功。由于恢复成功时区分算法必然正确，对输入取平均可得，在任一分布下成功概率均至少为 0.99。

□

4.2 化简为杠杆分数

为给出全变差距离的上界，我们将使用以下关于某个分布及其平移版本之间海林格距离的初等界。

引理 4.3。设 X 是一个实值随机变量，其在 $\mathbb{R}$ 上的密度为 p，满足 $u := \sqrt{p}$ 是绝对连续的，且 $\int_{\mathbb{R}} |u'(z)|^2 \mathrm{d}z < \infty$。对于每个 $h \in \mathbb{R}$，

$$H^2(X, X+h) \leq \frac{1}{2} h^2 \int_{\mathbb{R}} \left|u'(z)\right|^2 \mathrm{d}z.$$

证明。 $X + h$ 的密度为 $p(z - h)$，因此根据海林格距离的定义可得

$$2H^2(X, X+h) = \int_{\mathbb{R}} \left|u(z) - u(z-h)\right|^2 \mathrm{d}z.$$

根据 u 的绝对连续性和微积分基本定理，

$$u(z) - u(z-h) = h \int_0^1 u'(z - th)\, \mathrm{d}t.$$

代入这一恒等式，并对关于 $t \in [0,1]$ 的积分应用柯西–施瓦茨不等式，我们得到

$$\begin{aligned} 2H^2(X, X+h) &= h^2 \int_{\mathbb{R}} \left| \int_0^1 u'(z - th)\, \mathrm{d}t \right|^2 \mathrm{d}z \\ &\leq h^2 \int_{\mathbb{R}} \int_0^1 \left|u'(z-th)\right|^2 \mathrm{d}t\, \mathrm{d}z = h^2 \int_0^1 \int_{\mathbb{R}} \left|u'(z-th)\right|^2 \mathrm{d}z\, \mathrm{d}t \\ &= h^2 \int_0^1 \left(\int_{\mathbb{R}} \left|u'(z)\right|^2 \mathrm{d}z \right) \mathrm{d}t = h^2 \int_{\mathbb{R}} \left|u'(z)\right|^2 \mathrm{d}z, \end{aligned}$$

其中，由于被积函数非负，根据托内利定理，在第二行交换了积分顺序。证明完毕。

□

下面的引理用杠杆分数对线性草图在是分布和否分布下输出之间的全变差距离给出上界。

对于该实验的统计量 X ， 记$(X)_{\mathsf{Yes}}$ 和$(X)_{\mathsf{No}}$ 为其在两种情形下的分布；记$(X)_{\pi,\ \mathsf{Yes}}$ 和$(X)_{\pi,\ \mathsf{No}}$ 为进一步对实现的排列进行条件化时的分布。

引理 4。 **4.** 对于任意固定的草图矩阵$\mathbf{A}$ 以及随机排列 π的实现 ，

$$\mathrm{TV}\big((\mathbf{A}\widetilde{\boldsymbol{w}})_{\pi,\mathsf{Yes}},(\mathbf{A}\widetilde{\boldsymbol{w}})_{\pi,\mathsf{No}}\big)\le\sqrt{\min\Big\{1,10\,\mathrm{lev}_{\boldsymbol{a}_q}\Big(\big(\mathbf{A}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big)\Big\}}.$$

证明。 注意，对于固定的 π， q 和 $\boldsymbol{w}$ 都是确定的。如果 $\boldsymbol{a}_q$ 不在$\mathbf{A}_{-q}$的值域中，则结论显然成立，因为全变差距离至多为 1。因此，假设 $\boldsymbol{a}_q$ 在$\mathbf{A}_{-q}$的值域中。回忆

$$\begin{aligned}\mathrm{lev}_{\boldsymbol{a}_q}\Big(\big(\mathbf{A}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big)&=\min\Big\{\sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}}c_e^2:\boldsymbol{a}_q=\sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}}c_e w_e\boldsymbol{a}_e\Big\}\\&=\min\Big\{\sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}}\frac{c_e^2}{w_e^2}:\boldsymbol{a}_q=\sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}}c_e\boldsymbol{a}_e\Big\}.\end{aligned}$$

因此，存在系数$(c_e)_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}}$，使得$\boldsymbol{a}_q=\sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}}c_e\boldsymbol{a}_e$且$\sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}}c_e^2/w_e^2=\mathrm{lev}_{\boldsymbol{a}_q}\big(\big(\mathbf{A}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\big)$。令 $c_q:=0$和 $\boldsymbol{c}:=(c_e)_{e\in\binom{\mathcal{R}}{2}}$，则$\mathbf{A}\boldsymbol{c}=\boldsymbol{a}_q$。

设 $\widetilde{\boldsymbol{w}}'=(\widetilde{w}'_e)_{e\in\binom{\mathcal{R}}{2}}$ 服从分布 $(\widetilde{\boldsymbol{w}})_{\pi,\ \mathsf{No}}$。其坐标相互独立，且 $\widetilde{w}'_e=w_eZ_e$。我们有

$$\begin{aligned}&\mathrm{TV}\big((\mathbf{A}\widetilde{\boldsymbol{w}})_{\pi,\mathsf{Yes}},(\mathbf{A}\widetilde{\boldsymbol{w}})_{\pi,\mathsf{No}}\big)\\=\;&\mathrm{TV}(\mathbf{A}\widetilde{\boldsymbol{w}}'+\boldsymbol{a}_q,\mathbf{A}\widetilde{\boldsymbol{w}}')=\mathrm{TV}\big(\mathbf{A}(\widetilde{\boldsymbol{w}}'+\boldsymbol{c}),\mathbf{A}\widetilde{\boldsymbol{w}}'\big)\\\le\;&\mathrm{TV}(\widetilde{\boldsymbol{w}}',\widetilde{\boldsymbol{w}}'+\boldsymbol{c})\le\sqrt{2}\,H(\widetilde{\boldsymbol{w}}',\widetilde{\boldsymbol{w}}'+\boldsymbol{c}).\end{aligned}$$

对所有边应用引理 3.3可得 $H^2(\widetilde{\boldsymbol{w}}',\ \widetilde{\boldsymbol{w}}'+\boldsymbol{c})\le\sum_{e\in\binom{\mathcal{R}}{2}}H^2(\widetilde{w}'_e,\ \widetilde{w}'_e+c_e)$。接下来，我们根据引理4.3计算每一项 $H^2(w_eZ_e,\ w_eZ_e+c_e)$。密度 Z的平方根为 $u(z):=\mathbf{1}\{\frac{1}{2}\le z\le\frac{3}{2}\}\sqrt{30}(z-\frac{1}{2})(\frac{3}{2}-z)$，它是绝对连续的。我们有

$$\int_{\mathbb{R}}\big|u'(z)\big|^2\,\mathrm{d}z=\int_{1/2}^{3/2}30(2-2z)^2\,\mathrm{d}z=10.$$

现在，根据引理4.3，对于每个基准权重 $w>0$ 和 $c\in\mathbb{R}$ ，我们有

$$H^2(wZ,wZ+c)=H^2\Big(Z,Z+\frac{c}{w}\Big)\le\frac{1}{2}\cdot\frac{c^2}{w^2}\int_{\mathbb{R}}\big|u'(z)\big|^2\,\mathrm{d}z=\frac{5c^2}{w^2}.$$

综上，我们有

$$\mathrm{TV}\big((\mathbf{A}\widetilde{\boldsymbol{w}})_{\pi,\mathsf{Yes}},(\mathbf{A}\widetilde{\boldsymbol{w}})_{\pi,\mathsf{No}}\big)\le\sqrt{10\sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}}\frac{c_e^2}{w_e^2}}=\sqrt{10\ \mathrm{lev}_{\boldsymbol{a}_q}\Big(\big(\mathbf{A}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big)},$$

结合全变差距离的界 1，即可完成证明。 □

下一个引理将我们的草图维度下界证明归约为界定查询列截断于 1 时的期望杠杆分数。我们还额外公开 π，以便区分算法能够应用引理 4.2，从而在恢复成功时识别正确的情形。对于固定的草图矩阵 $\mathbf{A}$，记 $(\pi,\ \mathbf{A}\widetilde{\boldsymbol{w}})_{\mathsf{No}}$ 以及 $(\pi,\ \mathbf{A}\,\widetilde{\boldsymbol{w}})_{\mathsf{Yes}}$ 为两种情形下排列与草图的联合分布。

引理 4.5。 考虑一个线性草图，并假设对于其草图矩阵的每个实现 $\mathbf{A} = \big[\boldsymbol{a}_e :\ \ e \in \binom{\mathcal{R}}{2}\big]$，

$$\mathbb{E}_{\pi}\Big[\min\Big\{1, \mathrm{lev}_{\boldsymbol{a}_q}\Big(\big(\mathbf{A}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big)\Big\}\Big] < \frac{1}{100}.$$

那么，对于该草图的任何恢复算法，都不可能以至少（1 ± 0.01）的概率，在否 或是的支撑集中的每个输入上输出一个谱稀疏化器。099

证明。 在给定的杠杆分数假设下，我们首先上界给定草图和 否 与 是 时，任何能够区分二者的算法的成功概率 π。然后我们证明，满足所述保证的恢复算法将与该上界矛盾。固定草图矩阵的一个实现 $\mathbf{A}$ 。在给定 π的条件下，草图 $\mathbf{A}\,\widetilde{\boldsymbol{w}}$ 在两种情形下分别具有分布 $(\mathbf{A}\widetilde{\boldsymbol{w}})_{\pi,\,\mathsf{No}}$ 以及 $(\mathbf{A}\,\widetilde{\boldsymbol{w}})_{\pi,\,\mathsf{Yes}}$ 由于 π 在是 和 否下具有相同的分布，因此联合分布之间的全变差距离等于条件总变差距离的平均值。根据引理 4.4，

$$\begin{aligned}
&\mathrm{TV}\big((\pi, \mathbf{A}\widetilde{\boldsymbol{w}})_{\mathsf{Yes}}, (\pi, \mathbf{A}\widetilde{\boldsymbol{w}})_{\mathsf{No}}\big)\\
&= \mathbb{E}_{\pi}\Big[\mathrm{TV}\big((\mathbf{A}\widetilde{\boldsymbol{w}})_{\pi,\mathsf{Yes}}, (\mathbf{A}\widetilde{\boldsymbol{w}})_{\pi,\mathsf{No}}\big)\Big]\\
&\le \mathbb{E}_{\pi}\left[\sqrt{\min\Big\{1, 10\,\mathrm{lev}_{\boldsymbol{a}_q}\Big(\big(\mathbf{A}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big)\Big\}}\right]\\
&\le \sqrt{10\,\mathbb{E}_{\pi}\Big[\min\Big\{1, \mathrm{lev}_{\boldsymbol{a}_q}\Big(\big(\mathbf{A}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big)\Big\}\Big]} < \frac{1}{\sqrt{10}}.
\end{aligned}$$

第二个不　　等式对 $u \in [0,\infty]$ 使用 $\min\{1, 10u\} \le 10\min\{1, u\}$，并使用詹森不等式　　.

由于两种情况出现的概率相同，(2) 表明，任何给定草图和 π 的随机化区分算法，其成功概率至多为 $\frac{1}{2}(1 + 1/\sqrt{10}) < \frac{3}{4}$。此界对每个固定的 $\mathbf{A}$ 都成立。由于草图矩阵是独立于输入实验采样的，对其选择取平均后，同一界仍然成立。

现在假设某个恢复算法对任一分布支撑集中的输入都满足所述谱保证。对 $\mathbf{A}$ 和 $\mathbf{A}\widetilde{\boldsymbol{w}}$运行该算法，然后使用 4.2 中引理证明所用的 π区分算法。只要恢复算法产生有效的谱近似，所得算法就能正确区分。对于每个输入，在对草图矩阵的选择和恢复算法的随机性取平均后，恢复算法的成功概率至少为0。99。因此，对输入分布取平均后，每种情况下的成功概率至少为0。99 ，这与上述上界矛盾。

□

5 关联草图的下界

定理 5.1。 任何维数为 N、以至少 0.99 的概率在每个具有 $w_{\max}/w_{\min} = O(n^5)$ 的 n 个顶点图上计算出 (1 ± 0.01)-谱稀疏化器的关联草图 都必须满足 $N = \Omega(n^{6/5})$。

证明依赖于以下对查询列的表示，我们也将在第 6 节中关于一般线性草图的证明里使用该表示。

引理 5.2. 设 t 为正整数，$\ell \geq 10$ 为偶整数，并令 $b := (10^8+1)t+1$。固定一个大小为 ℓb 的顶点集 $\mathcal{R}$，以及一个维数为 $N \leq 10^{-9}\ell bt$ 的关联草图矩阵 $\mathbf{B} = [\boldsymbol{b}_e : e \in \binom{\mathcal{R}}{2}]$。在环槽位的均匀随机标记下，存在一个事件 $\mathcal{E}$，使得在该事件上，查询列可以表示为

$$\boldsymbol{b}_q = \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} f_e \boldsymbol{b}_e.$$

系数向量 $\boldsymbol{f} = (f_e)_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}}$ 仅取决于环标记，并且在 $\mathcal{E}$ 之外设为零。当 $\boldsymbol{b}_e = \mathbf{0}$ 且 $\|\boldsymbol{f}\|_2^2 \leq 2\ell b$ 时，系数满足 $f_e = 0$。此外，

$$\mathbb{E}\left[\sum_{0 \leq i < j < \ell} L(i,j)^8 \sum_{\substack{u \in \mathcal{B}_i, v \in \mathcal{B}_j, \\ uv \neq q}} f_{uv}^2 \,\middle|\, \mathcal{E} \right] \leq 25\ell. \tag{5}$$

换言之，该引理保证，在随机环标记下，以高概率查询列 $\boldsymbol{b}_q$ 可以表示为其他列的线性组合；并且在期望意义下，表示系数按循环块距离的八次方加权后的平方和有界。

这两个系数界对应（3）中的不同基准权重。对于块内边，其基准权重为 $\|\boldsymbol{f}\|_2^2$，此处关于 $w_{\text{int}} = n^2$的界已足够。对于块间边，$w_{uv} = w_{\text{ext}} L(i,j)^{-4}$，因此它们的贡献 f_{uv}^2 / w_{uv}^2为 $w_{\text{ext}}^{-2} L(i,j)^8 f_{uv}^2$。因此，我们需要（5）中的距离加权界。

5.1 查询列的望远镜求和分解

令 $b = t + s + 1$，其中 t和 s稍后选定。在指定顶点标签之前，先指定 t个保留槽位、s 个辅助槽，并在每个块中指定一个备用槽位。记 $\mathcal{T}_i$，$\mathcal{S}_i \subseteq \mathcal{B}_i$ 为 π 分配给块 i中保留槽位和辅助槽的标签，因此 $|\mathcal{T}_i| = t$ 且 $|\mathcal{S}_i| = s$。查询端点 x和 y 分别占据块 0 和 $\ell/2$中的备用槽位。槽位组预先固定，只有它们的标签是随机的。这些集合仅用于分析，不会改变困难分布。

对于每个保留集 $\mathcal{T}_i$，令$\mathbf{B}_{\mathcal{T}_i}$ 表示由 $\mathcal{T}_i$内部边上的列组成的矩阵，并令$\mathbf{B}_{v,\,\mathcal{T}_i}$ 表示由连接 $v \notin \mathcal{T}_i$ 和 $\mathcal{T}_i$的边上的列组成的矩阵。以 y 为公共根，如果$\mathrm{lev}_{\boldsymbol{b}_{vy}}\big([\mathbf{B}_{v,\,\mathcal{T}_i}\, \mathbf{B}_{y,\,\mathcal{T}_i}\, \mathbf{B}_{\mathcal{T}_i}]\big) \leq 1$，则称 $v \notin \mathcal{T}_i \cup \{y\}$保留良好。

引理5.3。 固定一个块索引 $i \in \{0, \ldots, \ell-1\}$。在均匀环标记下，令 $\mathcal{T}_i$ 由块 i 的指定保留槽位以及 x 给出，令 y 由块0和$\ell/2$中的两个备用槽位给出，则有

$$\Pr\{x \text{ is not reserve-good for } \mathcal{T}_i \text{ with respect to } y\} \leq \frac{4N}{|\mathcal{R}|(t+1)}. \tag{6}$$

证明。对于固定的关联草图矩阵$\mathbf{B}$，我们将每一行分配给其支撑集中每条边所关联的某个顶点（零行可以任意分配）。注意，我们在随机环标记之前就固定了行的分配。令 $\mathcal{I}_v \subseteq [N]$ 表示分配给顶点 v 的行索引集合，并令 $k_v := |\mathcal{I}_v|$。则有 $\mathrm{supp}(\boldsymbol{b}_{uv}) \subseteq \mathcal{I}_u \cup \mathcal{I}_v$ 且 $\sum_{v \in \mathcal{R}} k_v = N$。

固定顶点集 $\mathcal{T}_i \cup \{x, y\}$。其诱导团的 $\boldsymbol{b}$ 列的秩至多为$\sum_{v\in\mathcal{T}_i\cup\{x,y\}} k_v$，因为这些列仅支撑在与这些顶点相关联的坐标上。由于查询边 xy 在 $\binom{t+2}{2}$ 个边槽位中均匀随机选取，引理 3.1 将杠杆分数超过一的条件概率上界为 $2\sum_{v\in\mathcal{T}_i\cup\{x,y\}} k_v/\binom{t+2}{2}$。集合 $\mathcal{T}_i \cup \{x, y\}$ 是 $\mathcal{R}$的均匀 $(t+2)$顶点子集，因此 $\mathbb{E}\big[\sum_{v\in\mathcal{T}_i\cup\{x,y\}} k_v\big] = (t+2)N/|\mathcal{R}|$。对条件上界取平均后，即得到上界 $2(t+2)$ $N/\big(|\mathcal{R}|\binom{t+2}{2}\big) = 4N/\big(|\mathcal{R}|(t+1)\big)$，正如所需。

□

对于任何对 $\mathcal{T}_i$ 而言保留良好的 $v \in \mathcal{R}\setminus(\mathcal{T}_i\cup\{y\})$，我们将其系数坐标按顶点标签排序后的最小范数表示写作

$$\boldsymbol{b}_{vy} = \mathbf{B}_{v,\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{v,\mathcal{T}_i} + \mathbf{B}_{y,\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{y,\mathcal{T}_i} + \mathbf{B}_{\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{\mathcal{T}_i}, \tag{7}$$

其中，三个系数向量分别对应于从 v 到 $\mathcal{T}_i$ 的边、从 y 到 $\mathcal{T}_i$ 的边，以及 $\mathcal{T}_i$ 内部的边。根据保留良好性，

$$\big\|\boldsymbol{\alpha}^{(v)}_{v,\mathcal{T}_i}\big\|_2^2 + \big\|\boldsymbol{\alpha}^{(v)}_{y,\mathcal{T}_i}\big\|_2^2 + \big\|\boldsymbol{\alpha}^{(v)}_{\mathcal{T}_i}\big\|_2^2 \le 1.$$

项 $\mathbf{B}_{y,\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{y}$、$\mathcal{T}_i$ 使用从 y 到 $\mathcal{T}_i$的星；我们称其为 根星项。对于 $v = x$ 和 $i = 0$，式（7）中的另外两项均表示在块 0内，如图 1b所示。我们将用一个使用 $\mathcal{T}_{\ell/2}$的根星项替换剩余的根星项，该项表示在包含 y 的块内。

我们将选择一条路径 $0 = i_0 < i_1 < \cdots < i_m = \ell/2$，使得对于每个选定的保留集 $\mathcal{T}_{i_0}$、$\mathcal{T}_{i_1}$、……、$\mathcal{T}_{i_m}$，x 都是保留良好的。沿着这条路径对根星项进行加减运算，得到查询列 $\boldsymbol{b}_{xy}$ 的如下望远镜分解：

$$\begin{aligned}\boldsymbol{b}_{xy} = {}& \big(\boldsymbol{b}_{xy} - \mathbf{B}_{y,\mathcal{T}_0}\boldsymbol{\alpha}^{(x)}_{y,\mathcal{T}_0}\big)\\ &+ \sum_{j=1}^{m}\big(\mathbf{B}_{y,\mathcal{T}_{i_{j-1}}}\boldsymbol{\alpha}^{(x)}_{y,\mathcal{T}_{i_{j-1}}} - \mathbf{B}_{y,\mathcal{T}_{i_j}}\boldsymbol{\alpha}^{(x)}_{y,\mathcal{T}_{i_j}}\big)\\ &+ \mathbf{B}_{y,\mathcal{T}_{\ell/2}}\boldsymbol{\alpha}^{(x)}_{y,\mathcal{T}_{\ell/2}}.\end{aligned} \tag{8}$$

剩下的工作是表示这个和中的差，并选择路径，使其总的距离加权系数代价较小。

5.2 望远镜求和项的表示

对于两个预备集 $\mathcal{T}_i$，$\mathcal{T}_j$，它们的根星项之差 $\mathbf{B}_{y,\mathcal{T}_i}\boldsymbol{\alpha}^{(x)}_{y,\mathcal{T}_i} - \mathbf{B}_{y,\mathcal{T}_j}\boldsymbol{\alpha}^{(x)}_{y,\mathcal{T}_j}$ 属于一个维数至多为 $2t$ 的空间。我们利用 x 与 $\mathcal{S}_i$ 中辅助顶点之间的对称性，将 x 的差表示为辅助顶点对应的差。将每个辅助顶点的两个预备集表示相减，会消去它到 y 的列，只留下预备集内部的边以及辅助顶点到预备集的边；见图 2。

引理 5.4。 固定不同的块索引 $i, j \in \{0, \ldots, \ell - 1\}$，并考虑顶点集 $\mathcal{T}_i, \mathcal{T}_j, \mathcal{S}_i$。在均匀环标记下，以至少 $1 - \frac{8N}{|\mathcal{R}|(t+1)} - \frac{4t}{s+1}$ 的概率，以下结论成立：

（1） x 对于 $\mathcal{T}_i$ 和 $\mathcal{T}_j$均为保留良好，且

(2) 存在系数 $\boldsymbol{f} = (f_e)_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}}$ 使得

$$\mathbf{B}_{y,\mathcal{T}_i}\boldsymbol{\alpha}^{(x)}_{y,\mathcal{T}_i} - \mathbf{B}_{y,\mathcal{T}_j}\boldsymbol{\alpha}^{(x)}_{y,\mathcal{T}_j} = \sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}} f_e \boldsymbol{b}_e,$$

此外，$\boldsymbol{f}$ 满足

$$\operatorname{supp}(\boldsymbol{f}) \subseteq \binom{\mathcal{T}_i}{2} \cup \binom{\mathcal{T}_j}{2} \cup \{uv : u \in \mathcal{S}_i, v \in \mathcal{T}_i \cup \mathcal{T}_j\}$$

and

$$\sum_{u\in\mathcal{S}_i}\sum_{v\in\mathcal{T}_j} f_{uv}^2 \le 1, \tag{9}$$

$$\sum_{e\in\binom{\mathcal{T}_i}{2}\cup\binom{\mathcal{T}_j}{2}} f_e^2 + \sum_{u\in\mathcal{S}_i}\sum_{v\in\mathcal{T}_i} f_{uv}^2 \le 2s+1. \tag{10}$$

注意，对于所述系数向量 $\boldsymbol{f}$中的非零项，只有对应于 $\mathcal{S}_i$ 与 $\mathcal{T}_j$ 之间的边的项跨越了不同的块，且它们的平方和在 1（9）中有界。

引理 5.4的证明。对于每个同时对两个预备集均为 reserve-good 的 $v \in \mathcal{S}_i \cup \{x\}$，定义 $\boldsymbol{\beta}_v := \mathbf{B}_{y,\,\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{y,\,\mathcal{T}_i} - \mathbf{B}_{y,\,\mathcal{T}_j}\boldsymbol{\alpha}^{(v)}_{y,\,\mathcal{T}_j}$。对于 $\mathcal{S}_i \cup \{x\}$中的其他顶点，令 $\boldsymbol{\beta}_v := \mathbf{0}_N$。这为 $\mathcal{S}_i \cup \{x\}$中的每个顶点定义了一个向量，因此我们可以利用查询槽 x 与所有辅助槽之间的对称性，而无需对 reserve-good 条件化。令 $\mathcal{E}_1$ 表示 x 同时对两个预备集均为 reserve-good 且 $\operatorname{lev}_{\boldsymbol{\beta}_x}\big((\boldsymbol{\beta}_v)_{v\in\mathcal{S}_i}\big) \le 1$的事件。该事件仅取决于 x、y、$\mathcal{T}_i$、$\mathcal{T}_j$和 $\mathcal{S}_i$。

我们首先给出 $\Pr\{\mathcal{E}_1\}$ 的下界。对 y、$\mathcal{T}_i$、$\mathcal{T}_j$ 以及顶点集 $\mathcal{S}_i \cup \{x\}$ 进行条件化。此时，向量 $(\boldsymbol{\beta}_v)_{v\in\mathcal{S}_i\cup\{x\}}$ 已固定，而 x 在该顶点集中均匀分布。每个 $\boldsymbol{\beta}_v$ 都属于值域$\big([\mathbf{B}_{y,\,\mathcal{T}_i}\mathbf{B}_{y,\,\mathcal{T}_j}]\big)$，因此矩阵 $\big[\boldsymbol{\beta}_v : v \in \mathcal{S}_i \cup \{x\}\big]$ 的秩至多为 $2t$。根据引理 3.1，$\operatorname{lev}_{\boldsymbol{\beta}_x}\big((\boldsymbol{\beta}_v)_{v\in\mathcal{S}_i}\big)$ 超过一的条件概率至多为 $4t/(s+1)$。对条件化取平均，并将引理 5.3 应用于两个预备集后，由并集界可得 $\Pr\{\mathcal{E}_1\} \ge 1 - \frac{8N}{|\mathcal{R}|(t+1)} - \frac{4t}{s+1}$。

在 $\mathcal{E}_1$ 上，选取最小范数系数 $(c_v)_{v\in\mathcal{S}_i}$，使得 $\boldsymbol{\beta}_x = \sum_{v\in\mathcal{S}_i} c_v\boldsymbol{\beta}_v$。于是 $\sum_{v\in\mathcal{S}_i} c_v^2 \le 1$，并且当 $\boldsymbol{\beta}_v = \mathbf{0}$ 时有 $c_v = 0$。对于每个满足 $c_v \ne 0$ 的 v，两个预备集表示都存在；将它们相减可消去 $\boldsymbol{b}_{vy}$，从而得到

$$\boldsymbol{\beta}_v = \mathbf{B}_{\mathcal{T}_j}\boldsymbol{\alpha}^{(v)}_{\mathcal{T}_j} - \mathbf{B}_{\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{\mathcal{T}_i} + \mathbf{B}_{v,\mathcal{T}_j}\boldsymbol{\alpha}^{(v)}_{v,\mathcal{T}_j} - \mathbf{B}_{v,\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{v,\mathcal{T}_i}.$$

将其代入 $\boldsymbol{\beta}_x = \sum_{v\in\mathcal{S}_i} c_v\boldsymbol{\beta}_v$，得到

$$\begin{aligned}
\boldsymbol{\beta}_x &= \sum_{v\in\mathcal{S}_i, c_v\ne 0} c_v\big(\mathbf{B}_{\mathcal{T}_j}\boldsymbol{\alpha}^{(v)}_{\mathcal{T}_j} - \mathbf{B}_{\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{\mathcal{T}_i} + \mathbf{B}_{v,\mathcal{T}_j}\boldsymbol{\alpha}^{(v)}_{v,\mathcal{T}_j} - \mathbf{B}_{v,\mathcal{T}_i}\boldsymbol{\alpha}^{(v)}_{v,\mathcal{T}_i}\big) \\
&= \mathbf{B}_{\mathcal{T}_j}\sum_{v\in\mathcal{S}_i, c_v\ne 0} c_v\boldsymbol{\alpha}^{(v)}_{\mathcal{T}_j} - \mathbf{B}_{\mathcal{T}_i}\sum_{v\in\mathcal{S}_i, c_v\ne 0} c_v\boldsymbol{\alpha}^{(v)}_{\mathcal{T}_i} \\
&\quad + \sum_{v\in\mathcal{S}_i, c_v\ne 0} \mathbf{B}_{v,\mathcal{T}_j}\big(c_v\boldsymbol{\alpha}^{(v)}_{v,\mathcal{T}_j}\big) - \sum_{v\in\mathcal{S}_i, c_v\ne 0} \mathbf{B}_{v,\mathcal{T}_i}\big(c_v\boldsymbol{\alpha}^{(v)}_{v,\mathcal{T}_i}\big).
\end{aligned}$$

读出每列的系数，定义 $\boldsymbol{f} = (f_e)_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}}$ 为

$$f_e \coloneqq \begin{cases} -\sum\limits_{v\in\mathcal{S}_i, c_v\neq 0} c_v \big(\boldsymbol{\alpha}_{\mathcal{T}_i}^{(v)}\big)_e, & e\in\binom{\mathcal{T}_i}{2}, \\ \sum\limits_{v\in\mathcal{S}_i, c_v\neq 0} c_v \big(\boldsymbol{\alpha}_{\mathcal{T}_j}^{(v)}\big)_e, & e\in\binom{\mathcal{T}_j}{2}, \\ -c_v\big(\boldsymbol{\alpha}_{v,\mathcal{T}_i}^{(v)}\big)_{uv}, & e=uv, u\in\mathcal{T}_i, v\in\mathcal{S}_i, c_v\neq 0, \\ c_v\big(\boldsymbol{\alpha}_{v,\mathcal{T}_j}^{(v)}\big)_{uv}, & e=uv, u\in\mathcal{T}_j, v\in\mathcal{S}_i, c_v\neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

因此，我们有 $\boldsymbol{\beta}_x = \sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}} f_e \boldsymbol{b}_e$，且 $\operatorname{supp}(\boldsymbol{f}) \subseteq \binom{\mathcal{T}_i}{2} \cup \binom{\mathcal{T}_j}{2} \cup \{uv\colon\ u\in\mathcal{S}_i,\ \ v\in\mathcal{T}_i\cup\mathcal{T}_j\}$。接下来，我们对系数的平方和进行界定。根据保留良好性，有

$$\sum_{v\in\mathcal{S}_i}\sum_{u\in\mathcal{T}_j} f_{vu}^2 = \sum_{v\in\mathcal{S}_i, c_v\neq 0} c_v^2 \big\|\boldsymbol{\alpha}_{v,\mathcal{T}_j}^{(v)}\big\|_2^2 \le \sum_{v\in\mathcal{S}_i} c_v^2 \le 1.$$

对于其余系数，我们有

$$\begin{aligned} &\sum_{e\in\binom{\mathcal{T}_i}{2}\cup\binom{\mathcal{T}_j}{2}} f_e^2 + \sum_{v\in\mathcal{S}_i}\sum_{u\in\mathcal{T}_i} f_{vu}^2 \\ &= \Big\|\sum_{v\in\mathcal{S}_i, c_v\neq 0} c_v \boldsymbol{\alpha}_{\mathcal{T}_i}^{(v)}\Big\|_2^2 + \Big\|\sum_{v\in\mathcal{S}_i, c_v\neq 0} c_v \boldsymbol{\alpha}_{\mathcal{T}_j}^{(v)}\Big\|_2^2 + \sum_{v\in\mathcal{S}_i, c_v\neq 0} c_v^2 \big\|\boldsymbol{\alpha}_{v,\mathcal{T}_i}^{(v)}\big\|_2^2 \\ &\le \Big(\sum_{v\in\mathcal{S}_i} c_v^2\Big) \sum_{v\in\mathcal{S}_i, c_v\neq 0} \Big(\big\|\boldsymbol{\alpha}_{\mathcal{T}_i}^{(v)}\big\|_2^2 + \big\|\boldsymbol{\alpha}_{\mathcal{T}_j}^{(v)}\big\|_2^2\Big) + 1 \le 2s+1, \end{aligned}$$

其中 我们使用了柯西–施瓦茨不等式和 $|\mathcal{S}_i| = s$。证明至此完成 f. □

5.3 块循环上的低成本路由

对于任意给定的整数 $t\ge 1$，令 $s \coloneqq 10^8 t$，并令 $b \coloneqq t+s+1$。现在我们来选择式 (8) 中的块索引。对于不同的块索引 i、j，如果引理 5.4 的证明中的事件 $\mathcal{E}_1$ 对这个有序对成立，就称 $i\to j$ 为一次转移。在此事件下，x 对两个预备集 $\mathcal{T}_i$、$\mathcal{T}_j$ 都是保留良好的，并且引理用 $\mathcal{S}_i$ 表示它们的根星项之差，且系数满足所述界。该转移仅取决于 x、y、$\mathcal{T}_i$、$\mathcal{T}_j$、$\mathcal{S}_i$，并且是有向的，因为它只使用源块的辅助集。

引理 5.5 (有界期望成本的路由)。 设 $\ell\ge 10$ 为偶数，t、s、b 如上所述。固定任意一个大小为 ℓb 的 $\mathcal{R}$，以及一个关联草图矩阵 $\mathbf{B} = \big[\boldsymbol{b}_e\colon\ \ e\in\binom{\mathcal{R}}{2}\big]$，其维数为 $N\le 10^{-9}\ell b t$。在均匀环标记下，存在一个事件 $\mathcal{E}$，满足 $\Pr\{\mathcal{E}\}\ge 0.999$，在该事件上我们可以选择一条递增路径

$$0 = i_0 < i_1 < \cdots < i_m = \ell/2$$

转移。该路径由标记确定，并满足

$$\mathbb{E}\Big[\sum_{j=1}^{m} L(i_j, i_{j-1})^8 \ \Big|\ \mathcal{E}\Big] \le 25\ell.$$

该引理中的代价来自（5）中的距离加权和。（3）中的跨块权重与 $L(i,\ j)^{-4}$成正比，因此它们的平方倒数贡献了一个 $L(i,\ j)^{8}$因子。对于一个满足 $i \to j$ 且 $0 \le i < j \le \ell/2$的转移，唯一的跨块系数位于 $\mathcal{S}_i$与 $\mathcal{T}_j$之间的边上，它们的循环距离均为 $L(i,\ j)$。根据（9），它们的平方和至多为一，因此它们对距离加权和的贡献满足

$$L(i,j)^8 \sum_{u\in\mathcal{S}_i}\sum_{v\in\mathcal{T}_j} f_{uv}^2 = (j-i)^8 \sum_{u\in\mathcal{S}_i}\sum_{v\in\mathcal{T}_j} f_{uv}^2 \le (j-i)^8.$$

证明。 如果在其他所有 $\ell-1$ 块 j中，至多有$(\ell-1)/100$个缺少转移$i \to j$，且至多有$(\ell-1)/100$个缺少转移 $j \to i$，则称块 i 可用。因此，可用块在任一方向上的缺失箭头都很少。令 $\mathcal{E}$ 表示这样一个事件：不可用块少于 $\ell/100$ 个，且块 0 和 $\ell/2$ 均可用。

对于任意固定的块 i，根据引理5.4 和期望的线性性，缺失的入箭头和出箭头数量的期望满足以下界。

$$\mathbb{E}\Big[\big|\{j\neq i : i \nrightarrow j\}\big| + \big|\{j\neq i : j \nrightarrow i\}\big|\Big] \le 2(\ell-1)\left(\frac{8N}{|\mathcal{R}|(t+1)} + \frac{4t}{s+1}\right).$$

如果块 i 不可用，则该总数超过$(\ell-1)/100$。因此，根据马尔可夫不等式可得 s

$$\Pr\{i \text{ is unusable}\} \le 200\left(\frac{8N}{|\mathcal{R}|(t+1)} + \frac{4t}{s+1}\right) \le 200\cdot 4.8\cdot 10^{-8},$$

其中我们使用了 $N \le 10^{-9}|\mathcal{R}|t$ 和 $s = 10^8 t$。对各块求和可得，不可用块数量的期望至多为 $200\cdot 4.8\cdot 10^{-8}\ell$。马尔可夫不等式给出了至少有$\ell/100$ 个不可用块的概率上界 $100\cdot 200\cdot 4.8\cdot 10^{-8}$。再加上两个端点块的失败概率，可得

$$1 - \Pr\{\mathcal{E}\} \le (100+2)\cdot 200\cdot 4.8\cdot 10^{-8} < 0.001.$$

在 $\mathcal{E}$ 中，如果中间块 i 可用且两条箭头都存在，则它给出一条路径 $0 \to i \to \ell/2$。不可用的块少于 $\ell/100$ 个。由于块 0 和 $\ell/2$ 可用，因此至多有 $(\ell-1)/100$ 个中间块 i 缺少箭头 $0 \to i$，至多有 $(\ell-1)/100$ 个缺少箭头 $i \to \ell/2$。因此，被排除的中间位置少于 $3\ell/100 < \ell/2 - 1$ 个，所以这样的路径存在。

在所有通过可用块和转移、从 0 到 $\ell/2$ 的递增路径中，选择访问块数量最多的一条，平局时任意选择。这一选择不允许在路径中两个相邻索引之间插入任何同时具备所需两条箭头的可用块。

只有很小一部分块无法插入，因此随机顺序使路径中相邻索引之间出现较大间隔的可能性很低。我们利用这一点来界定其成本之和的期望值 $(j-i)^8$。

选定间隔的概率。 首先对查询端点 x、y 以及无序集合 $\{(\mathcal{T}_i,\ \mathcal{S}_i):\ 0 \le i < \ell\}$进行条件化。这确定了顶点如何分组成保留–辅助对，但不确定这些对在环上的排序方式。箭头仅取决于 x、y以及这些顶点集，因此对之间的所有箭头以及每一对的可用性都已确定。各对的排序仍然是均匀的，因为在保持所有备用槽标签不变的情况下置换整个块，可在任意两个指定排序的标记之间建立双射。接下来固定位置 0、$\ell/2$处的各对。这些选择决定了 $\mathcal{E}$ 是否成立，因为我们知道不可用对的数量以及两个端点对是否可用。考虑任意满足 $\mathcal{E}$ 的此类选择。在进一步条件化下，其余各对仍然是均匀排序的。

对于固定位置 $0 \le i < j \le \ell/2$，令 X_{ij} 表示选定路径使用步 $i \to j$。进一步固定 i、j处的保留–辅助对，但不对 $X_{ij} = 1$是否成立进行条件化。如果任一对不可用，或缺少箭头 $i \to j$，则 $X_{ij} = 0$。否则，假设选定路径使用此步。任何满足 $i < k < j$ 和 $i \to k \to j$ 的可用块 k 都可以插入，同时保持其他所有路径索引不变。所得路径将访问多一个块，这与路径的选取相矛盾。因此，$X_{ij} = 1$ 要求严格位于 i 和 j 之间的每一对至少不满足一个插入条件。不可用对的数量少于 $\ell/100$，而位于 i、j 处的可用对通过缺少所需箭头，各自至多排除 $(\ell-1)/100$ 个其他对。因此，未满足插入条件的其他对少于 $3\ell/100$ 个。在当前条件化下，这些对的集合是固定的。

最多有四对已固定，位置为 0，$\ell/2$，i，j，且 i 与 j 之间没有其他位置。$j-i-1$ 内部的配对是从至少 $\ell-4$ 个剩余配对中无放回均匀抽取的样本。因此，所有这些配对都不满足插入条件的条件概率至多为

$$\Big(\frac{3\ell}{100(\ell-4)}\Big)^{j-i-1} \le 20^{-(j-i-1)},$$

其中 $\ell \ge 10$。在每次先前抽取都不满足插入条件的条件下，剩余配对中不满足条件的比例只会减少。将这些条件界相乘即可得到所示不等式；如果此类配对数量太少，无法填满该区间，则概率为零。对于所有与 $\mathcal{E}$ 一致的固定数据选择，该界均成立，因此对这些数据取平均可得 $\mathbb{E}[X_{ij} \mid \mathcal{E}] \le 20^{-(j-i-1)}$。

期望成本。在 $\mathcal{E}$ 上，选定路径的每一步都恰好对应一个指示变量 X_{ij}。因此，

$$\sum_{a=1}^{m} L(i_a, i_{a-1})^8 = \sum_{0 \le i < j \le \ell/2} (j-i)^8 X_{ij}.$$

取条件期望，应用关于 $\mathbb{E}[X_{ij} \mid \mathcal{E}]$ 的界，并按$d = j - i$对配对分组，可得

$$\begin{aligned}
\mathbb{E}\left[\sum_{a=1}^{m} L(i_a, i_{a-1})^8 \,\middle|\, \mathcal{E}\right] &= \sum_{0 \le i < j \le \ell/2} (j-i)^8 \; \mathbb{E}[X_{ij} \mid \mathcal{E}] \\
&\le \sum_{0 \le i < j \le \ell/2} (j-i)^8 \, 20^{-(j-i-1)} \\
&= \sum_{d=1}^{\ell/2} \sum_{i=0}^{\ell/2-d} d^8 \, 20^{-(d-1)} \\
&= \sum_{d=1}^{\ell/2} (\ell/2 + 1 - d) \, d^8 \, 20^{-(d-1)} \\
&\le \frac{\ell}{2} \sum_{d=1}^{\ell/2} d^8 \, 20^{-(d-1)} \le \frac{\ell}{2} \sum_{d \ge 1} d^8 \, 20^{-(d-1)}.
\end{aligned}$$

这里，$\ell/2 + 1 - d$ 统计距离为 d 的位置对。从 $d = 3$ 开始，连续求和项的比值至多为 $(4/3)^8/20 < 1/2$，因此

$$\sum_{d \ge 1} d^8 \, 20^{-(d-1)} \le 1 + \frac{256}{20} + \frac{2 \cdot 6561}{400} < 50.$$

由此可得 $\mathbb{E}\big[\sum_{j=1}^{m} L(i_j,\ i_{j-1})^8 \mid \mathcal{E}\big] \le 25\ell$，如所述。 □

引理的证明 *5.2*。令 $\mathcal{E}$ 为引理中的成功事件5.5，因此$\Pr\{\mathcal{E}\} \ge 0.999$。在 $\mathcal{E}$上，取该引理给出的路径 $0 = i_0 < \cdots < i_m = \ell/2$。根据保留集分解（7），（8）中的首项和末项分别在块 0和 $\ell/2$内有表示，且各自的系数平方范数至多为一。

对（8）中的每次转移应用引理 5.4。将所得系数向量与两个端点项对应的系数向量相加，得到 $\boldsymbol{f}$ 满足

$$\boldsymbol{b}_q = \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} f_e \boldsymbol{b}_e.$$

当 $\boldsymbol{b}_e = \mathbf{0}$时，令 $f_e = 0$。

系数平方范数。 保留集内的一条边至多出现在两个表示中：分别对应进入和离开转移的表示；或者在块 0中，出现在源项和第一次转移的表示中。由于保留集与辅助集互不相交，并且 x，y 占据备用槽位，因此其他边至多出现一次。路径单调递增，故其包含 $m \le \ell/2$ 次转移。对重叠系数应用 $(a+b)^2 \le 2a^2 + 2b^2$，并使用引理 5.4 中的界，得到

$$\|\boldsymbol{f}\|_2^2 \le 2\big(2 + m(2s+2)\big) \le 4 + 2\ell(s+1) \le 2\ell b.$$

最后一个不等式用到了 $b = t + s + 1$ 和 $2\ell t \ge 4$。

距离加权系数和。不同转移所使用的跨块边彼此不相交：每条边都连接源辅助集和目标保留集，而沿递增路径，各源辅助集互不相同。对于转移 j，它们的循环块距离为 $L(i_j, i_{j-1}) = i_j - i_{j-1}$，因为该路径位于环的前半部分。结合 (9) 得到

$$\sum_{0 \le i < j < \ell} L(i,j)^8 \sum_{\substack{u \in \mathcal{B}_i, v \in \mathcal{B}_j, \\ uv \ne q}} f_{uv}^2 \le \sum_{j=1}^{m} L(i_j, i_{j-1})^8.$$

此比较对 $\mathcal{E}$中的每种标记方式都成立。取条件期望并应用引理 5.5 即可证明 (5)。系数由标记方式确定，因为坐标顺序和打破平局规则都是固定的；令 $\boldsymbol{f} = \mathbf{0}$ off $\mathcal{E}$。 □

5.4 证明关联草图的下界

定理 *5.1* 的证明。任选一个整数 $t \ge 1$ 并令 $b := (10^8 + 1)t + 1$, $\ell := 10^7 b^4$, 并令 $n := \ell b = \Theta(t^5)$。假设该草图的维数为 $N \le 10^{-9} nt$。注意 $nt = \Theta(n^{6/5})$。由引理4.5，因此只需证明：对于每个这样的固定关联草图矩阵 $\mathbf{B}$,

$$\mathbb{E}_{\pi}\Big[\min\Big\{1, \mathrm{lev}_{\boldsymbol{b}_q}\Big(\big(\mathbf{B}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big)\Big\}\Big] < \frac{1}{100}.$$

令 $\mathcal{E}$ 和 $\boldsymbol{f} = (f_e)_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}}$ 分别为引理5.2所给出的成功事件和系数向量。则$\Pr\{\mathcal{E}\} \ge 0.999$，并且当 $\boldsymbol{f} = \mathbf{0}$ 不发生时 $\mathcal{E}$。在 $\mathcal{E}$上，这些系数使用其他列表示 $\boldsymbol{b}_q$，并满足 $\|\boldsymbol{f}\|_2^2 \le 2\ell b$。根据杠杆分数的定义，在此事件上我们有

$$\mathrm{lev}_{\boldsymbol{b}_q}\Big(\big(\mathbf{B}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big) \le \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} \frac{f_e^2}{w_e^2} \le \frac{2}{\ell^3 b^3} + \sum_{0 \le i < j < \ell} \sum_{\substack{u \in \mathcal{B}_i, v \in \mathcal{B}_j, \\ uv \ne q}} \frac{f_{uv}^2}{w_{uv}^2},$$

其中第二个不等式使用 $w_{\text{int}} = \ell^2 b^2$ 和 $\|\boldsymbol{f}\|_2^2 \le 2\ell b$ 来界定块内贡献。由于 $\boldsymbol{f} = \mathbf{0}$ 在 $\mathcal{E}$之外，跨块基础权重和引理5.2 给出

$$\begin{aligned}
&\mathbb{E}_\pi\left[\sum_{0\le i<j<\ell}\ \sum_{\substack{u\in\mathcal{B}_i, v\in\mathcal{B}_j,\\ uv\neq q}} \frac{f_{uv}^2}{w_{uv}^2}\right] \\
&= \frac{\Pr\{\mathcal{E}\}}{w_{\text{ext}}^2}\,\mathbb{E}_\pi\left[\sum_{0\le i<j<\ell} L(i,j)^8 \sum_{\substack{u\in\mathcal{B}_i, v\in\mathcal{B}_j,\\ uv\neq q}} f_{uv}^2 \,\middle|\, \mathcal{E}\right] \le \frac{25\ell}{w_{\text{ext}}^2} = \frac{6400\,b^4}{\ell}. \qquad (11)
\end{aligned}$$

如果 $\mathcal{E}$ 失败，我们就对截断杠杆得分使用平凡上界 1。因此，将这一失败概率与上述界结合起来，可得

$$\mathbb{E}_\pi\left[\min\left\{1, \operatorname{lev}_{\boldsymbol{b}_q}\left(\left(\mathbf{B}\operatorname{diag}(\boldsymbol{w})\right)_{-q}\right)\right\}\right] \le \left(1-\Pr\{\mathcal{E}\}\right) + \frac{2}{\ell^3 b^3} + \frac{6400\,b^4}{\ell} < \frac{1}{100},$$

其中，最后一个不等式由 $\ell = 10^7 b^4$ 和 $b \ge 10^8$得出。

最后，通过向困难实例图添加孤立顶点，我们可以将下界扩展到所有足够大的 n。定理 5.1的证明至此完成。 □

6 任意线性草图的下界

本节 证明了任意线性草图的下界，具体如定理 1。 1.

6.1 向困难分布添加锚点团

我们在第 4 节的环构造中加入一个锚点团，且团与环之间没有边。锚点列使我们能够将查询列表示问题归约为环上的关联草图。我们令团的规模与环相当，以使该关联草图的期望维数保持在原草图维数的常数倍以内。较大的锚点权重使额外的锚点杠杆贡献保持较小。

给定整数 $b \ge 2$ 和偶整数 $\ell \ge 2$，令 $n := 2\ell b$。这里 n 统计环顶点和锚点顶点。取一个模板，其中包含 ℓb 个锚点槽位，以及 ℓ 个环块，每个环块有 b 个槽位，并设有与第 4 节相同的两个指定查询槽位。按照固定顺序为这 n 个槽位编号，并将顶点 $\pi(t)$ 放入槽位 t，其中 π 是 $[n]$的均匀随机置换。记 $\mathcal{C}$ 为锚点顶点集，$\mathcal{R} := [n] \setminus \mathcal{C}$ 为环顶点集，并将环块记为 $\mathcal{B}_0$、……、$\mathcal{B}_{\ell-1}$。因此$|\mathcal{C}| = |\mathcal{R}| = \ell b$。查询边为 $q := xy$，其中 x、y 分别占据块 0 和 $\ell/2$ 中的指定槽位。

我们仍使用 $w_{\text{ext}} := \ell/(16b^2)$，但根据扩大的顶点数量设置 $w_{\text{int}} := n^2$。现在，边集和基准权重为

$$\begin{aligned}
E &:= \binom{\mathcal{C}}{2} \cup \binom{\mathcal{R}}{2}, \\
w_e &:= \begin{cases} w_{\text{int}}, & e \in \binom{\mathcal{C}}{2} \text{ or } e \in \binom{\mathcal{B}_i}{2} \text{ for some } i, \\ w_{\text{ext}}\, L(i,j)^{-4}, & e = uv, u \in \mathcal{B}_i, v \in \mathcal{B}_j, i \neq j, \\ 0, & \text{otherwise.} \end{cases} \qquad (12)
\end{aligned}$$

在给定 π的条件下，对每个 $e \in E$独立地从密度为（4）的分布中采样 Z_e。独立地以 1/2的概率选择是 或否，每种情况的概率均为 1/2。在 否 的情况下，设置 $\widetilde{w}_e := w_e Z_e$ 对于 $e \in E$；在 是 的情况下，设置

$\widetilde{w}_e := w_e Z_e + \mathbf{1}\{e = q\}$。在两种情况下， E 之外的所有坐标均为零。在本节中，是和否均指这些分布，而 $\widetilde{\boldsymbol{w}}$ 表示由此得到的输入权重向量。

任一分布中的图，其正权重比至多为 $4n^5$，这与引理 4.1证明中的相同计算一致。事实上，(12)给出 $1/n^3 \le 1/(\ell^3 b^2) \le w_e \le n^2$ 对每个 $e \in E$均成立。由于$Z_e \in [1/2, 3/2]$ ，且查询边权重可能增加 $\mathbf{1}$，因此每个正输入权重都位于$[1/(2n^3)\ \ 2n^2]$中。

下一个引理表明，在添加锚点团后，引理 4.5 中的杠杆分数约简仍然有效。

引理 6.1。 如果一个线性草图满足

$$\mathbb{E}_{\pi}\left[\min\left\{1, \mathrm{lev}_{\boldsymbol{a}_q}\left(\left(\mathbf{A}\,\mathrm{diag}(\boldsymbol{w})\right)_{-q}\right)\right\}\right] < \frac{1}{100} \tag{13}$$

对于其草图矩阵的每个实现$\mathbf{A}$，并且 (12)中的基础权重为(1 ± 0.01)-谱近似，那么该草图的任何恢复算法都无法以至少 0.99的概率，在锚定分布的支撑集中的每个输入上进行计算。

证明。 我们检查引理 4.5 证明中的两个要素仍然有效。将引理 4.2 证明中的势 ϕ 在 $\mathcal{C}$上定义为零。锚点边和块内边对其二次型的贡献为零，因为它们的端点具有相同的势。在$\mathcal{C}$ 与 $\mathcal{R}$之间不存在边，并且跨块权重以及 q 上的单位增量均未改变。因此，引理 4.2 中的谱分离仍然成立，只要恢复成功，其证明中构造的区分算法就仍然正确。

固定一个实现 $\mathbf{A}$ 并对其进行条件化 π。坐标 E 对草图没有任何贡献，并对应于 $\mathbf{A}\mathrm{diag}(\boldsymbol{w})$ 的零列，因此删除这些坐标不会改变杠杆分数。在 E 上，所有锚点噪声和环噪声均相互独立，并且是 与否 仅在 q 上增加了单位值。因而，引理 4.4 的证明在限制到这些坐标后仍给出相同的全变差距离界。

按照引理 π 4.5 的证明对 进行平均，杠杆假设给出 $1/\sqrt{10}$ 与 $(\pi,\ \mathbf{A}\widetilde{\boldsymbol{w}})_{\mathsf{Yes}}$ 之间的全变差距离小于 $(\pi,\ \overline{\mathbf{A}\,\widetilde{\boldsymbol{w}}})_{\mathsf{No}}$，对于每个固定的 $\mathbf{A}$ 都成立。由于两种情况等可能，(2) 表明，对于给定 $\mathbf{A}$、**草图和** π 的任意随机化区分算法，其成功概率小于 $3/4$ 。对独立采样的草图矩阵进行平均后，该界仍然成立。

现在假设某个恢复算法满足所述保证。根据上述谱分离，将区分算法应用于其输出和 π 时，只要恢复成功，该算法就是正确的。因此，对输入进行平均后，在每个分布下成功概率至少为 0.99，这就产生了矛盾。

□

6.2 通过关联草图表示任意草图

对于 $\mathcal{C}$ 和 $\mathcal{R}$ 的任意实现，我们将通过关联草图矩阵 $\mathbf{B} = [\boldsymbol{b}_e$ 表示任意草图矩阵 $\mathbf{A}$：$e \in \binom{\mathcal{R}}{2}]$ 在 $\mathcal{R}$ 上，这意味着 $\mathbf{B}$ 的每一行的支撑集均为 $\binom{\mathcal{R}}{2}$ 中与 $\mathcal{R}$ 中某个单一顶点关联的边（不同的行可能对应不同的顶点）。

对于 $\mathcal{C}$ 和 $\mathcal{R}$ 的任意实现，我们定义 $\mathbf{A}_{\mathcal{C}} := \left[\boldsymbol{a}_e :\ e \in \binom{\mathcal{C}}{2}\right]$ 和 $\mathbf{A}_{v,\ \mathcal{C}} := [\boldsymbol{a}_{vw} :\ w \in \mathcal{C}]$ ，其中$v \in \mathcal{R}$。如果一条边 $uv \in \binom{\mathcal{R}}{2}$ 满足 $\mathrm{lev}_{\boldsymbol{a}_{uv}}\left([\mathbf{A}_{u,\ \mathcal{C}}\,\mathbf{A}_{v,\ \mathcal{C}}\,\mathbf{A}_{\mathcal{C}}]\right) \le 1$，我们就称其为锚点良好。利用查询边 q 在给定 $\mathcal{C} \cup \{x, y\}$ 条件下的均匀性，我们立即得到 q 非锚点良好的概率上界。

引理 6.2。 对于任意固定的草图矩阵 $\mathbf{A}$，有

$$\Pr_\pi\{q \text{ is not anchor-good}\} \le 2N \Big/ \binom{\ell b + 2}{2}. \tag{14}$$

证明。 给定顶点集 $\mathcal{C}\cup\{x,y\}$，查询边 $q = xy$是其所在团中的一条均匀随机边，因此应用引理3.1 可得

$$\Pr_\pi\Big\{\mathrm{lev}_{\boldsymbol{a}_q}\big([\mathbf{A}_{x,\mathcal{C}}\ \mathbf{A}_{y,\mathcal{C}}\ \mathbf{A}_{\mathcal{C}}]\big) > 1\Big\} \le 2\,\mathrm{rank}(\mathbf{A}) \Big/ \binom{|\mathcal{C}|+2}{2} \le 2N \Big/ \binom{\ell b+2}{2}. \qquad \square$$

为了通过关联草图表示任意草图，我们首先介绍一些额外的记号，用于定义关联草图的维数。对于 $\mathcal{C}$ 的任意实现，我们定义 $k_v(\mathcal{C}) := n_+\big(\mathbf{A}_{v,\mathcal{C}}\mathbf{A}_{v,\mathcal{C}}^\top - \mathbf{A}_{\mathcal{C}}\mathbf{A}_{\mathcal{C}}^\top\big)$ 对于 $v \in \mathcal{R}$，其中 $n_+(\mathbf{M})$ 表示对称矩阵 $\mathbf{M}$ 的正特征值数量。随后我们定义 $K(\mathcal{C}) := \sum_{v\in\mathcal{R}} k_v(\mathcal{C})$。

下一个引理表明，草图矩阵$\mathbf{A}$中对应于锚点-良好环边的每一列，都可以分解为 $\mathcal{R}$ 上维数为 $K(\mathcal{C})$ 的关联草图中相应列的线性变换，加上锚点列的线性组合。

引理 6.3。 考虑 $\mathcal{C}$ 和 $\mathcal{R}$ 的任意实现以及任意草图矩阵 $\mathbf{A}$。存在一个关联草图矩阵 $\mathbf{B} = \big[\boldsymbol{b}_e : e \in \binom{\mathcal{R}}{2}\big]$，其定义在 $\mathcal{R}$ 上且维数为 $K(\mathcal{C})$；还存在一个矩阵 $\mathbf{T} \in \mathbb{R}^{N\times K(\mathcal{C})}$ 以及向量 $(\boldsymbol{c}_e)_{e\in\binom{\mathcal{R}}{2}}$，它们属于 $\mathbb{R}^{\binom{\ell b}{2}}$，均由 $\mathbf{A}$ 和 $\mathcal{C}$ 决定，并满足以下条件：

(1) 对于每条锚点良好环边 e, $\boldsymbol{a}_e = \mathbf{T}\,\boldsymbol{b}_e + \mathbf{A}_{\mathcal{C}}\boldsymbol{c}_e$ 和 $\|\boldsymbol{c}_e\|_2^2 \le 3$;

(2) 对于每条非锚点良好环边 e, $\boldsymbol{b}_e = \mathbf{0}_{K(\mathcal{C})}$ 和 $\boldsymbol{c}_e = \mathbf{0}_{\binom{\ell b}{2}}$。

证明。 对于任意环顶点 v，令 $\mathbf{T}_v \in \mathbb{R}^{N\times k_v(\mathcal{C})}$ 的列为对应于$\mathbf{A}_{v,\mathcal{C}}\mathbf{A}_{v,\mathcal{C}}^\top - \mathbf{A}_{\mathcal{C}}\mathbf{A}_{\mathcal{C}}^\top$正特征值的正交归一特征向量，并将每个向量按其特征值的平方根缩放。根据特征分解，这保证了$\mathbf{A}_{v,\mathcal{C}}\mathbf{A}_{v,\mathcal{C}}^\top - \mathbf{A}_{\mathcal{C}}\mathbf{A}_{\mathcal{C}}^\top \preceq \mathbf{T}_v\mathbf{T}_v^\top$，因此我们有

$$\mathbf{A}_{v,\mathcal{C}}\mathbf{A}_{v,\mathcal{C}}^\top \preceq \mathbf{T}_v\mathbf{T}_v^\top + \mathbf{A}_{\mathcal{C}}\mathbf{A}_{\mathcal{C}}^\top. \tag{15}$$

按顶点标签固定 $\mathcal{R}$ 中顶点的顺序为 $v_1,\dots,v_{\ell b}$，并将所有矩阵连接为$\mathbf{T} := [\mathbf{T}_{v_1}\ \cdots\ \mathbf{T}_{v_{\ell b}}] \in \mathbb{R}^{N\times K(\mathcal{C})}$。令 $\mathcal{I}_v$ 为$\mathbf{T}_v$所占据的 $k_v(\mathcal{C})$ 个坐标构成的集合。

对于任意非锚点良好环边 uv，我们可以直接按要求设置 $\boldsymbol{b}_{uv} := \mathbf{0}$ 以及 $\boldsymbol{c}_{uv} := \mathbf{0}$。现在考虑任意锚点良好环边 uv。根据锚点良好性的定义，有$\mathrm{lev}_{\boldsymbol{a}_{uv}}\big([\mathbf{A}_u,\ \mathcal{C}\ \mathbf{A}_v,\ \mathcal{C}\ \mathbf{A}_{\mathcal{C}}]\big) \le 1$，因此根据引理3.2，

$$\begin{aligned}\boldsymbol{a}_{uv}\boldsymbol{a}_{uv}^\top &\preceq \mathbf{A}_{u,\mathcal{C}}\mathbf{A}_{u,\mathcal{C}}^\top + \mathbf{A}_{v,\mathcal{C}}\mathbf{A}_{v,\mathcal{C}}^\top + \mathbf{A}_{\mathcal{C}}\mathbf{A}_{\mathcal{C}}^\top \\ &\preceq \mathbf{T}_u\mathbf{T}_u^\top + \mathbf{T}_v\mathbf{T}_v^\top + 3\mathbf{A}_{\mathcal{C}}\mathbf{A}_{\mathcal{C}}^\top,\end{aligned}$$

其中我们对 u 和 v 都代入（15）。现在，根据引理 3.2，$\mathrm{lev}_{\boldsymbol{a}_{uv}}\big([\mathbf{T}_u\ \mathbf{T}_v\ \sqrt{3}\mathbf{A}_{\mathcal{C}}]\big) \le 1$，因此存在 $\boldsymbol{\beta}_u \in \mathbb{R}^{k_u(\mathcal{C})}$、$\boldsymbol{\beta}_v \in \mathbb{R}^{k_v(\mathcal{C})}$ 和 $\boldsymbol{c}_{uv} \in \mathbb{R}^{\binom{\ell b}{2}}$，使得

$$\boldsymbol{a}_{uv} = \mathbf{T}_u\boldsymbol{\beta}_u + \mathbf{T}_v\boldsymbol{\beta}_v + \mathbf{A}_{\mathcal{C}}\boldsymbol{c}_{uv}$$

以及$\|\boldsymbol{\beta}_u\|_2^2 + \|\boldsymbol{\beta}_v\|_2^2 + \frac{1}{3}\|\boldsymbol{c}_{uv}\|_2^2 \le 1$。鉴于此，定义 $\boldsymbol{b}_{uv} \in \mathbb{R}^{K(\mathcal{C})}$：将 $\boldsymbol{\beta}_u$ 置于 $\mathcal{I}_u$ 中，将$\boldsymbol{\beta}_v$ 置于 $\mathcal{I}_v$ 中，并将所有其他坐标置零。这保证了 $\mathrm{supp}(\boldsymbol{b}_{uv}) \subseteq \mathcal{I}_u \cup \mathcal{I}_v$，因此 $\mathbf{B}$ 是定义在 $\mathcal{R}$ 上的关联草图矩阵，其维数为 $K(\mathcal{C})$。于是，根据构造可得 $\boldsymbol{a}_{uv} = \mathbf{T}\,\boldsymbol{b}_{uv} + \mathbf{A}_{\mathcal{C}}\,\boldsymbol{c}_{uv}$，并且有 $\|\boldsymbol{c}_{uv}\|_2^2 \le 3$，正如所需。 $\square$

作为这一端点分解方案的结果，下面的引理表明，如果 $\boldsymbol{b}_q$ 可以表示为其他 $\boldsymbol{b}$ 列的线性组合，且系数较小，那么 $\boldsymbol{a}_q$ 也可以表示为其他 $\boldsymbol{a}$列的线性组合，且系数较小。

引理 *6.4*（提升）。固定 $\mathcal{C}$ 和 $\mathcal{R}$，令 $B = [\boldsymbol{b}_e : e \in \binom{\mathcal{R}}{2}]$ 为引理 *6.3* 提供的关联草图矩阵。令 $q \in \binom{\mathcal{R}}{2}$ 为任意一条锚点良好环边。假设系数 $\boldsymbol{f} = (f_e)_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}}$ 满足 $\boldsymbol{b}_q = \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} f_e \boldsymbol{b}_e$，并且当 $\boldsymbol{b}_e = \boldsymbol{0}$ 时满足 $f_e = 0$。那么存在一个向量 $\boldsymbol{c} \in \mathbb{R}^{\binom{\ell b}{2}}$，使得

$$\boldsymbol{a}_q = \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} f_e \boldsymbol{a}_e + \mathbf{A}_{\mathcal{C}} \boldsymbol{c} \tag{16}$$

and

$$\|\boldsymbol{c}\|_2^2 \le 6\big(1 + \ell^2 b^2 \|\boldsymbol{f}\|_2^2\big).$$

证明。令 $\mathbf{T}$ 和 $(\boldsymbol{c}_e)$ 为引理 $(\boldsymbol{b}_e)$ 中附带的矩阵和向量，6.3我们有$\boldsymbol{a}_e = \mathbf{T}\,\boldsymbol{b}_e + \mathbf{A}_{\mathcal{C}} \boldsymbol{c}_e$ 对于每一条锚点良好环边 e。由于 $\boldsymbol{b}_e = \boldsymbol{0}$ 对于任意非锚点良好环边e，根据引理陈述中对 $\boldsymbol{f}$ 的要求，$f_e \neq 0$ 意味着 e 是锚点良好的。因此，我们有

$$\begin{aligned}\sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} f_e \boldsymbol{a}_e &= \mathbf{T} \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} f_e \boldsymbol{b}_e + \mathbf{A}_{\mathcal{C}} \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} f_e \boldsymbol{c}_e \\ &= \mathbf{T}\,\boldsymbol{b}_q + \mathbf{A}_{\mathcal{C}} \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} f_e \boldsymbol{c}_e.\end{aligned}$$

由于 q 是锚点良好，我们有 $\boldsymbol{a}_q = \mathbf{T}\,\boldsymbol{b}_q + \mathbf{A}_{\mathcal{C}} \boldsymbol{c}_q$。将两个方程相减后，$\mathbf{T}\,\boldsymbol{b}_q$ 项相互抵消，重新整理可得

$$\boldsymbol{a}_q = \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} f_e \boldsymbol{a}_e + \mathbf{A}_{\mathcal{C}} \left(\boldsymbol{c}_q - \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} f_e \boldsymbol{c}_e \right).$$

现在令 $\boldsymbol{c} := \boldsymbol{c}_q - \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} f_e \boldsymbol{c}_e$，于是方程（16）成立。由于 $\|\boldsymbol{c}_e\|_2^2 \le 3$ 对每条环边 e均成立，应用三角不等式和柯西–施瓦茨不等式可得

$$\|\boldsymbol{c}\|_2^2 \le 3\left(1 + \sum_{e \in \binom{\mathcal{R}}{2} \setminus \{q\}} |f_e|\right)^2 \le 6\big(1 + \ell^2 b^2 \|\boldsymbol{f}\|_2^2\big),$$

至此，证明完成。 □

接下来证明 $\mathbb{E}\big[K(\mathcal{C})\big]$，即引理 **B** 6.3 中构造的关联草图的期望维数 不超过 。

引理 6.5。对于任意固定的草图矩阵 $\mathbf{A}$，我们有 $\mathbb{E}_{\mathcal{C}}\big[K(\mathcal{C})\big] \le 4N$。

为证明引理 6.5，我们需要以下初等引理：它通过迹给出了两个半正定矩阵之差的正特征值数量的上界。

引理 6.6。对于任意维数相同的矩阵 $\mathbf{M}$、$\mathbf{N} \succeq \mathbf{0}$，我们有

$$n_+(\mathbf{M} - \mathbf{N}) \le 2\,\mathrm{tr}\big((\mathbf{M} + \mathbf{N})^{\dagger}\mathbf{M}\big).$$

证明。令 $\mathbf{W} \coloneqq (\mathbf{M}+\mathbf{N})^{\dagger/2}\,\mathbf{M}\,(\mathbf{M}+\mathbf{N})^{\dagger/2}$。由于 $\mathbf{M}$、$\mathbf{N} \succeq \mathbf{0}$，我们有 $\operatorname{range}(\mathbf{M}) \subseteq \operatorname{range}(\mathbf{M}+\mathbf{N})$。由于 $(\mathbf{M}+\mathbf{N})^{1/2}(\mathbf{M}+\mathbf{N})^{\dagger/2}$ 是 $\operatorname{range}(\mathbf{M}+\mathbf{N})$ 上的正交投影，这意味着 $(\mathbf{M}+\mathbf{N})^{1/2}\,\mathbf{W}\,(\mathbf{M}+\mathbf{N})^{1/2} = \mathbf{M}$。因此，我们可以写成

$$\mathbf{M}-\mathbf{N} = 2\mathbf{M}-(\mathbf{M}+\mathbf{N}) = (\mathbf{M}+\mathbf{N})^{1/2}\big(2\mathbf{W}-\mathbf{I}\big)(\mathbf{M}+\mathbf{N})^{1/2}. \tag{17}$$

注意，$(\mathbf{M}+\mathbf{N})^{1/2}$ 在 $\operatorname{range}(\mathbf{M}+\mathbf{N})$ 上可逆，并且 $\mathbf{M}-\mathbf{N}$ 和 $\mathbf{M}$ 在 $\operatorname{range}(\mathbf{M}+\mathbf{N})^{\perp}$ 上都是零映射。因此，我们有

$$\begin{aligned} n_+(\mathbf{M}-\mathbf{N}) &= n_+\big((\mathbf{M}-\mathbf{N})|_{\operatorname{range}(\mathbf{M}+\mathbf{N})}\big) \\ &= n_+\Big(\big(2\mathbf{W}-\mathbf{I}\big)\big|_{\operatorname{range}(\mathbf{M}+\mathbf{N})}\Big) = n_+\big(2\mathbf{W}-\mathbf{I}\big), \end{aligned}$$

其中，第二个等式是将西尔维斯特惯性定律应用于 $\operatorname{range}(\mathbf{M}+\mathbf{N})$ 上的公式（17）得到的。因此，我们已经证明，$n_+(\mathbf{M}-\mathbf{N})$ 等于 $\mathbf{W}$ 的大于 $1/2$ 的特征值个数。由于 $\mathbf{W}$ 的特征值之和 $\succeq \mathbf{0}$ 为 $\operatorname{tr}(\mathbf{W}) = \operatorname{tr}\big((\mathbf{M}+\mathbf{N})^{\dagger}\mathbf{M}\big)$，我们得出结论：$n_+(\mathbf{M}-\mathbf{N}) \le 2\operatorname{tr}\big((\mathbf{M}+\mathbf{N})^{\dagger}\mathbf{M}\big)$，如所需。 □

引理的证明 *6.5*。我们选择一个均匀顶点 $v \in \mathcal{R}$，条件是 $\mathcal{C}$，并旨在给出 $\mathbb{E}\big[k_v(\mathcal{C})\big]$ 的上界。考虑在集合 $\mathcal{C}' \coloneqq \mathcal{C} \cup \{v\}$ 的条件下，此时顶点 v 是 $\mathcal{C}'$ 中的一个均匀顶点，且锚点集为 $\mathcal{C}' \setminus \{v\}$。记 $\mathbf{A}_{\mathcal{C}'} \coloneqq \big[\boldsymbol{a}_e : e \in \binom{\mathcal{C}'}{2}\big]$。对于每个 $u \in \mathcal{C}'$，我们可以推导出

$$\begin{aligned} k_u\big(\mathcal{C}' \setminus \{u\}\big) &= n_+\Bigg(\sum_{w \in \mathcal{C}'\setminus\{u\}} \boldsymbol{a}_{uw}\boldsymbol{a}_{uw}^{\top} - \sum_{e \in \binom{\mathcal{C}'\setminus\{u\}}{2}} \boldsymbol{a}_e\boldsymbol{a}_e^{\top}\Bigg) \\ &\le 2\operatorname{tr}\Bigg(\big(\mathbf{A}_{\mathcal{C}'}\mathbf{A}_{\mathcal{C}'}^{\top}\big)^{\dagger} \sum_{w \in \mathcal{C}'\setminus\{u\}} \boldsymbol{a}_{uw}\boldsymbol{a}_{uw}^{\top}\Bigg) \\ &= 2\sum_{w \in \mathcal{C}'\setminus\{u\}} \operatorname{tr}\Big(\boldsymbol{a}_{uw}^{\top}\big(\mathbf{A}_{\mathcal{C}'}\mathbf{A}_{\mathcal{C}'}^{\top}\big)^{\dagger}\boldsymbol{a}_{uw}\Big) \\ &= 2\sum_{w \in \mathcal{C}'\setminus\{u\}} \operatorname{lev}_{\boldsymbol{a}_{uw}}(\mathbf{A}_{\mathcal{C}'}), \end{aligned}$$

其中第一行使用了 $k_u\big(\mathcal{C}' \setminus \{u\}\big)$ 的定义，第二行应用了引理 6.6，第三行使用了迹的性质，最后一行使用了杠杆分数的公式。因此，对 $u \in \mathcal{C}'$ 求和可得

$$\sum_{u \in \mathcal{C}'} k_u\big(\mathcal{C}' \setminus \{u\}\big) \le 4\sum_{e \in \binom{\mathcal{C}'}{2}} \operatorname{lev}_{\boldsymbol{a}_e}(\mathbf{A}_{\mathcal{C}'}) = 4\operatorname{rank}(\mathbf{A}_{\mathcal{C}'}) \le 4N.$$

因此，先对均匀选取的指定顶点 $v \in \mathcal{C}'$ 取平均，再对 $\mathcal{C}'$ 取平均，可得 $\mathbb{E}\big[k_v(\mathcal{C})\big] \le 4N/(\ell b+1)$。由于在给定 $\mathcal{C}$ 的条件下，v 在 $\mathcal{R}$ 中均匀分布，我们有 $\mathbb{E}_{\mathcal{C}}\big[K(\mathcal{C})\big] = \ell b\,\mathbb{E}\big[k_v(\mathcal{C})\big] \le 4N$，正如所需。 □

6.3 证明任意线性草图的下界

现在我们已经准备好证明主要结果，即任意线性草图的加权谱稀疏化草图维数下界。

定理 *1.1* 的证明。 任取一个整数 $t \ge 1$，并设 $b := (10^8+1)t+1$，$\ell := 10^7 b^4$，以及 $n := 2\ell b = \Theta(t^5)$。假设草图的维数为 $N \le 10^{-12} nt$。注意，$nt = \Theta(n^{6/5})$。所构造输入的正权重比为 $O(n^5)$，因此根据引理 6.1，只需证明：对于任意这样固定的草图矩阵 $\mathbf{A}$，

$$\mathbb{E}_\pi\Big[\min\Big\{1, \mathrm{lev}_{\boldsymbol{a}_q}\Big(\big(\mathbf{A}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big)\Big\}\Big] < \frac{1}{100}.$$

首先揭示锚点集 $\mathcal{C}$，并保持环标记均匀。根据引理6.5和马尔可夫不等式，我们有

$$\Pr_{\mathcal{C}}\big\{K(\mathcal{C}) > 10^{-9}\ell bt\big\} \le \frac{10^9\,\mathbb{E}_{\mathcal{C}}[K(\mathcal{C})]}{\ell bt} \le \frac{8\cdot 10^9 N}{nt}.$$

对于每个固定的 $\mathcal{C}$，满足 $K(\mathcal{C}) \le 10^{-9}\ell bt$，引理 6.3 给出了一个维数为 $K(\mathcal{C})$ 的固定关联草图矩阵 B，在均匀环标记下可应用引理 5.2。令 $\mathcal{E}$ 和 $\boldsymbol{f} = (f_e)_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}}$ 分别表示引理 5.2 为每个此类$\mathcal{C}$提供的成功事件和系数向量。对于所有其他锚点集，声明 $\mathcal{E}$ 为假，并在 $\mathcal{E}$ 不发生时令 $\boldsymbol{f} = \mathbf{0}$。那么当 $K(\mathcal{C}) \le 10^{-9}\ell bt$ 时，$\Pr\{\mathcal{E} \mid \mathcal{C}\} \ge 0.999$，从而

$$1 - \Pr\{\mathcal{E}\} \le \frac{8\cdot 10^9 N}{nt} + 10^{-3}.$$

引理 5.2 提供的系数满足 $f_e = 0$ 当 $\boldsymbol{b}_e = \mathbf{0}$ 且 $\|\boldsymbol{f}\|_2^2 \le 2\ell b$时。因此，如果 $\mathcal{E}$ 发生，且 q 为锚点良好，则应用引理 6.4 可得到系数 $\boldsymbol{c} = (c_e)_{e\in\binom{\mathcal{C}}{2}}$，使得

$$\boldsymbol{a}_q = \sum_{e\in\binom{\mathcal{R}}{2}\setminus\{q\}} f_e \boldsymbol{a}_e + \mathbf{A}_{\mathcal{C}}\boldsymbol{c},$$
$$\|\boldsymbol{c}\|_2^2 \le 6\big(1 + \ell^2 b^2 \|\boldsymbol{f}\|_2^2\big) \le 6 + 12\ell^3 b^3.$$

所有锚点边和块内边的基准权重均为 $w_{\mathrm{int}} = n^2$，因此，缩放列 $w_e\boldsymbol{a}_e$ 对系数平方范数的总贡献至多为

$$\frac{\|\boldsymbol{c}\|_2^2 + \|\boldsymbol{f}\|_2^2}{w_{\mathrm{int}}^2} \le \frac{6 + 12\ell^3 b^3 + 2\ell b}{16\ell^4 b^4} \le \frac{1}{\ell b}.$$

因此，在该事件下，根据杠杆分数的定义可得

$$\mathrm{lev}_{\boldsymbol{a}_q}\Big(\big(\mathbf{A}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big) \le \frac{1}{\ell b} + \sum_{0\le i<j<\ell}\ \sum_{\substack{u\in\mathcal{B}_i, v\in\mathcal{B}_j,\\ uv\ne q}} \frac{f_{uv}^2}{w_{uv}^2}.$$

对于每个固定的 $\mathcal{C}$ 满足 $K(\mathcal{C}) \le 10^{-9}\ell bt$，在给定11的条件下，跨块估计（）成立，因为 $\mathcal{C}$，由于$\mathbf{B}$是固定的，环标记是均匀的，且跨块基础权重保持不变。由于$\boldsymbol{f} = \mathbf{0}$ 对于所有其他锚点集，关于 $\mathcal{C}$ 取平均可得

$$\mathbb{E}_\pi\left[\sum_{0\le i<j<\ell}\ \sum_{\substack{u\in\mathcal{B}_i, v\in\mathcal{B}_j,\\ uv\ne q}} \frac{f_{uv}^2}{w_{uv}^2}\right] \le \frac{6400\,b^4}{\ell}.$$

将这一非负贡献限制在 $\mathcal{E}$ 发生且 q 锚点良好的事件上，只会降低其期望值。如果 $\mathcal{E}$ 失败或 q 不是锚点良好的，我们对

截断杠杆得分。回顾引理 6.2 ，查询边 q 不是锚点良好的概率至多为 $2N/\binom{\ell b+2}{2} \le 16N/n^2 \le 16N/(nt)$。因此，根据并集界和上述界限，我们有

$$\begin{aligned}
&\mathbb{E}_\pi\Big[\min\Big\{1, \mathrm{lev}_{\boldsymbol{a}_q}\Big(\big(\mathbf{A}\,\mathrm{diag}(\boldsymbol{w})\big)_{-q}\Big)\Big\}\Big] \\
&\le \big(1-\Pr\{\mathcal{E}\}\big) + \Pr\{q \text{ is not anchor-good}\} + \frac{1}{\ell b} + \frac{6400\,b^4}{\ell} \\
&\le \frac{8\cdot 10^9 N}{nt} + 10^{-3} + \frac{16N}{nt} + \frac{1}{\ell b} + \frac{6400\,b^4}{\ell} < \frac{1}{100},
\end{aligned}$$

其中，最后一个不等式由 $N \le 10^{-12} nt$、$\ell = 10^7 b^4$ 和 $b \ge 10^8$ 得出。

最后，通过为困难实例图补充孤立顶点，我们可以将下界推广到所有足够大的 n。由此完成定理 1.1 的证明。 □

参数和指数 $6/5$ 的选取。 我们取 $b = \Theta(t)$，如引理 5.2 所示，并取 $\ell = \Theta\big(b^4\big)$为一个足够大的常数，以使跨块贡献保持较小。这些选择给出$n = 2\ell b = \Theta\big(t^5\big)$，因此 $nt = \Theta\big(n^{6/5}\big)$。由于证明排除了至多为 nt 的一个较小常数倍的草图维度，因此我们得到下界 $\Omega\big(n^{6/5}\big)$。

7 致谢

本研究的部分工作是在 Mingji Yang 访问新加坡国立大学期间完成的，并获得了中国国家留学基金委员会 202506360304 号项目的部分资助。

8 人工智能方法论

我们研究结果的证明最初由 GPT-6 Astra 找到，随后通过与该模型进行广泛互动而得到了显著简化。作者已对证明进行了核验和重写，以提升其可读性和表述效果。本文中的图表是在 GPT-6 Astra 的协助下制作的。

参考文献

[ADKKP16] Ittai Abraham、David Durfee、Ioannis Koutis、Sebastian Krinninger 和 Richard Peng。“全动态图稀疏化器”。载于：第57届 IEEE计算机科学基础年会论文集。2016年，第335–344页。doi: 10.1109/FOCS.2016.44（引用于第2页）。[AGM12] Kook Jin Ahn、Sudipto Guha 和 Andrew McGregor。“图草图：稀疏化器、生成器和子图”。载于：第31届 ACM SIGMOD-SIGACT-SIGART 数据库系统原理研讨会论文集。2012年，第5–14页。doi: 10.1145/2213556.2213560（引用于第2页）。[AHLW16] Yuqing Ai、Wei Hu、Yi Li 和 David P. Woodruff。“转闸流中的新刻画及其应用”。载于：第31届计算复杂性会议论文集。第50卷。2016年，20:1–20:22。doi: 10 .4230 /LIPICS . CCC.2016.20（引用于第2、3页）。[BK15] András A. Benczúr 和 David R. Karger。“容量图上的割与流的随机化近似方案”。载于：SIAM 计算杂志 44.2（2015），第290–319页。doi: 10.1137/070705970（引用于第1、2页）。

[BK96] András A. Benczúr 和 David R. Karger。“在 s-t 时间内近似$\tilde{O}(n^2)$最小割”。载于：第28届 ACM 年度计算理论研讨会论文集。1996年，第47–55页。doi: 10.1145/237814.237827（引用于第1页）。[BSS12] Joshua Batson、Daniel A. Spielman 和 Nikhil Srivastava。“双重拉马努金稀疏化器”。载于：SIAM 计算杂志 41.6（2012），第1704–1721页。doi: 10.1137/090772873（引用于第2页）。[CKL22] Yu Chen、Sanjeev Khanna 和 Huan Li。“通过线性草图进行加权图稀疏化”。载于：第63届 IEEE计算机科学基础年会论文集。2022年，第474–485页。doi: 10.1109/FOCS54457.2022. 00052（引用于第2–4、7页）。[CST22] Antares Chen、Jonathan Shi 和 Luca Trevisan。“超越拉马努金界的团的割稀疏化：割稀疏化与谱稀疏化的分离”。载于：2022年 ACM-SIAM 离散算法研讨会论文集。2022年，第3693–3731页。doi: 10.1137/1.9781611977073.146（引用于第3页）。[CSWZ16] Jiecao Chen、He Sun、David P. Woodruff 和 Qin Zhang。“通信最优的分布式聚类”。载于：神经信息处理系统进展 29。2016年，第3720–3728页。url: https://papers.nips.cc/paper_files/paper/2016/hash/7503cfacd12053d309b6bed5c89de212-Abstract.html（引用于第1页）。[CWXZ26] Vincent Cohen-Addad、David P. Woodruff、Shenghao Xie 和 Samson Zhou。“仅插入数据流中近乎空间最优的图与超图稀疏化”。载于：第14届国际学习表征会议论文集。2026年。url: https://proceedings.iclr.cc/paper_files/paper/2026/hash/51d06be858cae042716d876e90ce505f-Abstract-Conference.html（引用于第2页）。[GLWYZ25] Elena Gribelyuk、Honghao Lin、David P. Woodruff、Huacheng Yu 和 Samson Zhou。“线性草图提升：最优界与对抗鲁棒性”。载于：第57届 ACM 年度计算理论研讨会论文集。2025年，第395–406页。doi: 10.1145/3717823.3718227（引用于第3页）。[JLY26] Cheng Jiang、Yinchen Liu 和 Huacheng Yu。“对于多项式长度数据流，转闸流算法可能（仍然）同样是线性草图”。载于：第67届 IEEE计算机科学基础年会论文集。待出版；预印本见 https://arxiv.org/abs/2604.22052。2026年（引用于第2、3页）。[JS25] Arun Jambulapati 和 Aaron Sidford。“超稀疏超稀疏化器与更快的拉普拉斯系统求解器”。载于：ACM 算法汇刊 21.3（2025），31:1–31:49。doi: 10.1145/3593809（引用于第1页）。[KL13] Jonathan A. Kelner 和 Alex Levin。“半流式设定下的谱稀疏化”。载于：计算理论系统 53.2（2013），第243–262页。doi: 10. 1007/S00224-012-9396-1（引用于第2页）。[KLMMS17] Michael Kapralov、Yin Tat Lee、Cameron Musco、Christopher Musco 和 Aaron Sidford。“动态数据流中的单遍谱稀疏化”。载于：SIAM 计算杂志 46.1（2017），第456–477页。doi: 10.1137/141002281（引用于第2、3页）。

[KLOS14] Jonathan A. Kelner、Yin Tat Lee、Lorenzo Orecchia 和 Aaron Sidford。“无向图近似最大流及其多商品推广的近线性时间算法”。载于：第25届ACM-SIAM离散算法研讨会论文集。2014年，第217–226页。doi：10.1137/1.9781611973402.16（引于第1页）。[KMMMNST20] Michael Kapralov、Aida Mousavifar、Cameron Musco、Christopher Musco、Navid Nouri、Aaron Sidford 和 Jakab Tardos。“动态数据流中快速且节省空间的谱稀疏化”。载于：2020年ACM-SIAM离散算法研讨会论文集。2020年，第1814–1833页。doi：10.1137/1.9781611975994.111（引于第2页）。[KP20] John Kallaugher 和 Eric Price。“转闸流与线性草图之间的分离与等价性”。载于：第52届ACM理论计算研讨会论文集。2020年，第1223–1236页。doi：10.1145/3357713.3384278（引于第2、3页）。[KX16] Ioannis Koutis 和 Shen Chen Xu。“谱图稀疏化的简单并行与分布式算法”。载于：ACM并行计算汇刊 3.2（2016年），14:1–14:14。doi：10.1145/2948062（引于第2页）。[LMW26] Honghao Lin、Vahab Mirrokni 和 David P. Woodruff。“转闸流中近乎最优的空间与时间割稀疏化”。载于：CoRR abs/2611.03604（2026年）。doi：10.48550/arXiv.2611.03604（引于第2页）。[LNW14] Yi Li、Huy L. Nguyen 和 David P. Woodruff。“转闸流算法或许同样可以是线性草图”。载于：第46届ACM理论计算研讨会论文集。2014年，第174–183页。doi：10.1145/2591796.2591812（引于第2、3页）。[Pen16] Richard Peng。“在 $O(m\text{polylog}(n))$时间内计算近似无向最大流”。载于：第27届ACM-SIAM离散算法研讨会论文集。2016年，第1862–1867页。doi：10.1137/1.9781611974331.CH130（引于第1页）。[She13] Jonah Sherman。“近线性时间内的近最大流”。载于：第54届IEEE计算机科学基础年会论文集。2013年，第263–269页。doi：10.1109/FOCS.2013.36（引于第1页）。[SS11] Daniel A. Spielman 和 Nikhil Srivastava。“通过有效电阻进行图稀疏化”。载于：SIAM 计算杂志 40.6（2011年），第1913–1926页。doi：10.1137/080734029（引于第2页）。[ST11] Daniel A. Spielman 和 Shang-Hua Teng。“图的谱稀疏化”。载于：SIAM 计算杂志 40.4（2011年），第981–1025页。doi：10.1137/08074489X（引于第1页）。[ST14] Daniel A. Spielman 和 Shang-Hua Teng。“用于预条件和求解对称对角占优线性系统的近线性时间算法”。载于：SIAM矩阵分析与应用杂志 35.3（2014年），第835–885页。doi：10.1137/090771430（引于第1页）。