arXiv:2302.00213v2 [cs.DS] 7 Jul 2023

红蓝集合覆盖与最小单调满足赋值的近似

Eden Chlamt´a˘c* Yury Makarychev† Ali Vakilian‡

摘要

我们为红蓝集合覆盖和电路最小单调满足赋值（MMSA）问题提出了新的近似算法。我们的红蓝集合覆盖算法实现了 $\tilde{O}(m^{1/3})$-近似，优于 Elkin 和 Peleg 提出的 $\tilde{O}(m^{1/2})$-近似（其中 m为集合数量）。对于 MMSA，我们的近似算法$_t$（针对深度为 t的电路）为 $\delta = \frac{1}{3}2^{3-\lceil t/2\rceil}$提供了 $\tilde{O}(N^{1-\delta})$ 近似，其中 N为门和变量的数量。此前，对于满足 $t \geq 4$ 的 MMSA$_t$，尚无非平凡的近似算法。

我们还为这些问题给出了相应的下界：对于红蓝集合覆盖，我们提出了一种几乎保持近似比的 Min k-并集归约；在稠密对随机猜想成立的条件下，该归约给出 $\tilde{\Omega}(m^{1/4-\varepsilon})$ 的困难性。对于 MMSA，我们概要说明了一个证明：经 Sherali–Adams 强化的半定规划松弛具有 $N^{1-\varepsilon}$ 的整数间隙，其中 $\varepsilon \to 0$，而电路深度为 $t \to \infty$。

1 引言

本文研究两个问题：红蓝集合覆盖及其推广电路最小单调满足赋值。红蓝集合覆盖是集合覆盖的一种自然推广，由Carr 等人 [2000]提出。电路最小单调满足赋值是一个与标签覆盖关系更为密切的问题，由Alekhnovich 等人 [2001] 和Goldwasser 和 Motwani [1997]提出。

定义 1.1。在红蓝集合覆盖问题中，给定一个包含$(k+n)$ 个元素的全集 U，该全集被划分为互不相交的红色元素集合（R），其大小为 n，以及蓝色元素集合（B），其大小为 k，即 $U = R \cup B$和 $R \cap B = \emptyset$；另给定一族集合 $\mathcal{S}$：$= \{S_1, \cdots, S_m\}$。目标是找到一个集合子族 $\mathcal{F} \subseteq \mathcal{S}$，使得其中集合的并集 $\mathcal{F}$ 覆盖所有蓝色元素，同时使被覆盖的红色元素数量最少。

除了红蓝集合覆盖问题，我们还考虑部分红蓝集合覆盖问题，其中ˆ ˆ还额外给定一个参数 k，目标是在覆盖至少 k 个蓝色元素的同时，尽量减少被覆盖的红色元素数量。

定义 1.2。深度为 t的电路最小单调满足赋值问题，记作 *MMSA$_t$*，定义如下。给定一个深度为 t 、包含布尔变量x_1、.....、x_n的电路 C 。电路 C包含与门和或门：与根节点距离为偶数的所有门（包括根节点处的输出门）都是与门；所有

*本-古里安大学计算机科学系。作者访问芝加哥丰田技术研究所期间完成了这项工作，并获得该研究所的支持。电子邮件：chlamtac@cs.bgu.ac.il

†芝加哥丰田技术研究所（TTIC）。本研究得到美国国家科学基金会资助，项目编号为 CCF-1955173 和 CCF-1934843。电子邮件：yury@ttic.edu

‡芝加哥丰田技术研究所（TTIC）。电子邮件：vakilian@ttic.edu

与根节点距离为奇数的门都是或门。目标是找到一个满足赋值，使被设置为 *1*（真）的变量 x_i 数量最少。

注意， C 计算的是一个单调函数，并且全为 1 的赋值始终是一个可行解。尽管这些问题的定义大不相同，红蓝集合覆盖与 MMSA$_t$ 却密切相关。具体而言，红蓝集合覆盖等价于 MMSA$_3$。[1] 对应关系如下：变量 $x_1, \ldots, x_n$ 表示红色元素；第三层的与门表示集合 $S_1, \ldots, S_m$；第二层的或门表示蓝色元素。表示集合 S_j 的门与表示蓝色元素的或门以及表示红色元素 S_j的变量 x_i 相连。容易看出，对 $x_1, \ldots, x_n$ 的赋值满足该电路，当且仅当存在一个子族 $\mathcal{F} \subseteq \mathcal{S}$ ，它覆盖所有蓝色元素，且仅覆盖与被赋值为 1 的变量 x_i 相对应的红色元素（但不一定覆盖所有这些元素）。

背景。 红蓝集合覆盖及其变体与组合优化领域中的若干著名问题相关，包括群斯坦纳问题 和有向斯坦纳问题 、最小单调满足赋值以及对称最小标签覆盖。可以说，人们对一般 MMSA$_t$ 问题的兴趣，主要源于它与复杂性和近似困难性的联系。

红蓝集合覆盖在异常检测、信息检索等多种场景中都有应用，尤其应用于析取式学习 [Carr 等人，2000]。在 $\{0,1\}^m$ 上学习析取式是 PAC模型中的基本问题之一。在这个问题中，给定 $\mathcal{D}$ 上的任意分布和 $\{0,1\}^m$ 上的目标函数 $h^*\colon \{0,1\}^m \to \{-1,+1\}$ （该函数表示样本的真实标签），目标是找到关于 $\mathcal{D}$ 和 h^* 的最佳析取式 $f^*\colon \{0,1\}^m \to \{-1,+1\}$ （即， $f^*(x)$ 计算由 x 的一部分坐标构成的析取式）。析取式学习问题可以表述为（部分）红蓝集合覆盖问题的一个实例 [Awasthi 等人，2010]: 我们可以将正例视为蓝色元素（即 $h^*(x) = 1$），将负例视为红色元素（即 $h^*(x) = -1$）。然后，我们为每个坐标 i 构造一个对应的集合 S_i ；若样本 x 的第 i 个坐标等于 1，则集合 S_i 包含该样本。令 $C \subset \{S_1, \cdots, S_m\}$。那么，与 C 对应的析取式 f_C，即 $f_C(x) := \bigvee_{S_i \in C} x_i$，在样本 x 上输出 1，当且仅当在构造出的红蓝集合覆盖实例中，对应于 x 的元素被 C 中的集合覆盖。

正如我们在第 5 节中所指出的，红蓝集合覆盖也与 Min k-Union 问题有关；后者是 Densest k-Subgraph 的推广 [Chlamt´aˇc 等人， 2016]。在 Min k-Union 问题中，给定一个由 m 个集合 $\mathcal{S}$ 组成的集合族，以及目标集合数量 k，目标是从 $\mathcal{S}$ 中选出 k 个集合，使其并集的基数最小。值得注意的是，在一个硬度假设下（该假设将针对 Densest k-Subgraph 的“稠密与随机”猜想推广到超图），Min k-Union 无法被近似到优于 $\tilde{\Omega}(m^{1/4-\varepsilon})$ [Chlamt´aˇc 等人， 2017]。本文通过构造从 Min k-Union 到红蓝集合覆盖的归约，证明了红蓝集合覆盖的近似困难性。

1.1 相关工作

Carr 等人 提出了红蓝集合覆盖问题，并给出了一个 $2\sqrt{m}$-近似算法，用于求解每个集合仅包含一个蓝色元素的情形。随后，Elkin 和 Peleg [2007]证明了在该问题的一般情形下，可以获得 $2\sqrt{m}\log(n+k)$ 近似。这是我们开展研究之前，红蓝集合覆盖已知的最佳上界。此前，对于任何 MMSA，都没有已知的非平凡算法 $t \geq 4$。

[1] 另请注意，MMSA$_2$ 等价于集合覆盖。

在困难性方面，Dinur 和 Safra [2004] 证明了 MMSA$_3$ 很难在 $O(2^{\log^{1-\epsilon} m})$ 因子范围内近似，其中 $\epsilon = 1/\log\log^c m$ 对任意常数 $c < 1/2$均成立，只要 $P \neq NP$。正如 Carr 等人 [2000], 所指出的，这也意味着红蓝集合覆盖存在 $O(2^{\log^{1-\epsilon} m})$ 因子困难性。即使在该问题的特例中，这一困难性结果仍然成立：每个集合仅包含一个蓝色元素和两个红色元素。

最后，Charikar 等人 [2016] 针对一种电路深度 t 不固定的 MMSA 变体给出了下界。基于稠密对随机猜想的一种变体，他们证明，对于每个 $\varepsilon > 0$，该问题都不存在 $O(n^{1/2-\varepsilon})$近似算法，其中 n 是变量的数量；此外，也不存在 $O(N^{1/3-\varepsilon})$近似算法，其中 N 是电路中门和变量的总数。

析取式学习。红蓝集合覆盖的算法返回的析取式在正例上不会出错，即它会覆盖所有“蓝色”元素；而要让这些算法适用于双侧错误的情形也很简单。该问题的双侧错误变体被正式定义为正负部分集合覆盖 [Miettinen, 2008] 作者证明，红蓝集合覆盖的一个 $f(m, n)$-近似算法可推出正负部分集合覆盖的一个 $f(m+n, n)$-近似算法。我们的结果对部分红蓝集合覆盖同样成立，而且，一个部分红蓝集合覆盖的 c-近似算法可用于输出一个 c正负部分集合覆盖的近似解。

Awasthi 等人 [2010] 设计了一种 $O(n^{1/3+\alpha})$-近似算法，适用于任意常数 $\alpha > 0$。尽管 Awasthi 等人提出的算法是一种无知学习析取式的算法（即解的形式并非析取式），但它采用红蓝集合覆盖的近似算法，并以析取式作为输出（即，红蓝集合覆盖算法是适当的学习器）。

几何红蓝集合覆盖。该问题已在几何模型中得到广泛研究。Chan 和 Hu [2015] 研究了如下情形：R 和 B 是 $\mathbb{R}^2$ 中的点集，且 $\mathcal{S}$ 是单位正方形的集合。他们证明，即使在这一受限实例中，该问题仍然是 NP 困难的，并为此问题提出了一种多项式时间近似方案。Madireddy 和 Mudgal [2022] 为该问题的另一种几何变体设计了一种 $O(1)$-近似算法；在该变体中，集合为单位圆盘。该问题也在高维情形下得到研究，其中集合为超平面和轴平行对象 [Ashok 等人，2017，Madireddy 等人，2021，Abidha 和 Ashok， 2022]。

1.2 我们的结果

本文针对红蓝集合覆盖、MMSA$_4$, 和一般 MMSA$_t$提出了新的近似算法。此外，我们还给出了红蓝集合覆盖的一项新的条件近似困难性结果。我们还讨论了在 $t \to \infty$ 时，经 Sherali–Adams 强化的 MMSA$_t$ 基本半定规划松弛的整数间隙。

我们首先介绍红蓝集合覆盖的结果。

定理 1.3. 存在一个 $O(m^{1/3}$ 对数$^{4/3}\, n$对数 $k)$-近似算法，用于求解红蓝集合覆盖问题，其中 m 是集合的数量， n 是红色元素的数量， k 是蓝色元素的数量。

正如我们稍后将展示的，我们的算法也适用于部分红蓝集合覆盖。我们的方法将实例划分为多个子实例，其中每个集合所含红色元素的数量都有界，例如介于 r 和 $2r$ 之间，并且每个红色元素出现于有界数量的集合中。利用这种划分的性质，我们证明总能找到一个小型集合族

，使红色元素与蓝色元素的比例保持适当，从而逐步构造出一个 $\tilde{O}(m^{1/3})$-近似算法。[2] 随后，通过迭代应用该算法，直至覆盖所有蓝色元素，我们得到定理 1.3 的保证。在每次迭代中，我们的分析都保证局部线性规划松弛可行，而通过简单的随机舍入即可得到所需的蓝色顶点与红色顶点之比。

现在我们描述 MMSA 问题的结果。

定理 1.4。 存在一个 $\tilde{O}(N^{1/3})$-近似算法，可用于 $MMSA_4$，，其中 N 是输入实例中门和变量的总数。

我们针对 MMSA_4 的算法受我们针对 MMSA_3，的算法启发，但由于该问题的复杂性，该算法涉及的内容要复杂得多。特别是，似乎不存在一种自然的预处理步骤，能够类比于我们对红蓝集合覆盖所应用的划分，因此我们需要依赖高阶矩线性规划松弛，以及一种谨慎设计的、基于线性规划的划分，并将其构建到算法中。

定理 1.5。 令 $t \geq 4$。定义 $\delta = \frac{1}{3} \cdot 2^{3-\lceil t/2 \rceil}$。存在一个 $\tilde{O}(N^{1-\delta})$-近似算法，可用于 $MMSA_t$，其中 N 是输入实例中门和变量的总数。

我们针对一般 MMSA_t 的算法对深度 t进行递归，并以我们针对红蓝集合覆盖和 MMSA_4 的算法作为递归的基础。每个递归步骤都依赖于一个初始的朴素线性规划松弛；当针对较小深度 MMSA 的算法调用发现新的违反约束时，我们便将这些约束加入其中。

我们还给出了近似困难性结果，与红蓝集合覆盖的上界相辅相成。

定理 1.6。 假设超图稠密对随机猜想成立，对于每个 $\varepsilon > 0$，不存在多项式时间算法能够实现优于 $O(m^{1/4-\varepsilon}/\text{对数}^2\, k)$ 的红蓝集合覆盖近似，其中 m 为集合数量， k 为蓝色元素数量。

为证明其困难性，我们给出从 Min k-Union 到红蓝集合覆盖的归约，该归约将近似比保持在至多相差 polylog(k) 倍的范围内。随后，困难性便由 Min k-Union 的标准猜想困难性 [Chlamt´ a˘c 等人， 2017]的结果推出。在我们的归约中，给定的 Min k-Union 实例中的所有元素都被视为所构造红蓝集合覆盖实例中的红色元素；此外，我们还从一个大小为 k的蓝色元素基集中，以有放回抽样的方式为每个集合补入 O(对数 k) 个元素。我们证明，该归约保持近似比，至多相差 polylog(k) 倍。

组织结构。 第 2 节中，我们重新表述红蓝集合覆盖问题，并介绍一些符号。第 3 节中，我们给出红蓝集合覆盖问题的算法。我们在附录 A 中将该算法调整用于部分红蓝集合覆盖问题。随后，在第 4 节中，我们给出 MMSA_t with$t \geq 5$ 的算法。该算法依赖于 MMSA_4，的算法，我们将在第 6 节中介绍后者。第 5 节中，我们给出从 Min k-Union 到红蓝集合覆盖的归约，由此得到红蓝集合覆盖问题的近似困难性结果。第 7 节中，我们讨论一般 MMSA_t 问题的近似困难性。我们还在该节给出证明概要，说明经 $\tilde{}\Omega(\log\, n)$ 轮 Sherali–Adams 强化的基本半定规划松弛的整数间隙为 $\Omega(N^{1-O((\log t)/t)})$（其中 N 是门和变量的总数）。

[2]在此，我们略微滥用 $\tilde{O}$ 符号来隐藏 n， k的多对数因子。

2 预备知识

为了简化红蓝集合覆盖近似算法的描述和分析，我们用图论术语重新表述该问题。本质上，我们将问题重新表述为 MMSA$_3$。具体而言，我们将一个红蓝集合覆盖实例视为一个三部图 (B, J, R, E)，其中所有边 (E) 都与 J 以及 B 或 R 相连。J 中的顶点代表集合索引，它们在 B（分别在 R）中的邻居代表这些集合中的蓝色（分别为红色）元素。因此，我们的目标是找出 J 中的一个顶点子集，使其成为 B 的支配集，并且在 R 中的邻居总数最少。简言之，我们用 $k = |B|$、$m = |J|$ 和 $n = |R|$ 表示这些不同顶点集的基数。

类似地，我们将一个 MMSA$_4$ 实例视为一个元组（B, J, R, S, E）。这里，B、J和 R 分别表示电路第二、第三和第四层中的门（其中第 i 层由距离根节点 $i-1$ 的门组成）；S 表示变量；E 表示门/变量之间的边。从组合角度看，目标是获得如上所述的 J 的一个子集，同时在 S 中为所选 J的红色邻居（位于 R中）求一个最小支配集。

符号。我们用 $\Gamma(\cdot)$表示顶点的邻域；对于顶点集 U，用$\Gamma(U)$表示 U中顶点邻域的并集，即 $\bigcup_{u\in U}\Gamma(u)$。我们还考虑受限邻域，记作 $\Gamma_T(u) := \Gamma(u) \cap T$或 $\Gamma_T(U) = \Gamma(U) \cap T$。我们将此类集合的基数，即 $|\Gamma_T(u)|$，称为 u的-度数 。

注释 2.1. 请注意，对于每个集合索引 $j \in J$，集合 $\Gamma(j)$ 就是该问题集合系统表述中的集合 S_j，而集合 $\Gamma_R(j)$（分别为 $\Gamma_B(j)$）就是索引为 j 的集合中的红色（分别为蓝色）元素子集。类似地，对于任意 $i \in R \cup B$，$\Gamma_J(i)$ 由索引 j 组成，这些索引表示包含元素 i 的集合 S_j。

对于红蓝集合覆盖算法，我们引入一种自然的进展概念 :

定义 2.2. 如果给定一个最优解包含 最优值 个红ˆ元素的实例后，红蓝集合覆盖算法能够找到一个子集 $J \subseteq J$，使得

$$\frac{|\Gamma_R(\hat{J})|}{|\Gamma_B(\hat{J})|} \leq \tilde{A} \cdot \frac{\mathrm{OPT}}{|B|}.$$

˜显然，如果我们有一个能够朝着 A-近似迈进的算法，那么我们就可以反复运行该算法（逐渐减小 $|B|$ 参数，初始时 $|B| = k$）˜，直到覆盖所有蓝色元素，并获得一个 $O(A \cdot \log k)$-近似解。

为简洁起见，除非另有说明，所有对数均默认以 2 为底，也就是说，我们将 $\log(\cdot) \equiv \log_2(\cdot)$ 记作 log。

3 红蓝集合覆盖近似算法

我们首先排除少量红色元素，并将集合 J 分入少量红度相同的分箱中。对于一个 $O(A)$-近似，目标是排除至多$A \cdot$ 最优值个红色元素（我们可以通过简单的线性搜索或二分搜索猜测最优值）。以下引理对此进行了处理:

引理 3.1。 存在一个多项式时间算法，给定输入 (B, J, R, E) 和参数 n_0，返回至多 $\log n$ 个数对 (J_α, R_α)，并满足以下性质：

- 集合 $J_\alpha \subseteq J$ 将集合 J划分。
- 集合 $R_\alpha \subseteq R$ 构成一个嵌套集合族，其中最小的集合（即它们的交集）的基数至少为 $n - n_0$。也就是说，这些集合中的任意一个最多排除 n_0 个红色元素。
- 对于每个 α，都存在某个 r_α，使得每个集合 $j \in J_\alpha$ 的 R_α度（或受限红色集合大小）

$$|\Gamma_{R_\alpha}(j)| \in [r_\alpha, 2r_\alpha],$$

- 并且对于每个 α，每个红色元素 $i \in R_\alpha$ 的 J_α度至多为（也就是说， J_α 中包含 i 的红色集合数量）

$$|\Gamma_{J_\alpha}(i)| \leq \frac{2mr_\alpha \log n}{n_0}.$$

证明。 考虑以下算法：

- 令 r 为最大红度（即 $\max_{j \in J} \deg_R(j)$）。
- 重复执行以下步骤，直到满足 $J \neq \emptyset$:
  - 从 R 中删除按度数排序最高的 $n_0/\log n$ J 个红色元素及其关联边。– 删除后，令 $J' = \{j \in J \mid \deg_R(j) \in [r/2, r]\}$。– 如果 J' 非空，则将当前数对 (J', R) 加入输出有序对列表（排除迄今为止从 R 中删除的所有元素）。– 从 J 中删除 J' 内的集合（及其关联边），并令 $r \leftarrow r/2$。

根据 r 的递减情况，可以很容易看出，这会将 J 划分为至多 $\log n$ 个集合（更具体地说，是初始最大红色集合大小 $\max_{j \in J} \deg_R(j)$ 的对数）。另请注意，在每次迭代开始时，所有红色集合的大小都至多为 r，因此连接到 R 的边至多有 mr 条，而度数最高的 $n_0/\log n$ J 个红色元素的平均度（特别是最小度）至多为 $mr/(n_0/\log n)$。因此，移除这些红色元素后，所有剩余红色元素的 J-度（特别是 J'-度）至多为所需上界 $2mr_\alpha \log_{n_0} n$，其中 $r_\alpha = r/2$。最后，由于迭代次数至多为 $\log n$，所有迭代中移除的红色元素总数至多为 n_0。

□

我们的算法通过多次迭代运行，在每次迭代中，都会覆盖并从 B 中移除蓝色元素的某个子集。然而，不会从 J 或 R 中移除任何内容。因此，上述引理在整个算法过程中都适用。注意，对于一个最优解 $J_{\mathrm{OPT}} \subseteq J$，在上述划分中的至少一个子集 J_α 中， $J_{\mathrm{OPT}} \cap J_\alpha$ 中的集合必须覆盖 B 的至少 $(1/\log n)$ 倍。因此，要实现 $O(A)$ 近似，只需应用上述引理，令参数为 $n_0 = \mathrm{OPT} \cdot A$，并反复在由 (B, J_α, R_α) 诱导出的某个子图中，朝着一个 A-近似解推进。在最终分析中，通过将注意力限制在这些子实例上，我们最多只需额外付出 $\mathrm{OPT} \cdot A$。

预先固定一个最优解 $J_{\mathrm{OPT}} \subseteq J$。对于上述划分中的任意 α，令 $J^\alpha_{\mathrm{OPT}} = J_\alpha \cap J_{\mathrm{OPT}}$ 表示集合 J_α 中也属于我们最优解的那些集合，并令 $B_\alpha = \Gamma_B(J^\alpha_{\mathrm{OPT}})$ 表示由 J^α最优值 中的集合所覆盖的蓝色元素。请注意，根据 J_{OPT} 的可行性，每个蓝色元素都必须至少属于一个 B_α。在此背景下，我们可以证明：

引理 3.2。 对于引理 *3.1* 所述划分中的任意 α，都存在一个红色元素 $i_0 \in R_\alpha$，使得其最优 J_α-受限邻居 $\Gamma_{J^\alpha_{\mathrm{OPT}}}(i_0)$ 至少覆盖 $|B_\alpha| r_\alpha$/最优值 个蓝色元素。

证明。考虑（B_α， J^α_{OPT}， $E(B_\alpha$， $J^\alpha_{\mathrm{OPT}})$）的如下子图。对于每个蓝色元素 $\ell \in B_\alpha$，保留恰好一条连接到 J^α_{OPT} 的边。令 F 为这组边。

ˆ 因此，蓝色元素 B_α 至少有 $|B_\alpha| r_\alpha$ 条路径经过 $F \times E(J^\alpha_{\mathrm{OPT}}$， $R_\alpha)$，通向 R_α 中 J^α_{OPT} 的红色邻居。由于这样的红色邻居至多有 OPT 个，其中至少有一个，设为 $i_0 \in \Gamma_{R_\alpha}(J^\alpha_{\mathrm{OPT}})$，必定出现在这些路径中至少占 $1/\mathrm{OPT}$ 的比例。也就是说，ˆ 至少有 $|B_\alpha| r_\alpha/\mathrm{OPT}$ 条路径。由于 J^α_{OPT} 中各顶点的 F-邻域互不相交（由构造可知），这些路径终止于不同的蓝色元素，因此，B_α 中至少有 $|B_\alpha| r_\alpha/\mathrm{OPT}$ 个元素。 □

当然，我们无法知道哪个红色元素具有这一性质，但算法可以尝试所有元素，并分别对每个元素运行算法的其余部分。现在，给定一个红色元素$i_0 \in R_\alpha$，我们的算法如下进行：首先求解以下线性规划。

$$\max \sum_{\ell \in B} z_\ell \tag{1}$$

$$\sum_{i \in R_\alpha} y_i \le \mathrm{OPT} \tag{2}$$

$$0 \le z_\ell \le \min\{1, \sum_{j \in \Gamma_{J_\alpha}(i_0) \cap \Gamma_{J_\alpha}(\ell)} x_j\} \qquad \forall \ell \in B \tag{3}$$

$$0 \le x_j \le y_i \le 1 \qquad \forall j \in \Gamma_{J_\alpha}(i_0), i \in \Gamma_{R_\alpha}(j) \tag{4}$$

在预期的解中，x_j 是表示事件 $j \in \Gamma_{J^\alpha_{\mathrm{OPT}}}(i_0)$ 的指示变量， y_i 是表示红色顶点 i 属于由 $\Gamma_{J^\alpha_{\mathrm{OPT}}}(i_0)$ 索引的红色集合并集（因此也属于最优解）的指示变量，而 z_ℓ 是表示蓝色顶点 $\ell \in B$由 $\Gamma_{J^\alpha_{\mathrm{OPT}}}(i_0)$ 中某个集合覆盖的指示变量。这个线性规划始终可行（例如，将所有变量都设为 0），不过，由于划分中至多有 $\log n$ 个子实例，因此至少存在一个 α，使得 $|B_\alpha| \ge |B|/\log n$；根据引理 3.2，还存在某个 $i_0 \in R_\alpha$ 的取值，使目标函数满足

$$\sum_{\ell \in B} z_\ell \ge \frac{|B_\alpha| r_\alpha}{\mathrm{OPT}} \ge \frac{|B| r_\alpha}{\mathrm{OPT} \cdot \log n}. \tag{5}$$

算法将选择使重新缩放后的目标函数 $\sum_{\ell \in B} z_\ell / r_\alpha$ 最大的 α 和 i_0，从而保证这一界。最后，在此阶段，我们执行简单的随机舍入：对每个集合 $j \in \Gamma_{J_\alpha}(i_0)$，独立地以概率 x_j 选取。整个算法见算法 1。

现在令 $J^* \subseteq J_\alpha$ 表示算法在此步骤中添加的集合。我们来分析 J^* 覆盖的蓝色元素数量，以及加入解中的红色元素数量。首先，由于该线性规划对蓝色元素构成最大覆盖松弛，覆盖的蓝色元素期望数量至少为

$$(1 - 1/e) \frac{|B| r_\alpha}{\mathrm{OPT} \cdot \log n},$$

根据标准分析和界限（5）。

算法 1 蓝红集合覆盖近似算法

输入： B，J，R，E

猜测 最优值 ▷ 例如使用二分搜索

$J_{\mathrm{ALG}} = \emptyset$ ▷ J_{ALG} 存储当前解

查找分解 $\{(J_\alpha,\ R_\alpha)\}_\alpha$，如引理 3.1 所述，参数为 $n_0 =$最优值$\cdot m^{1/3}$对数$^{4/3}\, n$对数$^2\, k$.

当 $B \neq \emptyset$ **时执行**

对于所有 α 和 $i_0 \in R_\alpha$，**执行**

求解线性规划 (1)-(4)。令 LP(α，i_0) 为其目标值

结束循环

选择使 LP(α，i_0)$/r_\alpha$ 最大的 α 和 i_0 的值

令 x，y，z 为该线性规划的最优解

使用解 x 和条件期望法求得 $J^* \subseteq J_\alpha$

使得 $|\Gamma_{R_\alpha}(J^*)|/|\Gamma_B(J^*)| \le O(1) \cdot m^{1/3}\log^{4/3} n \cdot$最优值$/|B|$ ▷ 详情请参见证明

令 $J_{\mathrm{ALG}} = J_{\mathrm{ALG}} \cup J^*$

令 $B = B \setminus \Gamma_B\ (J^*)$

移除 E中与已删除顶点关联的边

结束 while 循环

返回 J_{ALG}

现在来界定所添加的红色元素的数量。令

$$J_+ = \left\{ j \in \Gamma_{J_\alpha}(i_0) \;\middle|\; x_j \ge \frac{\mathrm{OPT}}{r_\alpha \hat{A}} \right\}$$

ˆ取一个稍后确定的值 A。根据约束条件（4），每个红色邻居 $i \in \Gamma_{R_\alpha}(J_+)$ 也将满足 ˆ ˆ $y_i \ge$最优值$/(r_\alpha A)$，因此根据约束条件（2），此类邻居至多有 $r_\alpha A$ 个。另一方面，剩余集合 $j \in J^* \setminus J_+$ 所添加的红色元素的期望数量有如下上界：

$$\begin{aligned}
\mathbb{E}\left[\left|\bigcup_{j\in J^*\setminus J_+} \Gamma_{R_\alpha}(j)\right|\right] &\le 2r_\alpha \mathbb{E}\left[|J^* \setminus J_+|\right] && \text{by } R_\alpha\text{-degree bounds for } j \in J_\alpha \\
&= 2r_\alpha \cdot \sum_{j\in\Gamma_{J_\alpha}(i_0)\setminus J_+} x_j \\
&\le 2r_\alpha \cdot \frac{\mathrm{OPT}}{r_\alpha \hat{A}} \left|\Gamma_{J_\alpha}(i_0) \setminus J_+\right| && \text{by definition of } J_+ \\
&\le \frac{2\mathrm{OPT}}{\hat{A}} \cdot \frac{2mr_\alpha \log n}{\mathrm{OPT}\cdot \hat{A}} && \text{by } J_\alpha\text{-degree bounds for } i \in R_\alpha \\
&= \frac{4mr_\alpha \log n}{\hat{A}^2}
\end{aligned}$$

当 r_αˆ$A = 4mr_\alpha$log $n/$ˆA^2 时，这两个界相等；也就是说，当 ˆ$A =$(4$m$log $n)^{1/3}$ 时，这一步添加的红色元素的期望数量总上界为

$$\mathbb{E}[|\Gamma_{R_\alpha}(J^*)|] \le 2r_\alpha(4m\log n)^{1/3} \le \frac{2r_\alpha(4m\log n)^{1/3}}{(1-1/e)|B|r_\alpha/(\mathrm{OPT}\cdot\log n)} \cdot \mathbb{E}[|\Gamma_B(J^*)|].$$

因此，

$$\mathbb{E}[|\Gamma_{R_\alpha}(J^*)|] - \frac{2(4m\log^4 n)^{1/3}\mathrm{OPT}}{(1-1/e)|B|} \cdot \mathbb{E}|\Gamma_B(J^*)| \le 0.$$

使用条件期望法，我们可以对算法进行去随机化，并找到一个蓝色邻居集非空的 J^*，使得

$$\frac{|\Gamma_{R_\alpha}(J^*)|}{|\Gamma_B(J^*)|} \le O(1)\cdot m^{1/3}\log^{4/3} n \cdot \frac{\mathrm{OPT}}{|B|}.$$

~因此，我们按照定义 2.2 推进，朝着为 $\tilde{A} = O\left(m^{1/3}\log^{4/3} n\right)$ 实现 $A\cdot k$ 的近似保证迈进。正如前文所述，这最终也会为红蓝集合覆盖带来相同的近似保证，从而证明定理 1.3。

4 近似 $\mathbf{MMSA}_t$ （$t \ge 5$）

现在，我们转而研究如何为任意大的（但固定的）t 近似求解一般的 MMSA_t 问题。我们将基于针对 MMSA_4 的近似算法（见第 6 节），递归调用针对更小 t 值的该问题的近似算法，并在某些情况下将此近似结果用作分离预言机。

我们将输入的总规模记为 N，并将 MMSA_t 的近似因子记为 A_t。我们只介绍偶数深度的算法。奇数深度的算法略微简单一些，但与之相当类似；然而，我们能够为 MMSA_{2t-1} 实现的保证与 MMSA_{2t} 的几乎相同（相差一个 $O(\log N)$ 因子）。由于 MMSA_{2t-1} 本质上是 MMSA_{2t} 的一个特例，我们只关注偶数层。

引理 4.1. 对于 $t \ge 2$，如果 MMSA_{2t} 可以在因子 A_{2t} 范围内近似，那么我们就可以在 $O(\sqrt{N}\cdot A_{2t}\log N)$ 范围内近似 MMSA_{2t+2}（因此也可以近似 MMSA_{2t+1}）。

证明。 将输入记为一个具有顶点层 $V_1,\dots,V_{2t+2}$ 的分层图。理想情况下，我们希望丢弃任何满足以下条件的顶点 $j \in V_{2t}$：覆盖其邻居 $\Gamma_{2t+1}(j)$ 需要的顶点数超过 V_{2t+2} 中的最优值；然而，要精确判断这一点就需要求解集合覆盖。我们改为丢弃这样的顶点 $j \in V_{2t}$：在 V_{2t+2} 中覆盖其邻居 $\Gamma_{V_{2t+1}}(j)$ 所需的最小分数集合覆盖值大于最优值。若包含这些顶点，代价就会超过最优值，因此我们知道它们不会出现在任何最优解中。我们从以下基本线性规划开始：

$$\sum_{h\in V_{2t+2}} w_h \le \mathrm{OPT} \tag{6}$$

$$y_i \le \sum_{h\in\Gamma_{2t+2}(i)} w_h \qquad \forall i \in V_{2t+1} \tag{7}$$

$$x_j \le y_i \qquad \forall j \in V_{2t}, i \in \Gamma_{V_{2t+1}}(j) \tag{8}$$

$$x_j, y_i, w_h \in [0,1] \qquad \forall j \in V_{2t}\forall i \in V_{2t+1}\forall h \in V_{2t+2} \tag{9}$$

请注意，按目前的表述，这个线性规划是平凡的。事实上，在没有任何附加约束的情况下，全零解是可行的。然而，随着算法的进行，我们会添加新的违反约束。

给定上述线性规划的一个解，我们的 MMSA_{2t+2} 算法如下：

- 令 $V_{2t}^+ = \{j \in V_{2t} \mid x_j \ge 2(1+\ln N)/A_{2t+2}\}$。将这些顶点加入解中。

[3] That is, $\min\left\{\sum_{h\in S} z_h \;\middle|\; \forall i \in \Gamma_{V_{2t+1}}(j) : \sum_{h\in\Gamma_{V_{2t+2}}(i)} z_h \ge 1;\ \forall h \in S : z_h \ge 0\right\}$.

- 令 $V_{2t+2}^{+}=\Gamma_{V_{2t+2}}(\Gamma_{V_{2t+1}}(V_{2t}^{+}))$ 为 V_{2t}^{+} 的邻居的邻居。
- 对集合覆盖应用贪心 $(1+\ln\ N)$-近似算法，求得一个针对$\Gamma_{V_{2t+1}}(V_{2t}^{+})$的集合覆盖（位于 V_{2t+2} 中），并将此集合覆盖也加入解中。
- 通过移除层 V_{2t+1}, V_{2t+2}、 V_{2t}^{+}中的所有顶点及其在 V_{2t-1} 中的邻居（即 $\Gamma_{V_{2t-1}}(V_{2t}^{+})$）来构造一个 MMSA_{2t} 实例，因为这些顶点已由 V_{2t}^{+} 覆盖。
- 对该实例应用 MMSA_{2t}的 A_{2t}-近似算法，并令 $U_{\mathrm{ALG}}\subseteq V_{2t}\setminus V_{2t}^{+}$为所得结果（或至少是属于第 $2t$ 层的部分）。
- 若$|U_{\mathrm{ALG}}|\leq A_{2t+2}/(2+2\ln\ N)$，则将 U_{ALG} 中的顶点加入解中，并为邻域 $\Gamma_{V_{2t+1}}(U_{\mathrm{ALG}})$ 求一个贪心集合覆盖（位于 V_{2t+2} 中）。
- 否则（如果$|U_{\mathrm{ALG}}|>A_{2t+2}/(2+2\ln\ N)$），则使用新的违反约束继续执行椭球算法

$$\sum_{j\in V_{2t}\setminus V_{2t}^{+}}x_j\geq\left\lfloor\frac{A_{2t+2}}{2(1+\ln N)A_{2t}}\right\rfloor+1,\tag{10}$$

并使用新的线性规划解重新启动算法（丢弃之前的解）。

现在分析该算法。根据（8），我们知道 V_{2t}^{+}的所有邻居 $i\in V_{2t+1}$ 的线性规划值均为 $y_i\geq 2(1+\ln\ N)/A_{2t+2}$。因此，根据（7），如果我们为每个顶点 $h\in V_{2t+2}$定义$w_h^{+}=w_h\cdot A_{2t+2}/(2+2\ln\ N)$，那么这就是 V_{2t+1}-邻域$\Gamma_{V_{2t+1}}(V_{2t}^{+})$的一个分数集合覆盖，并且根据（6），其总分数值至多为 $\mathrm{OPT}\cdot A_{2t+2}/(2+2\ln\ N)$。因此，贪心集合覆盖 $(1+\ln\ N)$-近似算法将使用至多 $\mathrm{OPT}\cdot A_{2t+2}/2$ 个顶点覆盖该邻域。完成此步骤后，我们最多可以向 V_{2t+2} 中添加 $\mathrm{OPT}\cdot A_{2t+2}/2$ 个额外顶点到我们的解中，从而获得一个 A_{2t+2}-近似。

现在，假设我们的 MMSA_{2t} 近似算法返回一个基数为 $|U_{\mathrm{ALG}}|\leq A_{2t+2}/(2+2\ln\ N)$ 的集合 U_{ALG}。显然，将 U_{ALG} 的顶点以及其邻域 $\Gamma_{V_{2t+1}}(U_{\mathrm{ALG}})$ 的一个 V_{2t+2}-集合覆盖加入我们的解中，就能得到该 MMSA_{2t+2} 实例的一个可行解。此外，由于经过预处理后，每个 $j\in U_{\mathrm{ALG}}$ 的邻域 $\Gamma_{V_{2t+1}}(j)$ 在 V_{2t+2} 中都有一个值至多为最优值的分数集合覆盖，因此，这些邻域的并集，即 $\Gamma_{V_{2t+1}}(U_{\mathrm{ALG}})$，在 V_{2t+2} 中也有一个值至多为最优值 $\mathrm{OPT}\cdot|U_{\mathrm{ALG}}|\leq\mathrm{OPT}\cdot A_{2t+2}/(2+2\ln\ N)$ 的分数集合覆盖。因此，对邻域 $\Gamma_{V_{2t+1}}(U_{\mathrm{ALG}})$ 应用贪心集合覆盖算法，至多会为我们的解在 V_{2t+2} 中额外增加 $\mathrm{OPT}\cdot A_{2t+2}/2$ 个顶点，符合要求。

最后，让我们考察最后一步（分离预言机）的有效性。如果针对 MMSA_{2t} 的 A_{2t}-近似算法未能找到大小至多为 $A_{2t+2}/(2+2\ln\ N)$ 的解，那么根据定义，我们的 MMSA_{2t} 实例中任何解的值都必须大于 $A_{2t+2}/((2+2\ln\ N)A_{2t})$。这是原始实例的一个子实例，因此，原始 MMSA_{2t+2} 实例中的任何解在 V_{2t} 中也必须包含多于 $A_{2t+2}/((2+2\ln\ N)A_{2t})$ 个顶点。因此，对于任何最优解，约束（10）都是有效的。但它何时会被违反?

根据 V_{2t}^{+} 的定义， $V_{2t}\setminus V_{2t}^{+}$ 当前的线性规划总值至多为 $2(1+\ln\ N)N/A_{2t+2}$。因此，当满足以下条件时，当前线性规划解违反（10）：

$$\frac{2(1+\ln N)N}{A_{2t+2}}\leq\frac{A_{2t+2}}{2(1+\ln N)A_{2t}}\qquad\Longleftrightarrow\qquad A_{2t+2}\geq 2(1+\ln N)\sqrt{N\cdot A_{2t}}.$$

因此，我们可以获得了 $A_{2t+2}=O(\sqrt{N\cdot A_{2t}}\log\ N)$ 的近似保证，正如所声称的 d. □

因此，以 t 为归纳变量，以 MMSA$_4$ 的定理 1.4 的保证作为归纳基础，并以引理 4.1 作为归纳步骤，我们得到一个适用于 MMSA$_t$ 的通用近似算法，其近似比为

$$O\left(N^{1-\frac{1}{3}2^{3-\lceil t/2\rceil}}\cdot(\log N)^{2+O(2^{-t/2})}\right).$$

5 从 Min k-Union 归约到红蓝集合覆盖

本节首先介绍从 Min k-Union 到红蓝集合覆盖的归约，然后证明红蓝集合覆盖的一个困难性结果。我们先正式定义 Min k-Union 问题。

定义 5.1 (Min k-Union)。 在 ***Min k-Union*** 问题中，给定一个大小为 n 的集合 X 、一个由 m 个集合 S_1、 ..... 、 S_m 构成的集合族 $\mathcal{S}$，以及一个整数参数 $k \geq 1$。目标是选择 k 个集合 S_{i_1}、.....、 S_{i_k}，使代价 $\left|\bigcup_{t=1}^{k} S_{i_t}\right|$ 最小。我们将最优解的代价记为 最优值$_{MU}(X, \mathcal{S}, k)$。

注意，Min k-Union 与红蓝集合覆盖问题相似：在这两个问题中，目标都是从给定的集合族 $\mathcal{S}$ 中选取一些子集 S_{i_1}，.....， S_{i_r}，使这些子集的并集中的元素数或红色元素数最小。然而，红蓝集合覆盖和 Min k-Union 对所选子集 S_{i_1}，.....， S_{i_r} 的可行性要求有所不同；前者要求所选集合覆盖所有 k 个蓝色点，而后者只要求所选集合的数量为 k。尽管存在这一差异，我们仍将证明，从 Min k-Union 到红蓝集合覆盖存在一个简单的归约。

命题 5.2. 存在一种从 ***Min k-Union*** 到红蓝集合覆盖的随机化多项式时间归约：给定 ***Min k-Union*** 的一个实例 $\mathcal{I} = (X, \mathcal{S}, k)$，该归约会返回红蓝集合覆盖的一个实例 $\mathcal{I}' = (R, B, \{S'_i\}_{i\in[m]})$，并满足以下两个性质：

1. 若 $S'_{j_1}, \ldots, S'_{j_r}$ 是 $\mathcal{I}'$ 的一个可行解，则 $k' \leq r$，且 ***Min k'-Union*** 的解 $S_{j_1}, \ldots, S_{j_{k'}}$ 的代价（其中 $k' = \lfloor k/\ell \rfloor$ 且 $\ell = \lceil \log_e k \rceil + 1$）不超过红蓝集合覆盖的解 $S'_{j_1}, \ldots, S'_{j_r}$ 的代价：

$$\mathrm{cost}_{MU}(S_{j_1}, \ldots, S_{j_{k'}}) \equiv \left|\bigcup_{t=1}^{k'} S_{i_t}\right| \leq \left|\bigcup_{t=1}^{r} (S'_{i_t} \cap R)\right| = \mathrm{cost}_{RB}(S'_{j_1}, \ldots, S'_{j_r}).$$

无论归约进行何种随机选择，这一点始终成立。

2. 以至少 $1 - 1/e$ 的概率，最优值$_{RB}(R, B, \{S'_i\}_i) \leq$ 最优值$_{MU}(X, \mathcal{S}, k)$。

证明。 我们定义如下实例 $\mathcal{I}'$。令 $R = X$ 和 $B = [k]$。对于每个 $i \in [m]$，令 $R_i = S_i$； B_i是从 $[k]$ 中有放回地随机抽取的 ℓ 个元素组成的集合，并且 $S'_i = R_i \cup B_i$。所有随机选择均相互独立。现在我们验证该归约满足所需的性质。

考虑 $\mathcal{I}'$的一个可行解 $S'_{j_1}, \ldots, S'_{j_r}$。由于该解是可行的， $\cup_{t=1}^{r} B_t = B$。现在$|B_t| \leq \ell$，因此 $r \geq |B|/\ell = k/\ell \geq k'$，符合要求。此外，

$$\mathrm{cost}_{MU}(S_{j_1}, \ldots, S_{j_{k'}}) \equiv \left|\bigcup_{t=1}^{k'} S_{j_t}\right| = \left|\bigcup_{t=1}^{k'} R_{j_t}\right| \leq \left|\bigcup_{t=1}^{r} R_{j_t}\right| \equiv \mathrm{cost}_{RB}(S'_{j_1}, \ldots, S'_{j_r}).$$

我们已经验证了第 1 项成立。现在，令 $S_{i_1}, \ldots, S_{i_k}$为 $\mathcal{I}$的最优解。我们声称， $S'_{i_1}, \ldots, S'_{i_k}$ 以至少 $1 - 1/e$的概率是 $\mathcal{I}'$的可行解。为验证这一声称，

我们需要给 $B_{i_1} \cup \cdots \cup B_{i_k} = B$发生的概率一个下界。事实上，集合 $B_{i_1} \cup \cdots \cup B_{i_k}$由从 B 中有放回抽取的 $k\ell$ 个元素组成。给定元素 $b \in B$ 不属于 $B_{i_1} \cup \cdots \cup B_{i_k}$的概率至多为 $(1 - 1/k)^{k\ell} \leq e^{-\ell} \leq \frac{1}{ek}$。根据并集界，存在某个 $b \in B \setminus (B_{i_1} \cup \cdots \cup B_{i_k})$ 的概率至多为 $k \times \frac{1}{ek} = \frac{1}{e}$。因此，以至少 $1 - 1/e$的概率，不存在这样的 b，从而 $B_{i_1} \cup \cdots \cup B_{i_k} = B$。在这种情况下，红蓝集合覆盖的解 $S'_{i_1}, \ldots, S'_{i_k}$的代价等于 $\left|\bigcup_{i=1}^{k} R'_{i_t}\right| = \left|\bigcup_{i=1}^{k} S_{i_t}\right|$，即 Min k-并集的最优解的代价。

□

推论 5.3. 假设存在一个用于红蓝集合覆盖的 $\alpha(m, n)$ 近似算法 $\mathcal{A}$（其中 α 是 m 和 n 的非递减函数）。那么，存在一个用于 Min k-Union 的随机多项式时间算法 $\mathcal{B}$，可以找到 k' 个集合 $S_{i_1}, \ldots, S_{i_{k'}}$，使得

$$\left|\bigcup_{t=1}^{k'} S_{i_t}\right| \leq \alpha(m, n)\mathrm{OPT}_{MU}(X, \mathcal{S}, k).$$

失败概率至多为 $1/n$。

证明。 我们只需将归约应用于 Min k-Union 的输入实例，然后使用算法 $\mathcal{A}$ 求解得到的红蓝集合覆盖实例。为确保失败概率至多为 $1/n$，我们重复此过程 $\lceil \log_e n \rceil$ 次，并输出找到的最佳解。

□

定理 5.4. 假设存在一个用于红蓝集合覆盖的 $\alpha(m, n)$ 近似算法 $\mathcal{A}$ （其中 α 是 m 和 n 的非递减函数）。那么，存在一个用于 ***Min*** k***-Union*** 的 $O(\log^2 k)\alpha(m, n)$近似算法。

证明。 我们的算法迭代地使用推论中的算法 $\mathcal{B}$ 来寻找近似解。首先，它在输入实例上运行 $\mathcal{B}$，得到 $k_1 = k'$ 个集合。然后，它从实例中移除找到的集合，并将参数 k 缩减为 $k - k_1$。接着，算法在所得实例上运行 $\mathcal{B}$，得到 k_2 个集合。它再次移除所得集合，并将 k 缩减为 $k - k_1 - k_2$（此处 k 是 k的原始值）。它反复执行这些步骤，直到总共找到 k 个集合。也就是说，$k_1 + \cdots + k_T = k$，其中 T是算法执行的迭代次数。

注意，在此过程中构造的每个 Min k-Union 实例的代价至多为 $\mathrm{OPT}_{MU}(X, \mathcal{S}, k)$。事实上，考虑我们在第 $t + 1$次迭代中求解的子实例 $\mathcal{I}_{t+1}$。考虑构成（$X, \mathcal{S}, k$）的最优解的 k个集合。其中至多有 $k_1 + \cdots + k_t$ 个已从 $\mathcal{I}_{t+1}$ 中移除，因此至少有 $k - k_1 - \cdots - k_t$ 个仍存在于 $\mathcal{I}_{t+1}$中。我们从中任意选择$k - k_1 - \cdots - k_t$ 个集合。所选集合构成 $\mathcal{I}_{t+1}$的一个可行解，其代价至多为 $\mathrm{OPT}_{MU}(X, \mathcal{S}, k)$。

因此，我们在每次迭代中找到的部分解的代价至多为 $\alpha(m, n)$·最优值$_{MU}(X, \mathcal{S}, k)$。总代价至多为 $\alpha(m, n) \cdot T \cdot$最优值$_{MU}(X, \mathcal{S}, k)$。还需证明 $T \leq O(\log^2 k)$。我们注意到，k 的值在每次迭代中至少缩小至原来的 $1 - 1/\ell$，因此经过 t次迭代后，它至多为 $(1 - 1/\ell)^t k$。由此可知，迭代总次数 T至多为$O(\ell \log k) = O(\log^2 k)$，如所需。

□

现在，我们根据超图稠密对随机猜想的一个推论，得到红蓝集合覆盖的一个条件性困难性结果。

推论 5.5 (Chlamt´aˇc 等人 [2017])。 假设超图稠密对随机猜想成立，对于任意 $\varepsilon > 0$，不存在多项式时间算法能够对最小 k-并集实现优于 $\Omega(m^{1/4-\varepsilon})$ 的近似。

定理 1.6 立即得出。

6 MMSA 的近似算法$_4$

考虑一个 MMSA$_4$ 的实例 $(B,\ J,\ R,\ S,\ E)$。和红蓝集合覆盖一样，我们将着眼于逐步改进，以获得良好的近似。

定义 6.1。 我们称 *MMSA*$_4$ 的算法朝着 $O(A)$-近似取得进展，如果对于一个最优解至多包含 最优值 个顶点的实例 S，该算法ˆ ˆ ˆˆ能找到一个子集 $J \subseteq J$ 和一个子集 $S \subseteq S$，使得 $\Gamma_R(J) \subseteq \Gamma_R(S)$ （构成一个有效的部分解），并且

$$\frac{|\hat{S}|}{|\Gamma_B(\hat{J})|} \leq A \cdot \frac{\mathrm{OPT}}{|B|}.$$

和之前一样，很容易看出，给定这样一个算法，我们可以反复运行该算法˜，从而得到 MMSA$_4$ 的一个实际 $O(A)$ 近似。事实上，在本节余下部分，我们将只 讨论一种朝着 $O(A)$-近似取得进展的算法。

为了构造具有高度均匀性的线性规划松弛，我们实际上将关注一种部分解，它以均匀方式覆盖大量蓝色元素:

引理 6.2。 对于蓝色元素 B的任意覆盖 $J_0 \subseteq J$，存在子集 $J' \subseteq J_0$ 和 $B' \subseteq B$以及参数 $\Delta > 0$，满足以下性质:

- 每个顶点 $j \in J'$ 的 B'-度数都在范围 $\deg_{B'}(j) \in [\Delta, 2\Delta]$内。
- 每个蓝色元素 $\ell \in B'$ 在 J'_Δ 中至少有一个邻居，且至多有 $2e\ln(2k)$ 个邻居。
- 我们有基数界 $|B'| \geq |B|/(\log k \log m)$。

证明。 按蓝色元素的 J_0-度数划分后，存在某个 D，使得集合 $B_D = \{\ell \in B \mid \deg_{J_0}(\ell) \in [D, 2D]\}$ 的基数为 $|B_D| \geq |B|/\log m$。如果以 $\ln(2k)/D$ 的概率独立地对每个 $j \in J_0$进行抽样，我们会得到某个子集 $J' \subseteq J_0$，并且以正概率，每个 $\ell \in B_D$ 仍会被 J'覆盖，且没有 $\ell \in B_D$ 会在 J'中拥有超过$2e\ln(2k)$个邻居。最后，如果我们按顶点在 J'中的 B_D-度数对其进行分组，那么某个分组 $J'_\Delta = \{j \in J' \mid \deg_{B_D}(j) \in [\Delta, 2\Delta]\}$会覆盖 $1/\log k$比例的元素；将这些元素记为 B'，则每个元素 $\ell \in B'$ 在 J'_Δ中仍会拥有（至少一个且）至多 $2e\ln(2k)$个邻居。

□

简化假设。 我们可以作出以下假设，这些假设对分析我们的算法很有用。首先，我们可以假设，对于每个 $j \in J$，其红邻域 $\Gamma_R(j)$ 在 S 中存在一个权重至多为 OPT 的分数集合覆盖。也就是说，使用 S 覆盖 $\Gamma_R(j)$ 的标准线性规划松弛的最优值至多为 OPT。如果我们猜对了 OPT 的值，那么我们就知道，任何红邻域无法以代价 OPT 覆盖的 $j \in J$ 都不可能属于最优解，因此可以将其丢弃。我们还可以假设，对于某个 $\varepsilon > 0$，上面的值 Δ 至多为 k/m^ε。原因是，否则，蓝色元素˜B' 可以用 J 中至多 $O(m^\varepsilon)$ 个顶点覆盖；并且我们知道，对于其中每一个顶点，其红色邻域˜都可以用 S 中大小为 $O(\mathrm{OPT})$ 的集合覆盖，因此我们可以朝着˜一个 $O(m^\varepsilon)$ 近似取得进展。

猜测上面的 $\Delta \in [k]$ 值以及最优值 OPT 后，我们可以写出以下线性规划松弛：

$$\sum_{h\in S} w_h \le \mathrm{OPT} \tag{11}$$

$$\sum_{\ell\in B} z_\ell \ge |B|/(\log k \log m) \tag{12}$$

$$z_\ell \le \sum_{j\in\Gamma_J(\ell)} x_j^\ell \le 2e\ln(2k) z_\ell \qquad \forall \ell \in B \tag{13}$$

$$\Delta x_j \le \sum_{\ell\in\Gamma_B(j)} x_j^\ell \le 2\Delta x_j \qquad \forall j \in J \tag{14}$$

$$0 \le x_j^\ell \le x_j, z_\ell \le 1 \qquad \forall \ell \in B \forall j \in J \tag{15}$$

$$\sum_{h\in\Gamma_S(i)} w_h \ge y_i \qquad \forall i \in R \tag{16}$$

$$0 \le x_j \le y_i \qquad \forall (j,i) \in E(J,R) \tag{17}$$

我们进一步通过部分提升上述约束来加强这个线性规划。具体来说，对于每个$a \in J \cup S$、 $j \in J$、 $h \in S$、 $i \in R$和 $\ell \in B$，我们引入变量 $X_h^{(a)}, X_\ell^{(a)}, X_\ell^{(a),j}, X_j^{(a)}, X_i^{(a)}$，并相应地提升上述所有约束。精确定义见附录 B。对于满足 $x_j > 0$ 的任意$j \in J$，或满足 $w_h > 0$的任意 $h \in S$，我们将“条件化”变量记作 $\hat{}w_h^{(j)} = X_h^{(j)}/x_j$、$\hat{}z_\ell^{(h)} = X_\ell^{(h)}/w_h$ 等。

注释 6.3。 上述线性规划是引理 6.2 所给部分解的一个松弛。具体而言，给定一个最优解（J_{OPT}, S_{OPT}），将引理应用于 $J_0 = J_{\mathrm{OPT}}$，可得到以下可行解：按照引理，将变量 z_ℓ 和x_j 分别设为 B' 和 J' 的指示变量，并将变量 x_j^ℓ设为 $J' \times B'$的指示变量。将变量 w_h设为 S_{OPT}的指示变量，并将变量 y_i设为 J'的红色邻居 Γ_R（J'）的指示变量。

让我们考察这个松弛的一些有用性质。首先，我们注意到它近似确定了 J 的总线性规划值 Θ（因为线性规划为 B 分配的总线性规划值为 $|B|$）：

命题 6.4。 任何满足约束 (13)-(15) 的解，其在 J 中的总 *LP* 权重有如下界：

$$\frac{1}{2\Delta}\sum_{\ell\in B} z_\ell \le \sum_{j\in J} x_j \le \frac{2e\ln(2k)}{\Delta}\sum_{\ell\in B} z_\ell.$$

证明。

$$\begin{aligned}
\frac{1}{2\Delta}\sum_{\ell\in B} z_\ell \le \frac{1}{2\Delta}\sum_{\ell\in B}\sum_{j\in\Gamma_J(\ell)} x_j^\ell &= \frac{1}{2\Delta}\sum_{j\in J}\sum_{\ell\in\Gamma_B(j)} x_j^\ell && \text{by (13)}\\
&\le \sum_{j\in J} x_j && \text{by (14)}\\
&\le \frac{1}{\Delta}\sum_{j\in J}\sum_{\ell\in\Gamma_B(j)} x_j^\ell && \text{by (14)}\\
&= \frac{1}{\Delta}\sum_{\ell\in B}\sum_{j\in\Gamma_J(\ell)} x_j^\ell \le \frac{2e\ln(2k)}{\Delta}\sum_{\ell\in B} z_\ell. && \text{by (13)}
\end{aligned}$$

□

这些约束还确定了一个有用的组合性质：在任意可行解中，子集 J 所拥有的蓝色邻居数量至少与该集合的线性规划值成正比。

命题 6.5。对于任何满足约束 (13)-(15)的解，以及任意顶点子集 $J \subseteq J$，J 的蓝色邻居数量有如下下界：

$$|\Gamma_B(\hat{J})| \geq \frac{1}{4e\ln(2k)\log k \log m} \cdot \frac{x(\hat{J})}{x(J)} \cdot |B|.$$

证明。

$$\begin{aligned}
|\Gamma_B(\hat{J})| &\geq \sum_{\ell \in \Gamma_B(\hat{J})} z_\ell \\
&\geq \frac{1}{2e\ln(2k)} \sum_{\ell \in \Gamma_B(\hat{J})} \sum_{j \in \Gamma(\ell)} x_j^\ell && \text{by (13)} \\
&\geq \frac{1}{2e\ln(2k)} \sum_{\ell \in \Gamma_B(\hat{J})} \sum_{j \in \Gamma_{\hat{J}}(\ell)} x_j^\ell && \text{由于 } \hat{J} \subseteq J \\
&= \frac{1}{2e\ln(2k)} \sum_{j \in \hat{J}} \sum_{\ell \in \Gamma_B(j)} x_j^\ell \\
&\geq \frac{\Delta}{2e\ln(2k)} \sum_{j \in \hat{J}} x_j && \text{by (14)} \\
&\geq \frac{1}{2e\ln(2k)} \cdot \frac{z(B)}{2x(J)} \cdot x(\hat{J}_0) && \Delta \geq \frac{z(B)}{2x(J)} \text{ by Claim 6.4} \\
&= \frac{1}{4e\ln(2k)} \cdot \frac{x(\hat{J}_0)}{x(J)} \cdot z(B) \\
&\geq \frac{1}{4e\ln(2k)\log k \log m} \cdot \frac{x(\hat{J})}{x(J)} \cdot |B| && \text{by (12)}
\end{aligned}$$

□

一个分数 蓝色顶点的上述覆盖性质的一个对偶变体如下：

断言 6.6. 对于满足约束 (13)-(15)的任意解，以及任意顶点子集 $J \subseteq J$，至少有 $\varepsilon|B|$ 个顶点 $\ell \in B$ 满足

$$\sum_{j \in \Gamma_{\hat{J}}(\ell)} x_j^\ell \geq \frac{1}{4\log k \log m} \cdot \frac{x(\hat{J})}{x(J)},$$

其中

$$\varepsilon = \frac{1}{8e\ln(2k)\log k \log m} \cdot \frac{x(\hat{J})}{x(J)}.$$

证明。设

$$B_0 = \left\{ \ell \in B \;\middle|\; \sum_{j \in \Gamma_{\hat{J}}(\ell)} x_j^\ell \geq \frac{1}{4\log k \log m} \cdot \frac{x(\hat{J})}{x(J)} \right\}.$$

我们需要证明 $|B_0| \geq \varepsilon|B|$。为引出矛盾，假设 $|B_0| < \varepsilon|B|$。一方面，我们知道

$$\begin{aligned}
\sum_{\ell\in B}\sum_{j\in\Gamma_j(\ell)} x_\ell^j &= \sum_{j\in\hat{K}}\sum_{\ell\in\Gamma_B(\ell)} x_\ell^j \\
&\geq \Delta\cdot\sum_{j\in\hat{J}} x_j && \text{by (14)} \\
&= \Delta x(J)\cdot\frac{x(\hat{J})}{x(J)} \\
&\geq \frac{1}{2}\sum_{j\in\hat{J}}\sum_{\ell\in\Gamma_B(\ell)} x_\ell^j\cdot\frac{x(\hat{J})}{x(J)} && \text{by (14)} \\
&= \frac{1}{2}\sum_{\ell\in B}\sum_{j\in\Gamma_J(\ell)} x_\ell^j\cdot\frac{x(\hat{J})}{x(J)} \\
&\geq \frac{1}{2}\sum_{\ell\in B} z_\ell\cdot\frac{x(\hat{J})}{x(J)} && \text{by (13)} \\
&\geq \frac{|B|}{2\log k\log m}\cdot\frac{x(\hat{J})}{x(J)}. && \text{by (12)}
\end{aligned}$$

另一方面，根据我们的假设，我们有

$$\begin{aligned}
\sum_{\ell\in B}\sum_{j\in\Gamma_j(\ell)} x_\ell^j &= \sum_{\ell\in B_0}\sum_{j\in\Gamma_j(\ell)} x_\ell^j + \sum_{\ell\in B\setminus B_0}\sum_{j\in\Gamma_j(\ell)} x_\ell^j \\
&< \sum_{\ell\in B_0}\sum_{j\in\Gamma_j(\ell)} x_\ell^j + \frac{|B|}{4\log k\log m}\cdot\frac{x(\hat{J})}{x(J)} && \text{by def. of } B_0 \\
&\leq 2e\ln(2k)\sum_{\ell\in B_0} z_\ell + \frac{|B|}{4\log k\log m}\cdot\frac{x(\hat{J})}{x(J)} && \text{by (13)} \\
&\leq |B_0|\cdot 2e\ln(2k) + \frac{|B|}{4\log k\log m}\cdot\frac{x(\hat{J})}{x(J)} \\
&< \frac{|B|}{2\log k\log m}\cdot\frac{x(\hat{J})}{x(J)}, && \text{assuming } |B_0| < \varepsilon|B|
\end{aligned}$$

wh这与前面的界矛盾。因此，我们必须有 $|B_0| \geq \varepsilon|B|$，如所需 . □

现在我们来分析算法 2 的近似保证。首先给出集合 J_0 的线性规划值总和以及该集合中顶点的线性规划值总和的简单下界。

引理 6.7. 算法 *2* 中定义的集合 J_0 的线性规划值至少为 $x(J)/(2\log m)$，且该集合中各元素的线性规划值下界不超过 $x_0 \geq 1/m$。

证明. 根据我们的 sim 在这些简化假设下，我们有 $\Delta \leq k/m^\varepsilon$，因此根据命题 6.4，集合 J

算法 2 MMSA 的近似算法[4]

输入： $B,\ J,\ R,\ S,\ E$

猜测 最优值 Δ 并求解线性规划 ▷ 例如，使用二分搜索

选择 参数 s，使桶 $J_s = \{j \in J \mid 2^{-s} \le x_j \le 2^{-(s-1)}\}$ 的 LP 权重即 $\sum_{j\in J_s} x_j$ 最大化，并令 $x_0 = 2^{-s}$ 和 $J_0 = J_s$。

对每个 $j \in J_0$ 和 $i \in \Gamma_R(j)$ **执行**

选择 一个新的参数 s，使桶 $S_s^{ji} = \{h \in S \mid 2^{-s} \le \hat{w}_h^{(j)} \le 2^{-(s-1)}\}$ 的条件线性规划权重（即 $\sum_{h\in S_s^{ji}} \hat{w}_h^{(j)}$）最大，并令 $\beta_{ji} = 2^{-s}$。

选择参数 t，使子桶 $\hat{S}_t^{ji} = \{h \in S_s^{ji} \mid 2^{-t} \le w_h \le 2^{-(t-1)}\}$ 的基数 $|\hat{S}_t^{ji}|$最大，并令 $\gamma_{ji} = 2^{-t}$ 和 $\hat{\Gamma}_j(i) = \hat{S}_t^{ji}$。

end for

对每个 $j \in J_0$ 和 $\beta,\ \gamma$ **执行**

令 $\Gamma^R_{\beta,\ \gamma}(j) = \{i \in \Gamma_R(j) \mid \beta_{ji} = \beta,\ \ \gamma_{ji} = \gamma\}$。

令 $\Gamma^S_{\beta,\ \gamma}(j) = \bigcup_{i\in\Gamma^R_{\beta,\ \gamma}(j)} \hat{\Gamma}_j(i)$。

结束循环

对于每个 $\beta,\ \gamma$ 和 $D \in \{2^{s-1} \mid s \in \lceil \log |S| \rceil\}$ **执行**

令 $J^D_{\beta,\gamma} = \{j \in J_0 \mid \Gamma^R_{\beta,\gamma}(j) \ne \emptyset,\ \ |\Gamma^S_{\beta,\gamma}(j)| \in [D, 2D]\}$。

令 $T^D_{\beta,\gamma} = \{\langle j, \Gamma^R_{\beta,\gamma}(j)\ \Gamma^S_{\beta,\gamma}(j)\rangle \mid j \in J^D_{\beta,\gamma}(j)\}$.

结束循环

令 $P_1 = \{\langle \beta, \gamma, D \mid \beta/\gamma > A, \beta D > A \cdot$最优值$/x(J_0)\rangle\}$，并且

$$J_1 = \bigcup_{\langle\beta,\gamma,D\rangle\in P_1} J^D_{\beta,\gamma}.$$

如果 $|J_1| < |J_0|/2$ **那么**

令 $J_{\mathrm{ALG}} = \emptyset$。对于所有 $j \in J_0 \setminus J_1$，独立地以概率 x_0 将 j 加入 J_{ALG}。结束循环。令 $S_{\mathrm{ALG}} = \emptyset$。对于每个 β，令 $S_\beta = \bigcup_{j\in J_{\mathrm{ALG}}} \bigcup_\gamma \Gamma^S_{\beta,\gamma}(j)$。对于所有 $h \in S_\beta$，独立地以概率 $\min\{1, \beta \cdot 12 \log |S| \log(|S|^2 m) \ln n\}$ 将 h 加入 S_{ALG}。结束循环。结束循环。否则，如果 $|J_1| \ge |J_0|/2$，则选择使基数 $|J^D_{\beta,\gamma}|$ 最大的 $\langle\beta, \gamma, D\rangle \in P_1$，并令 $J_2 = J^D_{\beta,\gamma}$。令 $S_{\tilde{D}} = \{h \in S \mid \{j' \in J_2 \mid h \in \Gamma^S_{\beta,\gamma}(j')\}| \in [\tilde{D}, 2\tilde{D}]\}$，对于每个 $\tilde{D} \in \{2^{s-1} \mid s \in \lceil \log |J_2| \rceil\}$。选择使基数 $|\{\langle j, h\rangle \in J_2 \times S_{\tilde{D}} \mid h \in \Gamma^S_{\beta,\gamma}(j)\}|$ 最大的 $\tilde{D}$，并令 $\tilde{S} = S_{\tilde{D}}$。选择使总线性规划值 $\sum_{j\in J_2:\Gamma^S_{\beta,\gamma}(j)\ni h_0} \hat{x}^{(h_0)}$ j 最大的 $h_0 \in \tilde{S}$。令 $J_{\mathrm{ALG}} = \{j \in J_2 \mid h_0 \in \Gamma^S_{\beta,\gamma}(j)\}$。令 $S_{\mathrm{ALG}} = \emptyset$。对于每个 $h \in \bigcup_{j\in J_{\mathrm{ALG}}} \Gamma_S(\Gamma_R(j))$，独立地以概率 $\min\{1, \hat{w}_h^{h_0} \cdot 4\gamma \ln n/(x_0\beta)\}$ 将 h 加入 S_{ALG}。结束循环。结束条件判断。

总 LP 权重至少为

$$\begin{aligned}\sum_{j\in J} x_j &\geq \frac{1}{2\Delta}\sum_{\ell\in B} z_\ell \\ &\geq \frac{k}{\Delta \log k \log m} && \text{by (12)} \\ &\geq \frac{m^{\varepsilon}}{\log k \log m},\end{aligned}$$

因此，特别是，$x(J) = \omega(1)$，且 J 中 LP 值至多为$1/m$ 的顶点的 LP 权重总和不超过 $\sum_{j\in J:\, x_j\leq 1/m} x_j \leq 1 \leq x(J)/2$；并且存在某个 $s \in [\log m]$，使得 J_s的 LP 权重至少为 $x(J)/(2\log m)$，这给出了我们对 J_0（最重的桶）的下界。此外，我们知道，对于 $2^{-s} < 1/m$，最重的桶不可能是 J_s，因为所有 LP 值不超过该值的顶点的 LP 权重总和至多为 $O(x(J)/m^{\varepsilon})$。因此，$x_0 \geq 1/m$。

□

接下来，我们考察 S中邻居的分桶，并给出这些分桶集合中顶点数量的下界。

引理 6.8。 在算法 *2* 中，对于每个顶点 $j \in J_0$ 以及每个红色邻居 $i \in \Gamma_R(j)$，其邻居的分桶集合 $\hat{\Gamma}_j(i)$ 的基数至少为 $|\hat{\Gamma}_j(i)| \geq 1/(6\beta_{ji}\log|S|\log(|S|^2 m))$。

证明。 固定顶点 $j \in J_0$ 和 $i \in \Gamma_R(j)$。我们先考察对 β_{ji}的选择。注意，提升约束 17 后，得到 $x_j = X_j^{(j)} \leq X_i^{(j)} (\leq x_j)$，因此 $\hat{y}^{(j)} = X_i^{(j)}/x_j = 1$。进而，提升约束 (16) 可得

$$\sum_{h\in\Gamma_S(i)} \hat{w}_h^{(j)} \geq 1.$$

注意，集合 $S'_{ji} = \{h \in \Gamma_S(i) \mid \hat{w}_h^{(j)} \leq 1/|S|^2\}$ 的 LP 权重总和至多为 $1/|S| \leq \hat{w}^{(j)}(\Gamma_S(i))/3$。因此，对于满足 $2^{-s} \geq 1/|S|^2$ 的 s，分桶后的集合 S_s^{ji}的总 $\hat{w}^{(j)}$ LP 权重至少为 $2/3\,\hat{w}^{(j)}(\Gamma_S(i))$，并且其中至少一个分桶后的集合的 LP 权重至少是该值的 $1/(2\log|S|)$倍，即至少为 $\hat{w}^{(j)}(\Gamma_S(i))/(3\log|S|) \geq 1/(3\log|S|)$。这给出了定义 β_{ji}的桶的 LP 权重下界。此外，最重的桶不可能是满足 $2^{-s} \leq 1/|S|^2$的 s，因为这些桶的总权重至多为 $1/|S| = o(1/(3\log|S|))$。特别地，这意味着 $\beta_{ji} \geq 1/|S|^2$。此外，对于满足 $2^{-s} = \beta_{ji}$的 s，由于 S_s^{ji}的条件 LP 权重总和至少为 $1/(3\log|S|)$，且集合中的每个顶点的条件 LP 值至多为 $2\beta_{ji}$，因此该集合的基数至少为 $1/(6\beta_{ji}\log|S|)$。

现在让我们考察分桶的第二阶段。注意，对于每个 $h \in \Gamma_S(i)$，我们都有

$$w_h \geq X_h^{(j)} = \hat{w}_h^{(j)} \cdot x_j \geq \beta_{ji} x_0 \geq 1/(|S|^2 m)$$

（当然，还有 $w_h \leq 1$）。因此，非空桶的数量 $\hat{S}_t^{ji}$ 至多为 $\log(|S|^2 m)$，且其中至少有一个的基数至少为$|S_s^{ji}|/\log(|S|^2 m)$；结合我们上面对 $|S_s^{ji}|$ 的下界，即可得到所需的 $|\hat{\Gamma}_j(i)|$下界。

□

注意，从上述证明中，我们还可以得到能够产生非空桶的 β_{ji} 和γ_{ji} 的取值数量上界。具体而言，我们得到以下界：

观察 6.9。 β_{ji}的可能取值总数至多为 $2\log|S|$，而 γ_{ji} 的可能取值总数至多为 $\log(|S|^2 m)$。加上 D的取值范围后，使 $J_{\beta,\gamma}^D$非空的三元组 $\langle\beta,\gamma,D\rangle$ 总数至多为 $2\log^2|S|\log(|S|^2 m)$。

算法通过分离与参数相对应的桶来进行；对于这些参数，˜简单舍入（从 J 中抽取大小为 $\Omega(J)$ 的随机子集）能够朝着达到 $O(A)$ 的˜近似保证迈进。如果 J_0 中很大一部分顶点只参与这类桶，那么算法就会应用这种舍入。下面的引理给出了算法在此情形下的分析。

引理 6.10. 在算法 *2* 中，如果 $|J_1| < |J_0|/2$，那么该算法以高概率采样出一个˜子集 $J_{\mathrm{ALG}} \subseteq J$，它覆盖 $\Omega(1)$ 比例的蓝色顶点；并采样出一个大小为 $O(A\cdot$最优值$)$ 的 S 子集，它覆盖 J_{ALG} 的所有红色邻居。

证明。 我们先分析被覆盖的蓝色顶点数量。首先，我们可以将集合 $J_0 \setminus J_1$ 的线性规划值界定为

$$\begin{aligned}
x(J_0 \setminus J_1) &= x(J) - x(J_1) && \text{since } J_1 \subseteq J_0 \\
&= x(J) - \frac{x(J_1)}{|J_1|} \cdot |J_1| \\
&\geq x(J) - 2 \cdot \frac{x(J_0 \setminus J_1)}{|J_0 \setminus J_1|} \cdot |J_1| && \text{since } \forall j \in J : x_j \in [x_0, 2x_0] \\
&> x(J) - 2 \cdot \frac{x(J_0 \setminus J_1)}{|J_0 \setminus J_1|} \cdot |J_0 \setminus J_1| && \text{since } |J_1| < |J_0|/2 \\
&= x(J) - 2x(J_0 \setminus J_1) \qquad \Longrightarrow \qquad x(J_0 \setminus J_1) > x(J_0)/3.
\end{aligned}$$

特别地，由引理 6.7 可得 $x(J_0 \setminus J_1) > x(J)/(6 \log m)$。因此，由断言 6.6 可知，存在一个 $B_0 \subseteq B$ 的蓝色子集，其基数为

$$|B_0| \geq \frac{|B|}{48e \ln(2k) \log k \log^2 m},$$

使得对于每个 $\ell \in B_0$，都有

$$\sum_{j \in \Gamma_{J \setminus J_0}(\ell)} x_j^\ell \geq \frac{1}{24 \log k \log^2 m}.$$

由于上文中的每个邻居 j 都满足 $x_j^\ell \leq x_j \leq 2x_0$，我们得到每个 $\ell \in B_0$ 至少有

$$|\Gamma_{J_0 \setminus J_1}(\ell)| \geq \frac{1}{48 \log k \log^2 m} \cdot \frac{1}{x_0}$$

$J_0 \setminus J_1$中的邻居。因此，当算法以概率 x_0独立地对每个顶点 $j \in J_0 \setminus J_1$ 进行采样时，以高概率至少覆盖

$$\Omega\left(\frac{1}{\log k \log^2 m}\right) \cdot |B_0| = \Omega\left(\frac{|B|}{\log^3 k \log^4 m}\right)$$

蓝色顶点。

在分析算法产生的成本之前，我们先指出，它以高概率输出一个可行解。事实上，根据引理 6.8，对于每个 $j \in J_{\mathrm{ALG}}$ 以及每个红色邻居 $i \in \Gamma_(j)$，红色顶点 i 至少有 $1/(6\beta_{ji} \log |S| \log(|S|^2 m))$ 个邻居属于 S；根据定义，这些邻居在算法当前步骤中都属于 S_β。因此，每个这样的邻居 i 被覆盖的概率至少为 $1 - 1/n^2$，并且以高概率， J_{ALG}的所有红色邻居都会被覆盖。

现在我们来分析这个解的成本 $|S_{\mathrm{ALG}}|$。根据 J_1 的定义，对于任意与 $J_0 \setminus J_1$ 相交的集合 $J^D_{\beta,\gamma}$，必有 $\beta/\gamma \le A$ 或 $\beta D \le A \cdot \mathrm{OPT}/(x(J_0))$。等价地，对于任意$j \in J_0 \setminus J_1$，如果 $\Gamma^S_{\beta,\gamma}(j) \neq \emptyset$，那么

$$\frac{\beta}{\gamma} \le A, \qquad \text{or} \tag{18}$$

$$\beta|\Gamma^S_{\beta,\gamma}(j)| \le \frac{2A \cdot \mathrm{OPT}}{x(J_0)}. \tag{19}$$

因此，对于每个 β，我们可以将集合 S_β 划分为两个（可能重叠的）部分 $S_\beta = S^1_\beta \cup S^2_\beta$，其中

$$S^1_\beta = \bigcup_{j \in J_{\mathrm{ALG}}} \bigcup_{\gamma:\beta/\gamma \le A} \Gamma^S_{\beta,\gamma}(j) \qquad \text{and} \qquad S^2_\beta = \bigcup_{j \in J_{\mathrm{ALG}}} \bigcup_{\gamma:\beta|\Gamma^S_{\beta,\gamma}(j)| \le 2A\cdot\mathrm{OPT}/x(J_0)} \Gamma^S_{\beta,\gamma}(j).$$

我们将分别分析从这两个部分中采样的顶点数量 y.

从 S^1_β 开始，对于每个满足 (18) 的 γ，考虑子集

$$S^1_{\beta,\gamma} = \bigcup_{j \in J_{\mathrm{ALG}}} \Gamma^S_{\beta,\gamma}(j).$$

注意，这些子集的并集恰好是 S^1_β。注意，对于每个顶点 $h \in S^1_{\beta,\gamma}$，我们有$w_h \ge \gamma$，因此，由于线性规划总值至多为最优值，我们有 $|S^1_{\beta,\gamma}| \le \mathrm{OPT}/\gamma$。以高概率，从该集合中采样的顶点数量（在对应于 β 的迭代中）至多为 $O(\beta \cdot \log(S) \log(|S|m) \log n \cdot \mathrm{OPT}/\gamma)$；并且由于根据观察 6.9，能够对 S_{ALG} 作出贡献的 $\langle \beta, \gamma \rangle$ 数对至多为 $O(\log |S| \log(|S|m)$，我们得到对这一贡献的总上界为 S_{ALG}

$$O(\beta \cdot \log^2 |S| \log^2(|S|m) \log n \cdot \mathrm{OPT}/\gamma) \le O(A \cdot \mathrm{OPT} \cdot \log^2 |S| \log^2(|S|m) \log n),$$

其中不等式由（18）得出。

现在让我们考察各个 S^2_β 的贡献。注意，对于满足（19）的任意 $j \in J_0 \setminus J_1$ andβ, γ, $\Gamma^S_{\beta,\gamma}$ 所对应的顶点的期望数量对 S_{ALG} 的贡献至多为

$$O\left(|\Gamma^S_{\beta,\gamma}(j) \cdot \beta \cdot \log |S| \log(|S|m) \log n|\right) \le O\left(\frac{A \cdot \mathrm{OPT}}{x(J_0)} \cdot \log |S| \log(|S|^2 m) \log n\right),$$

并且由于以高概率有 $|J_{\mathrm{ALG}}| = O(x(J_0))$，结合上述关于可能产生贡献的数对 $\langle \beta, \gamma \rangle$ 的上界，我们得到总上界

$$O(A \cdot \mathrm{OPT} \cdot \log^2 |S| \log^2(|S|m) \log n)$$

与先前对 S_{ALG} 的贡献类似，这就完成了对 $|S_{\mathrm{ALG}}|$ 的上界证明。

□

最后，我们转向算法 2 中的剩余情况，即 $|J_1| \le |J_0|/2$。对此情况的分析基于一种反向度数论证，类似于（尽管复杂得多）引理 3.2 中针对红蓝集合覆盖的论证。事实上，我们证明如下:

引理 6.11. 如果 $|J_1| \geq |J_0|/2$，那么对于算法在此情况下定义的 β、γ、D、h_0 和集合 J_{ALG}，我们有

$$\sum_{j \in J_{\mathrm{ALG}}} \hat{x}_j^{(h_0)} \geq \frac{|J_0| D x_0 \beta}{\mathrm{OPT}} \cdot \frac{1}{4 \log^2 |S| \log(|S|^2 m) \log m}.$$

此外，对于每个 顶点 $j \in \tilde{J}$（由算法定义），我们有 $\hat{x}_j^{h_0} \in [x_0\beta/(2\gamma) \quad 4x_0\beta/\gamma]$.

˜ ˜证明。我们从第二个结论开始。根据 $S\hat{}_D$和 J的定义，对于每个 $j \in J$，我们有$w_{h_0} \in [\gamma, 2\gamma]$、$\hat{}w_h^{(j)} \in [\beta, 2\beta]$，当然还有 $x_j \in [x_0, 2x_0]$。由于根据定义我们有 $\hat{}$ $w_{h_0}^{(j)} = X_{h_0}^{(j)}/x_j$和 $\hat{}x^{(h_0)}$ $_j$ $/ = X^{(h_0)}$ $_j$ $/w_{h_0} = X_{h_0}^{(j)}/w_{h_0}$，因此有 $\hat{}x_j^{(h)} = x_j \hat{}w_h^{(j)}/w_h$，于是 $\hat{}$ $x_j^{(h)}$ 的界立即由 x_j、 w_h和 $\hat{}w_h^{(j)}$ 的相应界得到。

现在来证明 J_{ALG} 的条件 LP 权重下界。首先注意，根据我们对 $\langle \beta, \gamma, D \rangle$ 的选择、观察 6.9 以及算法当前的情形，集合 J_2 的基数可以界定为

$$|J_2| \geq \frac{|J_1|}{|P_1|} \geq \frac{|J_0|}{2|P_1|} \geq \frac{|J_0|}{4 \log^2 |S| \log(|S|^2 \log m)}.$$

现在考虑数对集合 $E_1 = \{\langle j, h \rangle \in J_2 \times S \mid h \in \Gamma^S_{\beta,\gamma}(j)\}$。由于 $J_2 \subseteq J^D_{\beta,\ \gamma}$，我们可以将该集合中的数对数量界定为 $|E_1| \geq |J_2|D$。因此，根据我们对 D 的选择，有

$$|\{\langle j, h \rangle \in J_2 \times \tilde{S} \mid h \in \Gamma^S_{\beta,\gamma}(j)\}| \geq \frac{|J_2|D}{\log m}.$$

现在可以使用平均论证来界定 S_{ALG}的条件 LP 值。首先，根据上文观察到的 x_0 和 $\hat{}x_j^h$ 的界，我们有

$$\begin{aligned} \sum_{h \in \tilde{S}} w_h \sum_{j \in J_2 : \Gamma^S_{\beta,\gamma}(j) \ni h} \hat{x}_j^{(h)} &\geq |\{\langle j, h \rangle \in J_2 \times \tilde{S} \mid h \in \Gamma^S_{\beta,\gamma}(j)\}| x_0 \beta \\ &\geq \frac{|J_2| D x_0 \beta}{\log m}. \end{aligned}$$

˜另一方面，我们知道 $\sum_{h \in \tilde{S}} w_h \leq$ 最优值，因此必定存在某个 $h \in S$，使得

$$\begin{aligned} \sum_{j \in J_2 : \Gamma^S_{\beta,\gamma}(j) \ni h} \hat{x}_j^{(h)} &\geq \frac{|J_2| D x_0 \beta}{\mathrm{OPT}} \cdot \frac{1}{\log m} \\ &\geq \frac{|J_0| D x_0 \beta}{\mathrm{OPT}} \cdot \frac{1}{4 \log^2 |S| \log(|S|^2 \log^2 m)}. \end{aligned}$$

根据我们对 J_2的界

特别地，我们对 h_0 的选择必须满足这一性质。 □

现在可以证明，在这种情况下，该算法朝着 $\tilde{O}(m/A^2)$-˜近似迈进。将这一进展与引理 6.10 保证的、朝着 $O(A)$-近似迈进的进展权衡后，通过设定 $A = m^{1/3}$，我们得到 $\tilde{O}(m^{1/3})$-近似。

引理 6.12. 在算法 *2* 中，如果 $|J_1| \geq |J_0|/2$，那么该算法以高概率朝着 $\tilde{O}(m/A^2)$ 的近似保证取得进展。

证明。 首先，我们来界定由 J_{ALG} 覆盖的蓝色顶点数量。根据引理 6.11，我们有

$$\begin{aligned}
\hat{x}^{h_0}(J_{\mathrm{ALG}}) &\geq \frac{|J_0| x_0 D\beta}{\mathrm{OPT}} \cdot \frac{1}{4\log^2|S|\log(|S|^2 m)\log m} \\
&\geq \frac{x(J_0) D\beta}{\mathrm{OPT}} \cdot \frac{1}{4\log^2|S|\log(|S|^2 m)\log m} && \text{since } \forall j \in J_0 : 2x_0 \geq x_j \\
&\geq \frac{x(J) D\beta}{\mathrm{OPT}} \cdot \frac{1}{4\log^2|S|\log(|S|^2 m)\log m} && \text{by Lemma 6.7} \\
&> x(J) \cdot \frac{A}{x(J_0)} \cdot \frac{1}{4\log^2|S|\log(|S|^2 m)\log m}. && \text{since } \forall \langle \beta, \gamma, D\rangle \in P_1 : \frac{\beta D}{\mathrm{OPT}} > \frac{A}{x(J_0)}
\end{aligned}$$

因此，由于条件化线性规划解满足基本线性规划，我们可以将命题 6.5 应用于该解，从而得到 J_{ALG} 的蓝色邻域大小可被界定为

$$\begin{aligned}
|\Gamma_B(J_{\mathrm{ALG}})| &\geq \frac{1}{4e\ln(2k)\log k\log m} \cdot \frac{A}{x(J_0)} \cdot \frac{1}{4\log^2|S|\log(|S|^2 m)\log m} \cdot |B| \\
&= \frac{1}{16\ln(2k)\log k\log^2|S|\log(|S|^2 m)\log^2 m} \cdot \frac{A}{x(J_0)} \cdot |B|.
\end{aligned}$$

注意，根据线性规划约束和引理 6.11，对于每个红色邻居 $i \in \Gamma_R(J_{\mathrm{ALG}})$，我们有

$$\hat{y}_i^{h_0} \geq x_i^{h_0} \geq x_0\beta/(2\gamma),$$

因此，根据 (16)，重新缩放的解 $(\hat{w}_h^{h_0} \cdot 2\gamma/(x_0\beta))_{h\in S}$ 是 $\Gamma_R(J_{\mathrm{ALG}})$ 的分数集合覆盖。于是，以 $\min\{1, 2\ln n \cdot \hat{w}_h^{h_0} \cdot 2\gamma/(x_0\beta)\}$ 的概率抽取每个 $h \in S$，高概率下即可得到有效的集合覆盖。接下来分析该集合覆盖的大小。事实上，由于 $\hat{w}^{h_0}(S) \leq$最优值，我们的抽样过程生成的集合的期望大小为

$$\begin{aligned}
\mathbb{E}[|S_{\mathrm{ALG}}|] &\leq \frac{4\gamma\ln n}{x_0\beta} \cdot \mathrm{OPT} \\
&\leq 8\ln n \cdot \frac{\gamma}{\beta} \cdot \frac{|J_0|}{x(J_0)} \cdot \mathrm{OPT} && \text{since } \forall j \in J_0 : x_j \leq 2x_0 \\
&\leq 8\ln n \cdot \frac{\gamma}{\beta} \cdot \frac{m}{x(J_0)} \cdot \mathrm{OPT} \\
&< 8\ln n \cdot \frac{1}{A} \cdot \frac{m}{x(J_0)} \cdot \mathrm{OPT}, && \text{since } \forall \langle \beta, \gamma, D\rangle \in P_1 : \frac{\beta}{\gamma} > A
\end{aligned}$$

因此，高概率下我们有 $|S_{\mathrm{ALG}}| = O($最优值 $\cdot m\log n/(A \cdot x(J_0))$.

综合我们的两个界限可知，在这种情况下，该算法能够逐步达到如下近似保证:

$$\frac{|B|}{|\Gamma_B(J_{\mathrm{ALG}})|} \cdot \frac{|S_{\mathrm{ALG}}|}{\mathrm{OPT}} = \tilde{O}(1) \cdot \frac{x(J_0)}{A} \cdot \frac{m}{A \cdot x(J_0)} = \tilde{O}(1) \cdot \frac{m}{A^2}.$$

□

7 MMSA 整数间隙示意图$_t$

本节给出 MMSA$_t$ 整数间隙的证明概要：将 Densest k-Subgraph 的困难实例拼接起来，作为 MMSA$_t$中的各层。我们的下界基于目前

已知的 Densest k-Subgraph 在 k不同参数范围内的最佳界。考虑 Erd¨os–R´enyi 随机图 $G(n,p)$，其中 $p=n^{\beta-1}$，并令 $k=n^{\alpha}$，其中某个 $\alpha,\beta\in(0,1)$。众所周知，对于任意 $\gamma>\alpha\beta$，基于对数密度的算法（以及 Sherali-Adams）都能检测出是否存在平均度为 n^{γ} 的植入式 k-子图；另一方面，对于 $\alpha\le 1/2$，即使是平方和也无法证明当 $\gamma<\alpha\beta$ [时不存在这样的子图（Chlamt´a˘c 和 Manurangsi, 2018]）。然而，对于 $\alpha>1/2$，情况有所不同。尽管 Sherali-Adams 和对数密度技术的局限仍为 $\gamma>\alpha\beta$，但简单的半定规划松弛（以及其他简单技术）可以证明当 $\gamma>\beta/2$时不存在稠密子图。

因此，我们可以将当前技术的局限（以及在简单的半定规划之上应用 Sherali-Adams 所得到的整数间隙）概括如下：对于 $p=n^{\beta-1}$ 中的 $G(n,\ p)$，当 $k=n^{\alpha}$ 时，我们能够证明不存在平均度为 $\omega(n^{\gamma})$ 的植入式 k-子图，当且仅当 $\gamma\ge\min\{\alpha,1/2\}\cdot\beta$。

整数间隙实例的构造。设 t 为正奇整数，且 $\varepsilon>0$。我们按如下方式构造一个分层图（表示深度为 t 的交替电路）：设 G_1、……、$G_{(t-1)/2}$ 为 $G(n,p)$ 的相互独立实例，且对于 $\beta=\varepsilon/(2-\varepsilon)$，有 $p=n^{-(2-2\varepsilon)/(2-\varepsilon)}=n^{\beta-1}$。对于每个 $i\in[(t-1)/2]$，令 U_{2i} 为与 G_i 的边对应的顶点集，并令 U_{2i+1} 为与 G_i 的顶点对应的顶点集。对于 G_i 中的每条边 $e=(u,v)$，在表示 U_{2i} 中边 e 的顶点与分别表示 U_{2i+1} 中 u 和 v 的顶点之间添加边。令 U_1 为大小为 $n^{(2-2\varepsilon)/(2-\varepsilon)}$ 的顶点集。最后，对于每个 $i\in[(t-1)/2]$，将每个顶点集 U_{2i} 任意划分为 $|U_{2i-1}|$ 个集合（因此，对于 $i=1$，划分为 $n^{(2-2\varepsilon)/(2-\varepsilon)}$ 个集合；对于 $i>1$，划分为 n 个集合），并将 U_{2i-1} 中第 j 个顶点连接到 U_{2i} 的划分中第 j 个集合内的所有顶点。下界实例的构造示例见图 1。

注意，顶点总数高度集中在 $n^{2/(2-\varepsilon)}\cdot(t-1)/2$附近。记 $m=n^{2/(2-\varepsilon)}\cdot(t-1)/2$。由于我们将 t 视为常数，因此将下界表示为 m （顶点总数，忽略常数因子）的函数来分析。为此，注意在所有图 G_i中，都有 $n\approx m^{1-\varepsilon/2}$，且 $p=n^{-(2-2\varepsilon)/(2-\varepsilon)}\approx m^{-(1-\varepsilon)}$。

整数间隙的证明概要。我们先分析最优值。我们需要覆盖 U_1 中的所有 $|U_1|\approx m^{1-\varepsilon}$ 个顶点，这迫使我们在 U_2 中选择同样数量的顶点。由于 U_2 中的每个顶点都通过与门恰好连接到 U_3 中的两个顶点，因此要满足在 U_2 中任意选择的 $|U_1|$ 个顶点，至多需要在 U_3 中选择 $2|U_1|$ 个顶点。然而，这同时也是一个下界：因为在 U_3 中任意选择 $m^{1-\varepsilon}$ 个顶点（对应于在 G_1 中选择 $k_1=m^{1-\varepsilon}$ 个顶点），会得到一个期望边数为 $pk_1^2\approx m^{-(1-\varepsilon)+2(1-\varepsilon)}=m^{1-\varepsilon}$ 的子图；切尔诺夫界表明，不存在规模显著更小、却含有这么多诱导边的子图。接下来可以对其余各层重复这一论证，因为每次我们最终都会得到 U_{2i-1} 中 $\Omega(m^{1-\varepsilon})$ 个需要满足的顶点，这迫使我们在 U_{2i} 中选择同样数量的顶点（因为对于 U_{2i-1} 中任意一对顶点，它们在 U_{2i} 中的公共邻居数为零，并且对应的 $U_{2i}\to U_{2i-1}$ 层是或层）；而在 G_i 中含有这么多边的子图必须有 $\Omega(m^{1-\varepsilon})$ 个顶点，这意味着我们在 U_{2i+1} 中至少需要一个这么大的集合。因此，最优值为 $\Theta(m^{1-\varepsilon})$。

现在来看看基本 SDP 和 Sherali-Adams 中哪些分数解能够存续。首先我们知道，对于 U_3 中以 $k_1^*\ge\sqrt{n}$ 个顶点为目标的情形，我们有一个分数解，它表示为一个含有 $k_1^*\sqrt{pn}$ 条边的 k_1^*-子图。因此，为了实现我们在 G_1 中（表示子图中的边）的目标分数权重 $m^{1-\varepsilon}$（位于 U_2 中），我们需要选择 k_1^*，使得 $k_1^*\sqrt{pn}=m^{1-\varepsilon}$，这意味着 $k_1^*\approx\frac{m^{1-\varepsilon}}{m^{-(1-\varepsilon)/2+(2-\varepsilon)/4}}=m^{1-5\varepsilon/4}$。因此，在我们的分数解中，LP 权重 $m^{1-5\varepsilon/4}$ 均匀分配在 U_3 的顶点上。我们可以重复这一论证，证明 U_{2i-1} 的 LP 权重为 $\approx m^{1-(3+i)\varepsilon/4}$；而要在 G_i 中获得具有该边权的分数解，

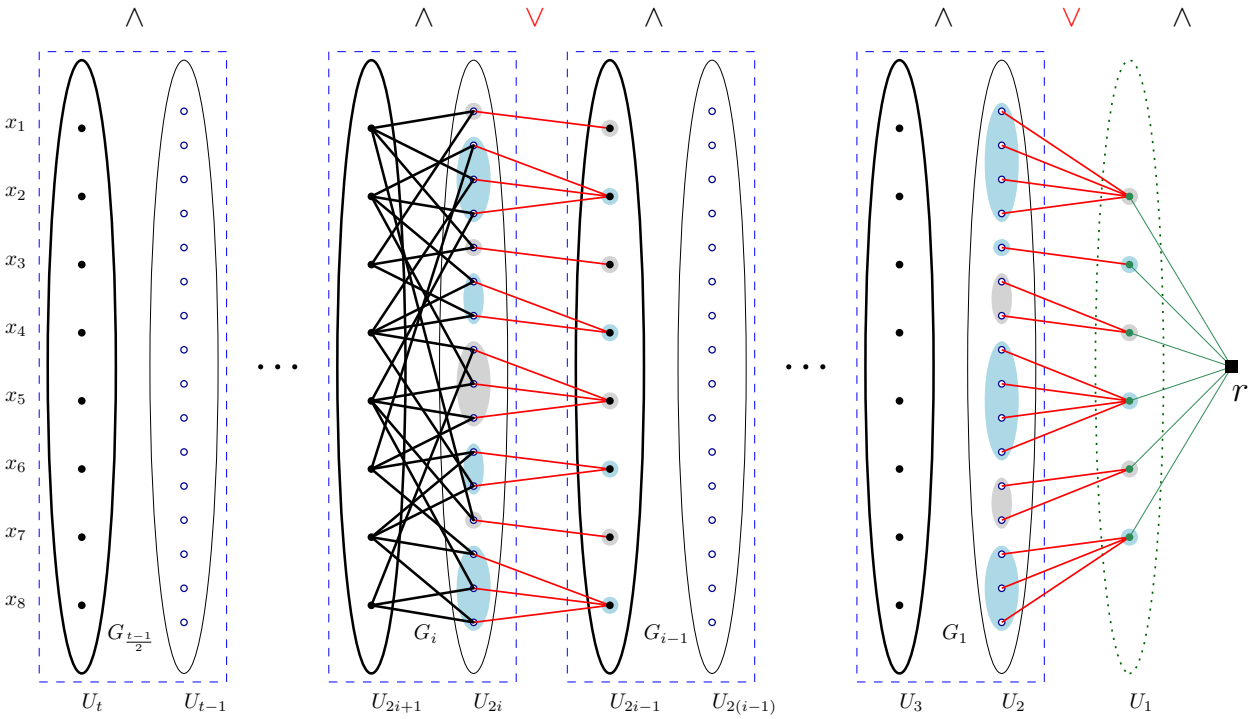


图 1：该图概述了困难性实例的构造。图 $t-1/2$ 的二部图表示 G_1，$\cdots$，$G_{\frac{t-1}{2}}$ 以及目标集合 U_1 构成了我们 MMSA_t 实例的各层。每个 G_i 都表示为一个二部图，其中包含两组节点 $U_{2i}=V(G_i)$ 和 $U_{2i+1}=E(G_i)$；当且仅当 v 是 e 在 G_i 中的一个端点时，$e\in U_{2i}$ 与 $v\in U_{2i+1}$ 之间才有一条边（这些是粗黑边，对应与门）。此外，不同 G_i 的层通过映射 G_i 的边节点（即 U_{2i} 中的顶点）到 G_{i-1} 的顶点节点（即 U_{2i-1} 中的顶点）拼接在一起，并添加 U_1 以处理与 G_1 的连接。这些中间边在图中以红色表示，对应或门。

只需将分数顶点权重设为 $k_i^*=m^{1-(4+i)\varepsilon/4}$，只要 $k_i^*\geq\sqrt{n}$ 即可。也就是说，只要 $m^{1-(4+i)\varepsilon/4}\geq m^{(2-\varepsilon)/4}$，这对应于 $i\leq\frac{2}{\varepsilon}-3$。一旦 $i\geq\frac{2}{\varepsilon}-3$，我们就可以利用对数密度下界来获得分数边权重 $k_{i-1}^*=m^{1-(3+i)/4}$，并将分数顶点权重设为 $k_i^*=m^{2\alpha_i/(1-2\varepsilon)}\approx n^{\alpha_i}$，使得

$$m^{1-(3+i)\varepsilon/4}=k_i^*n^{\alpha_i\beta}=k_i^*n^{\alpha_i\cdot\varepsilon/(2-\varepsilon)}\approx n^{\alpha_i}n^{\alpha_i\cdot\varepsilon/(2-\varepsilon)}=n^{2\alpha_i/(2-\varepsilon)}\approx m^{\alpha_i}.$$

回想一下，只要满足 $\alpha_i\leq 1/2$，这个下界就成立。事实上，此时由于 $i\geq\frac{2}{\varepsilon}-3$，我们得到$\alpha_i=1-(3+i)_4^{\varepsilon}\leq\frac{1-2}{4}=\frac{1}{2}$。从这里开始，$k_i^*$ 的序列持续递减，因此始终不超过 $\sqrt{n}$，我们也就始终可以使用对数密度下界。

现在我们进行类似的计算。若松弛将分数权重 $k_{i-1}^*=m^{\delta_{i-1}}$放在 U_{2i-1}上，那么要在 G_i中得到这一分数边权重，我们需要分数顶点权重 $k_i^*=n^{\alpha_i}$，使得

$$m^{\delta_{i-1}}\approx n^{2\alpha_i/(2-\varepsilon)}\approx m^{\alpha_i}$$

U_{2i+1} $k_i^*=n^{\delta_{i-1}}=()$，因此，$m^{\delta_{i-1}\cdot(2-\varepsilon)/2}$放在上的分数权重将为$()$。因此，每一步指数都会缩小一个$(2-\varepsilon)/2$的因子，再经过额外的 $\log(1/(2\varepsilon))/\log(2/(2-\varepsilon))$ 步，就会降至 m^{ε}。由此可知，对于任意 $t\geq\frac{2}{\varepsilon}-3+\frac{\log(1/(2\varepsilon))}{\log(2/(2-\varepsilon))}$，$\mathrm{MMSA}_t$ 的整数间隙至少为 $m^{1-2\varepsilon}$。因此，随着 t 增大，整数间隙确实趋近于线性，其依赖关系至多为 $t=O((1/\varepsilon)\log(1/\varepsilon))$，换言之，

整数间隙为 $m^{1-O((\log t)/t)}$。

参考文献

V. Abidha 和 P. Ashok。轴平行超平面及其他对象上的红蓝集合覆盖问题。*arXiv* 预印本 *arXiv:2209.06661*，2022年。M. Alekhnovich、S. Buss、S. Moran 和 T. Pitassi。最小命题证明长度的 np 困难性：无法对其进行线性近似。《符号逻辑杂志》，66(1):171–191，2001年。P. Ashok、S. Kolay 和 S. Saurabh。几何红蓝集合覆盖的多变量复杂性分析。*Algorithmica*，79(3):667–697，2017年。P. Awasthi、A. Blum 和 O. Sheffet。析取式的无关学习的改进保证。载于学习理论会议论文集，第359–367页，2010年。R. D. Carr、S. Doddi、G. Konjevod 和 M. Marathe。关于红蓝集合覆盖问题。载于离散算法研讨会论文集，第345–353页，2000年。T. M. Chan 和 N. Hu。适用于单位正方形的几何红蓝集合覆盖及相关问题。计算几何，48(5):380–385，2015年。M. Charikar、Y. Naamad 和 A. Wirth。关于目标集选择的近似。载于近似、随机化与组合优化国际研讨会论文集，2016年。E. Chlamt´aˇc 和 P. Manurangsi。与对数密度阈值匹配的 Sherali–Adams 整数间隙。近似、随机化与组合优化：算法与技术，2018年。E. Chlamt´aˇc、M. Dinitz、C. Konrad、G. Kortsarz 和 G. Rabanca。最稠密 k-子超图问题。载于近似、随机化与组合优化国际研讨会论文集：算法与技术，2016年。E. Chlamt´aˇc、M. Dinitz 和 Y. Makarychev。最小化并集：小集合二部图顶点扩张的紧确近似。载于离散算法研讨会论文集，第881–899页，2017年。I. Dinur 和 S. Safra。关于标签覆盖近似的困难性。《信息处理快报》，89(5):247–254，2004年。M. Elkin 和 D. Peleg。稀疏化器问题近似的困难性。《计算系统理论》，41(4):691–729，2007年。M. Goldwasser 和 R. Motwani。装配序列规划的棘手性：平面上的单位圆盘。载于*WADS* 论文集，第97卷，第307–320页，1997年。R. R. Madireddy 和 A. Mudgal。针对单位圆盘红蓝集合覆盖的常数因子近似算法。*Algorithmica*，第1–33页，2022年。R. R. Madireddy、S. C. Nandy 和 S. Pandit。关于几何红蓝集合覆盖问题。载于算法与计算国际研讨会论文集，第129–141页，2021年。

P. Miettinen。《正负部分集合覆盖问题》。《信息处理快报》，108(4):219–221，2008。

A 适配并应用我们的算法解决部分红蓝集合覆盖问题

ˆ 现在考虑这样一种变体：给定参数 k，并且只要求可行解至少覆盖 k 个元素。除以下几点外，算法及其分析几乎无需改动。

在算法中，循环的停止条件当然不再是覆盖所有蓝色元素，而是至少覆盖其中 k 个。

一个稍微微妙的改动涉及最后一次迭代中的线性规划舍入分析。如果红色元素与蓝色元素的覆盖比例仍然低至所要求的程度，但新增的蓝色元素数量远超所需，那么关于朝着某个特定近似保证取得进展的概念可能并不成立。无需对舍入过程进行去随机化；可以证明，尽管存在这个问题，该过程仍以高概率成功。下面简要概述这一论证。

首先请注意，我们总是可以预先丢弃包含超过最优值个红色元素的集合，因此可以假设 $r_\alpha \leq \mathrm{OPT}$。假设我们还需要覆盖 k^* 个元素，才能达到总共 k 个蓝色元素的目标。由于我们对 $\mathbb{E}[|\Gamma_{R_\alpha}(J^*)|]$ 的界是我们对 $\mathbb{E}[|J^* \setminus J_+|]$ 的界的线性˜函数，根据切尔诺夫界，除去指数级小的概率外，有 $|\Gamma_{R_\alpha}(J^*)| = O(r_\alpha A)$。另一方面，$|\Gamma_B(J^*)|$ 总是至多为 $|B|$，因此根据马尔可夫不等式，有

$$\mathrm{Prob}\left[|\Gamma_B(J^*)| \leq \frac{\mathbb{E}[|\Gamma_B(J^*)|]}{2}\right] \leq \frac{|B| - \mathbb{E}[|\Gamma_B(J^*)|]}{|B| - \mathbb{E}[|\Gamma_B(J^*)|]/2} \leq 1 - \frac{1}{2|B|}.$$

因此，在给定的一次迭代中，将舍入过程重复多项式次后，除去指数级小的概率外，我们可以找到一个集合 $J \subseteq J$，使其同时满足

$$|\Gamma_B(\hat{J})| \geq \frac{\mathbb{E}[|\Gamma_B(J^*)|]}{2} \qquad \text{and} \qquad |\Gamma_{R_\alpha}(\hat{J})|] = \tilde{O}(r_\alpha A).$$

现在，如果 $\mathbb{E}[|\Gamma_B(J^*)|] \leq 2k^*$，那么根据前面的分析，我们就能得到所需的比率，并将新增红色元素的数量限制在相应范围内。如果 $\mathbb{E}[|\Gamma_B(J^*)|] > 2k^*$，那么这将是最后一次迭代，因为我们至少会覆盖所需的 k^* 个额外蓝色元素，而在˜ ˜这一最后阶段新增的红色元素数量至多为 $O(r_\alpha A) \leq O(A \cdot \mathrm{OPT})$，因此我们仍能保持所需的近似比率。

B MMSA$_4$的其他线性规划约束

以下是除 MMSA$_4$ 的基本线性规划松弛之外所使用的全部提升约束：

$$
\begin{aligned}
&X_j^{(h)} = X_h^{(j)} && \forall j \in J, \forall h \in S \\
&X_j^{(j)} = x_j && \forall j \in J \\
&\sum_{h \in \Gamma_S(i)} X_h^{(j)} \geq X_i^{(j)} && \forall j \in J \forall i \in R \\
&X_j^{(j)} \leq X_i^{(j)} && \forall j \in J \forall i \in \Gamma_R(j) \\
&0 \leq X_a^{(j)} \leq x_j && \forall j \in J \forall a \in \{j\} \cup R \cup S \\
&\sum_{h'} w_{h'}^{(h)} \leq \text{最优值} \cdot w_h && \forall h \in S \\
&\sum_{h' \in S} X_\ell^{(h)} \geq |B| / (\log\ k \log\ m) && \forall h \in S \\
&X_\ell^{(h)} \leq \sum_{\ell \in B} X_\ell^{(h)}, && \forall h \in S \forall \ell \in B \\
&\Delta X_j^{(h)} \leq \sum_{j \in \Gamma_J(\ell)} X_\ell^{(h)}, && \forall h \in S \forall j \in J \\
&0 \leq X_\ell^{(h)} \leq 2\Delta X_j^{(h)} && \forall h \in S \forall \ell \in \Gamma_B(j) \\
&0 \leq X_j^{(h)}, X_\ell^{(h)} \leq w_h && \forall h \in S \forall \ell \in B \forall j \in J \\
&\sum_{h' \in S} X_{h'}^{(h)} \geq X_i^{(h)} && \forall h \in S \forall i \in R \\
&X_j^{(h)} \leq X_i^{(h)} && \forall h \in S \forall (j,i) \in E(J,R) \\
&0 \leq X_a^{(h)} \leq w_h && \forall h \in S \forall a \in B \cup J \cup R \cup S
\end{aligned}
$$