arXiv:2304.02752v2 [math.CO] 7 May 2023

DENSE CLUSTERS IN HYPERGRAPHS

YULY BILLIG

Abstract. 本文解决了在给定加权超图中寻找密度尽可能大的子超图这一问题。我们引入支撑矩阵的概念，并证明最优子超图的密度等于 $\|A^TA\|$，其中 A 是一个最优支撑矩阵。或者，子超图的最大密度等于支撑矩阵列和的极小极大问题的解。我们引入超图的谱分解，并表明它是对 Dulmage-Mendelsohn 分解的重要细化。我们的理论结果给出了一个高效算法，用于寻找最大密度子超图；更一般地，还可用于求给定加权超图的谱分解。

1. Introduction

超图是图的推广，其中超边可以包含任意数量的顶点。本文介绍用于超图聚类的新工具。我们开发的方法适用于数据科学。例如，视频流媒体公司可能希望根据客户观看过的视频对客户进行聚类。这里的数据可以表示为一个超图，其中视频是顶点，客户由超边表示，而超边的支撑集是相应客户观看过的视频集合。兴趣相似的客户群体对应于所有客户记录构成的超图中的一个高密度子超图。

我们将超图 H 的密度 $\alpha(H)$ 定义为其超边数与顶点数之比。本文解决以下问题。

最密子图问题：在给定的超图中寻找密度最大的子超图。

2020年数学学科分类。主要分类 05C42、05C65；次要分类 05C85、90C35。

当然，本文获得的结果也适用于图，因为图是超图的特殊情形。为简洁起见，本文将子超图称为子图。

对于图，Goldberg [5] 在1984年给出了最密子图问题（DSP）的一种解法。Goldberg算法通过迭代过程求得DSP的最优解，其中每一步都需要针对与某个图相关联的特定网络求解最大流最小割问题，并在每次迭代时更新网络的容量。

Hu、Wu和Chan在 [8]将Goldberg算法推广到加权超图的情形。对于具有 n个顶点和 m 条超边的超图 H ， [8]中构造的网络族具有以下参数：顶点数为 $N = n+m+2$，边数为 $M = D+2n$，其中 D 是超图 H中各顶点度数之和。

Goldberg和Tarjan在 [7]中综述了已知最快的最大流问题算法。迄今为止，求解最大流问题的最佳算法由Goldberg和Rao提出 [6] ，其复杂度为$O(M\min(\sqrt{M}, N^{\frac{2}{3}})\log(N^2/M))$。因此，基于最大流的DSP算法的一次迭代可在复杂度为$O(D\min(\sqrt{D}, (n+m)^{\frac{2}{3}})\log((n+m)^2/D))$的时间内实现。该算法所需的迭代次数为 $O(\log(n))$。

本文介绍了一种用于 DSP 的新迭代算法，其每次迭代的复杂度大大降低，为 $O(D)$，空间需求也为 $O(D)$。对于具有整数权重的超图，我们为该算法给出了一个终止条件，该条件可保证构造的子图是最优的。仍需进一步研究以确定该算法的收敛速度，不过实验表明，对于一般超图，该算法在 $O(\log(n))$ 次迭代后终止。

也有一些用于DSP的快速贪心算法，但它们无法得出最优解，参见例如 [2]。

我们引入支撑矩阵的概念，它将是我们解决这一问题的主要工具。令 $V = \{v_j \mid j = 1, \ldots, n\}$, $E = \{e_i \mid i = 1, \ldots, m\}$ 为 H 的顶点和超边。

定义1.1。对于超图 H，支撑矩阵 A 是一个 $m \times n$实矩阵，具有以下性质：

(1) $a_{ij} \geq 0$ 对于所有 $i = 1$、……、 m、 $j = 1$、……、 n。
(2) 如果顶点 v_j 不属于超边 e_i ，则 $a_{ij} = 0$。 n(3)
$\sum a_{ij} = 1$ 对于所有 $i = 1$、……、 m。 $j=1$

当我们对超边进行聚类时，自然会关注它们所共有的顶点。然而，在评估两条超边彼此之间的接近程度时，与另一条超边共享一个属于许多其他超边的顶点，并不如与其共享一个度较低的顶点那么重要。支撑矩阵将使我们能够自动将这一考量纳入其中。

对于给定的超图 H，其支撑矩阵集合构成一个单纯形 $\mathcal{S}(H)$。我们的方法是在 $\mathcal{S}(H)$ 中执行优化过程。

对于 $A \in \mathcal{S}(H)$，矩阵 A^TA 是一个具有非负实特征值的对称矩阵。范数 $\|A^TA\|$ 等于其主特征值。我们将 H 中的最大密度子图与最优支撑矩阵 A的范数 A^TA 联系起来。

Theorem 1.2.

$$\max_{H' \subset H} \alpha(H') = \min_{A \in \mathcal{S}(H)} \|A^TA\|.$$

在 $\mathcal{S}(H)$ 中还有另一种优化过程，它会得到相同的最优支撑矩阵。根据定义，支撑$_m$矩阵中的行和等于 1。我们来考察列和 $s_j(A) = \sum a_{ij}$.$_{i=1}$令 $s_{\max}(A) = \max_{j=1,\ldots,n} {}^{s_j}(A)$。

Theorem 1.3.

$$\max_{H' \subset H} \alpha(H') = \min_{A \in \mathcal{S}(H)} s_{\max}(A).$$

容易看出， H 包含唯一的最大密度为 α_1的极大子图 $H_{\alpha_1} = (V_{\alpha_1}, E_{\alpha_1})$。考虑商超图$H/H_{\alpha_1}$（从 H 中移除 H_{α_1} 的所有超边，并从剩余超边的支集里删除 H_{α_1} 的所有顶点而得到）。令$H_{\alpha_2} = (V_{\alpha_2}, E_{\alpha_2})$ 为H/H_{α_1} 中密度为 α_2的极大子图。重复这一过程，我们构造出 H的谱分解，它是对 H的顶点集和超边集的一种划分

$$V = V_{\alpha_1} \dot\cup \ldots \dot\cup V_{\alpha_k}, \quad E = E_{\alpha_1} \dot\cup \ldots \dot\cup E_{\alpha_k},$$

从而在 H中得到一条子图链:

$$H_{\alpha_1} \subset H_{\alpha_1} \cup H_{\alpha_2} \subset H_{\alpha_1} \cup H_{\alpha_2} \cup H_{\alpha_3} \subset \ldots \subset H.$$

我们证明存在一个支撑矩阵 A，使得对于每个顶点 $v_j \in V_{\alpha_r}$，我们有 $s_j(A) = \alpha_r$。因此，可以从这个最优支撑矩阵中提取 H 的谱分解。

这样的最优支撑矩阵必然具有块结构，其中$a_{ij} \neq 0$ 仅当 $v_j \in V_{\alpha_r}$、 $e_i \in E_{\alpha_r}$ 对于某个 r。

超图 $H = (V, E)$ 有一个对偶超图 H^*，其中 E 是 H^* 的顶点集， V 是超边集；当且仅当 $v \in \mathrm{supp}_H(e)$ 时， $e \in \mathrm{supp}_{H^*}(v)$。我们证明，$H^*$ 的谱分解是 H 的谱分解的对偶。

定理1.4。设 $\{(V_{\alpha_1}, E_{\alpha_1}),(V_{\alpha_2}, E_{\alpha_2}).., (V_{\alpha_k}, E_{\alpha_k})\}$是超图 H 的谱分解，其因子的密度为 $\alpha_1 > \alpha_2 > \ldots > \alpha_k > 0$。那么 $\{(E_{\alpha_k}, V_{\alpha_k}). . ., (E_{\alpha_2}, V_{\alpha_2}),(E_{\alpha_1}, V_{\alpha_1})\}$是对偶超图 H^* 的谱分解，其因子的密度为 $\alpha_k^{-1} > \ldots > \alpha_2^{-1} > \alpha_1^{-1} > 0$。

超图的谱分解是其Dulmage-Mendelsohn分解的一项重要细化 [3] （定义见第6节）。

定理 1.5。设(H^+, H^0, H^-) 是超图 H 的杜尔马奇-门德尔松分解。则

$$H^+ = \bigcup_{\alpha>1} H_\alpha, \quad H^0 = H_1, \quad H^- = \bigcup_{\alpha<1} H_\alpha.$$

设 H 是一个密度为 α 的超图，使得H 中的任何子图的密度都不超过 α。那么，最优支撑矩阵的行和等于 1，列和等于 α。因此，最优支撑矩阵可以看作双随机矩阵的类似物。众所周知，双随机矩阵构成一个单纯形，其顶点是置换矩阵。最优支撑矩阵也构成一个单纯形。

Question 1.6. *What are the vertices of the simplex of optimal support matrices?*

基于定理 1.3，我们开发了一种简单、高效且可并行化的算法，用于寻找给定超图的最大密度子图（更一般地说，进行谱分解）。该算法是一个应用于支撑矩阵 A 的迭代过程，逐步最小化 $s_{\max}(A)$。即使支撑矩阵尚未达到最优， A 中与谱分解相对应的分块结构也会显现出来。因此，可以从近似最优的支撑矩阵中提取最大密度子图。同时，给定的支撑矩阵为子图的密度提供了一个上界。因此，我们不仅可以从支撑矩阵中提取一个高密度子图，还可以利用支撑矩阵证明所找到的子图已是最优。确实，由于密度可能取值为分母有界于数量的有理数。

对于顶点，只需构造一个密度为 α 且支撑矩阵为 A 的子图，使得 $s_{\max}(A)$低于超过 α 且具有可接受分母的下一个有理数。该算法使我们能够找到一个子图，并可证明其在包含数百万条超边的超图中具有最大密度。

在本文中，我们在加权超图的框架下开展研究，为每个顶点和每条超边赋予正权重。为便于阐述，我们在本引言中介绍了研究结果的非加权版本。

本文的结构如下。第2节给出主要定义，并陈述一个关于最大密度子图的关键定理。第3节给出这一关键结果的证明。第4节介绍超图的谱分解。第5节展示超图及其对偶的谱分解之间的关系，并在第6节说明谱分解是Dulmage-Mendelsohn分解的重要细化。最后，本文描述一种用于寻找给定超图的最大密度子图，以及更一般地求其谱分解的高效算法。

致谢。本研究部分得到了加拿大自然科学与工程研究委员会的资助。

2. Weighted hypergraphs and their support matrices

超图是一个三元组 $H = (V, E, \mathrm{supp})$，其中 $V = \{v_j \mid j = 1, \ldots, n\}$ 是顶点的集合，$E = \{e_i \mid i = 1, \ldots, m\}$ 是超边的集合，而到 V的幂集的支撑函数 $\mathrm{supp}: E \to \mathcal{P}(V)$表示哪些顶点属于给定的超边。在加权超图中，我们为每个顶点和超边赋予权重$\mathrm{wt}(v_j) = w_j > 0$，$\mathrm{wt}(e_i) = u_i > 0$。

For subsets $V' \subset V$, $E' \subset E$ we define their weights as

$$\mathrm{wt}(V') = \sum_{v \in V'} \mathrm{wt}(v), \quad \mathrm{wt}(E') = \sum_{e \in E'} \mathrm{wt}(e).$$

对偶超图 H^* 是三元组 $(V^*, E^*, \mathrm{supp}^*)$，其中 $V^* = E$ 和 $E^* = V$。超边 $e^* = v$ 在 H^* 中的支集定义为包含顶点 v 的 H 的超边集合，即，当且仅当 $e \in \mathrm{supp}^*(v)$ 时，$v \in \mathrm{supp}(e)$。H 中顶点的权重 w_j 会变为 H^* 中超边的权重，反之亦然。

我们将子超图 $H' \subset H$ 定义为一对子集 (V', E')，其中$V' \subset V$，$E' \subset E$，且对每条超边 $e \in E'$都有 $\mathrm{supp}(e) \subset V'$。$H'$ 的权重继承自 H。

我们将商超图 H'' 定义为一对子集（V'', E''），使得对于每个 $v \in V''$ 以及每条包含v的超边 $e \in E$，都有 $e \in E''$。H'' 上的超图结构由 $\mathrm{supp}_{H''}(e) = \mathrm{supp}_H(e) \cap V''$定义。

A quotient of a subgraph in H will be called a factor of H.
The following Lemma is elementary, and we omit its proof:

引理 2.1. (a) $H' = (V', E')$ 是 H 中的子图，当且仅当$H'' = (V \backslash V', E \backslash E')$ 是 H的商超图。(b) $H' = (V', E')$ 是 H 中的子图，当且仅当 (E', V') 是 H^* 的商超图。

Definition 2.2. *The density of a weighted hypergraph is defined as*

$$\alpha(H) = \frac{\sum\limits_{i=1}^{m} u_i}{\sum\limits_{j=1}^{n} w_j} = \frac{\mathrm{wt}(E)}{\mathrm{wt}(V)}.$$

Lemma 2.3. *Let H be a weighted hypergraph.*
(a) The dual hypergraph has density $\alpha(H^*) = \alpha(H)^{-1}$.

(b) 设 $H' = (V', E')$ 是 H 中的一个真子图，其商$H'' =$为$(V'' = V \backslash V', E'' = E \backslash E')$。$H$ 的密度是 H' 和 H''的密度的加权平均值:

$$\alpha(H) = \frac{\mathrm{wt}(V')}{\mathrm{wt}(V') + \mathrm{wt}(V'')}\alpha(H') + \frac{\mathrm{wt}(V'')}{\mathrm{wt}(V') + \mathrm{wt}(V'')}\alpha(H'').$$

(c)假设 H 没有子图 $H' \subset H$ 满足 $\alpha(H') > \alpha(H)$。那么 $\tilde{}$ $\tilde{}$其对偶超图没有子图 $H \subset H^*$ 满足密度 $\alpha(H) > \alpha(H)^{-1}$.

证明。(a) 和 (b) 部分的结论显然。我们用反证法证明 (c) 部分。若 H^* 中有一个子图的密度大于 $\alpha(H)^{-1}$，则 H 有一个密度小于 $\alpha(H)$ 的商超图。但根据 (b) 部分，H 中的补子图的密度将超过 $\alpha(H)$，这就产生了矛盾。□

本文旨在解决以下
最密子图问题：给定加权超图 H，求一个子图 H'，使其密度 $\alpha(H')$ 最大。

出于技术上的原因，以下我们假设每个顶点都属于某条超边，且每条超边的支撑集均非空。

解决上述问题的主要工具是支撑矩阵这一概念 A。

定义 2.4. 对于加权超图 H，支撑矩阵 A 是一个满足以下性质的 $m \times n$ 实矩阵：

(1) 对所有 $i = 1, \ldots, m$, $j = 1, \ldots, n$, $a_{ij} \geq 0$。(2) 如果顶点 v_j 不属于超边 e_i，那么 $a_{ij} = 0$。(3) 对所有 $i = 1, \ldots, m$, $\sum_{j=1}^{n} w_j a_{ij} = u_i$。

我们用 $\mathcal{S}(H)$表示 H 的所有支撑矩阵的集合。我们将通过在集合 $\mathcal{S}(H)$ 中运行优化过程，求解在 H 中寻找最优子图的问题。

我们指出，支撑矩阵告诉我们关于 H 中（隐藏）子图密度的一些信息。给定一个支撑矩阵 A，让我们用 s_j 表示列 j 中条目的总和：

$$s_j = \sum_{i=1}^{m} a_{ij}.$$

Set $s_{\max}(A)$ to be the maximum of the column sums:

$$s_{\max}(A) = \max_{j=1,\ldots,n} s_j.$$

命题 2.5。设 H 是一个加权超图，且 A是其任意一个支撑矩阵。则任意子图 $H' \subset H$ 的密度不超过 $s_{\max}(A)$：

$$\alpha(H') \leq s_{\max}(A).$$

证明。设 $H' = (V', E')$ 是 H 中的一个子图。让我们计算 A 中对应于 E' 的行中条目的加权和：

$$\sum_{i \in E'} \sum_{j=1}^{n} w_j a_{ij} = \sum_{i \in E'} \sum_{j \in V'} w_j a_{ij} = \sum_{i \in E'} u_i.$$

让我们也计算对应于 V' 的列中的加权和：

$$\sum_{j \in V'} w_j s_j = \sum_{j \in V'} \sum_{i=1}^{m} w_j a_{ij} \geq \sum_{j \in V'} \sum_{i \in E'} w_j a_{ij} = \sum_{i \in E'} u_i.$$

然后我们得到

$$s_{\max}(A) = \frac{\sum\limits_{j \in V'} w_j s_{\max}(A)}{\sum\limits_{j \in V'} w_j} \geq \frac{\sum\limits_{j \in V'} w_j s_j}{\sum\limits_{j \in V'} w_j} \geq \frac{\sum\limits_{i \in E'} u_i}{\sum\limits_{j \in V'} w_j} = \alpha(H').$$

□

下面陈述我们的主要结果：

定理 2.6. 令 H 为一个加权超图。则

$$\max_{H' \subset H} \alpha(H') = \min_{A \in \mathcal{S}(H)} s_{\max}(A).$$

3. Proof of the Main Theorem.

It follows from Proposition 2.5 that

$$\max_{H' \subset H} \alpha(H') \leq \min_{A \in \mathcal{S}(H)} s_{\max}(A).$$

The opposite inequality will follow from

命题3.1。(a)设 H 是一个密度为 α的加权超图，使得 H 中没有任何子图的密度超过 α。那么存在一个支撑矩阵 A，用于 H，使得每个列和都等于 α:

$$s_j(A) = \sum_{i=1}^{m} a_{ij} = \alpha \quad \text{for all } j = 1, \ldots, n.$$

(b) 设 H 是一个加权超图，且 H 中没有任何子图（包括 H 本身）的密度超过 α。那么存在一个支撑矩阵 A，使得 H 满足♀♀♀♀♀♀

$$s_j(A) = \sum_{i=1}^{m} a_{ij} \leq \alpha \quad \text{for all } j = 1, \ldots, n.$$

注3.2。如果我们成功构造出一个最优支撑矩阵 A，使 H 最小化 $s_{\max}(A)$，那么我们将能够在 H 中识别出一个具有最大密度的子图$H' = (V', E')$：通过取 A 中满足 $s_j = s_{\max}(A)$ 的列来构造集合 V'，并将集合E' 定义为

$$E' = \{e_i \in E \,|\, \exists v_j \in V' \quad a_{ij} \neq 0\}.$$

因此，可以从最优支撑矩阵，甚至是最优支撑矩阵的近似中，提取出 H 中的最优子图。更多详情请参见第 7 节。

Hu、Wu 和 Chan 证明了（[8], 引理 2.2，另见 [2], 引理 1），DSP 的一个解可以通过求解以下线性规划问题获得：

引理 3.3. 设 H 是一个加权超图。考虑一个关于变量 $p_i \geq 0, q_j \geq 0$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ 的线性优- 化问题，受约束 $p_i \leq q_j$每当顶点 v_j 属于超-n m边 e_i 且 $\sum_{i=1} w_j q_j = 1$，目标函数为 $\sum_{j=1} u_i p_i \to \max$.

则在这些约束下

$$\max \sum_{i=1}^{m} u_i p_i = \max_{H' \subset H} \alpha(H').$$

推论 3.4。设 H 是一个加权超图，使得 H 中没有子图（包括其自身 H ）的密度超过 α。设 $p_i \geq 0$、 $i = 1, \ldots, m$、 $q_j \geq 0$、 $j = 1$、...、 n satisfy $p_i \leq q_j$，只要顶点 v_j 属于超边 e_i。然后

$$\sum_{i=1}^{m} u_i p_i \leq \alpha \sum_{j=1}^{n} w_j q_j.$$

现在我们可以开始证明命题 3.1 了。我们将使用线性规划的方法。对于部分 (a)，我们将把寻找所需支撑矩阵 A 的任务视为一个以未知量 $\{a_{ij} \mid v_j \in \text{supp}(e_i)\}$ 为变量、具有约束 $a_{ij} \geq 0$ 的线性规划问题，

$$[e_i]: \quad \sum_{j=1}^{n} w_j a_{ij} = u_i, \quad i = 1, \ldots, m,$$

$$[v_j]: \quad \sum_{i=1}^{m} a_{ij} = \alpha, \quad j = 1, \ldots, n.$$

If a desired support matrix A does not exist then this linear programming problem is infeasible.

Farkas' 引理（[9], 第10.4节）关于不可行线性规划问题指出，在这种情况下，存在约束方程的一个线性组合

$$\sum_{i=1}^{m} f_i [e_i] + \sum_{j=1}^{n} g_j [v_j]$$

使得其中所有 a_{ij} 的系数均为非负数，但右端项为负数。由此得到不等式

$$w_j f_i + g_j \geq 0 \quad \text{whenever vertex } v_j \text{ belongs to hyperedge } e_i, \tag{1}$$

and

$$\sum_{i=1}^{m} f_i u_i + \alpha \sum_{j=1}^{n} g_j < 0. \tag{2}$$

我们将 V 和 E 根据 f_i 和 g_j的符号分成子集：

$$V' = \{j \in V \mid g_j \geq 0\}, \quad V'' = \{j \in V \mid g_j < 0\}.$$
$$E' = \{i \in E \mid f_i < 0\}, \quad E'' = \{i \in E \mid f_i \geq 0\},$$

条件(1)意味着，对于每个 $i \in E'$， $j \in V''$ 顶点 v_j 都不属于超边 e_i，而 (1) 对于 $i \in E''$、$j \in V'$ 显然成立。

然后我们得出结论，(V', E') 是 H 中的一个子图，而 (E'', V'') 是 H^* 中的一个子图。

设 $p_i = |f_i|$ 和 $q_j = \frac{|g_j|}{w_j}$。然后假设顶点 v_j 属于超边 e_i，条件 (1) 变为

$$q_j \geq p_i \geq 0 \text{ if } i \in E', j \in V',$$

$$p_i \geq q_j \geq 0 \text{ if } i \in E'', j \in V'',$$

而条件 (2) 变为

$$\left(\alpha \sum_{j \in V'} q_j w_j - \sum_{i \in E'} u_i p_i\right) + \left(\sum_{i \in E''} u_i p_i - \alpha \sum_{j \in V''} q_j w_j\right) < 0. \quad (3)$$

由于 (V', E') 及其子图的密度至多为 α，根据推论 3.4，我们有

$$\sum_{i \in E'} u_i p_i \leq \alpha \sum_{j \in V'} q_j w_j.$$

根据引理 2.3(c)，H^* 没有密度超过 α^{-1} 的子图。将推论 3.4 应用于 $(E'', V'') \subset H^*$，我们得到

$$\sum_{j \in V''} q_j w_j \leq \alpha^{-1} \sum_{i \in E''} u_i p_i.$$

Combining the last two inequalities, we get a contradiction to (3).

命题 3.1 的 (b) 部分的证明与 (a) 部分的证明思路相同。线性规划问题将被修改为在约束 $[v_j]$ 中使用不等式而不是等式，而约束 $[e_i]$ 将保持为等式。因此我们将得到一个额外的限制 $g_j \geq 0$，并且集合 V'' 将变为空。将推论 3.4 应用于子图 $(V', E') \subset H$ 我们将得到与 (2) 的矛盾，这意味着线性规划问题是可行的并且所需的支撑矩阵 A 存在。这完成了命题 3.1 的证明并且我们的主定理 2.6 随之成立。

注记3.5.设 $H' = (V', E')$ 是 H 中密度为 α 的子图，并设 A 是 H 的支撑矩阵，且对于每个顶点 $v_j \in V'$ 都有 $s_j = \alpha$。那么由命题 2.5 的证明可知，A 具有分块分解，其分块为 (V', E') 和 $(V \backslash V', E \backslash E')$，也就是说，当 $v_j \in V'$ 和 $e_i \notin E'$，或当 $v_j \notin V'$ 和 $e_i \in E'$ 时，我们必须有 $a_{ij} = 0$。

4. Spectral decomposition of a hypergraph

在本节中，我们将介绍最优支撑矩阵的特征值解释，并引入超图的谱分解。

我们回顾，实 $n \times n$ 矩阵 B 的范数定义为

$$\|B\| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Bx\|}{\|x\|}.$$

这等于 B的复特征值的最大范数。如果 B是对称且非负定的，则 B 的所有特征值均为实数且非负，因此 B 的范数等于其最大特征值。在这种情况下，范数也可以表示为

$$\|B\| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{(x, Bx)}{(x, x)}.$$

此外，如果 B 分解为 $B = Q^T Q$，对于某个 $m \times n$ 实矩阵Q，则

$$\|B\| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{(Qx, Qx)}{(x, x)}.$$

考虑权重 W 和 U的对角 $n \times n$ 和 $m \times m$ 矩阵，其中 W 对角线上的第 j个条目为 w_j， U 对角线上的第 i个条目为 u_i。设$W^{\frac{1}{2}}$ 和 $U^{\frac{1}{2}}$ 为对角条目分别为 $\sqrt{w_j}$ 和 $\sqrt{u_i}$ 的对角矩阵。

For a support matrix A define

$$\widetilde{A} = U^{-\frac{1}{2}} A W^{\frac{1}{2}}.$$

Theorem 4.1. *Let H be a weighted hypergraph. Then*

$$\max_{H' \subset H} \alpha(H') = \min_{A \in \mathcal{S}(H)} \|\widetilde{A}^T \widetilde{A}\|.$$

在证明这个定理之前，我们先介绍超图的谱分解。我们从一个初等引理开始：

引理 4.2. 设 H 是一个加权超图，其中不存在密度超过 α 的子图。设 $H_1 = (V_1, E_1)$ 和 $H_2 = (V_2, E_2)$ 是 H 中密度为 α 的两个子图。那么 $H_1 \cup H_2 = (V_1 \cup V_2, E_1 \cup E_2)$ 和$H_1 \cap H_2 = (V_1 \cap V_2, E_1 \cap E_2)$ 的密度也为 α。

证明。我们有

$$\begin{aligned} \mathrm{wt}(E_1 \cup E_2) &= \mathrm{wt}(E_1) + \mathrm{wt}(E_2) - \mathrm{wt}(E_1 \cap E_2) \\ &\geq \alpha \mathrm{wt}(V_1) + \alpha \mathrm{wt}(V_2) - \alpha \mathrm{wt}(V_1 \cap V_2) = \alpha \mathrm{wt}(V_1 \cup V_2). \end{aligned}$$

由于 H 中没有任何子图的密度超过 α，上述不等式实际上是一个等式，从而得到引理的结论。□

推论 4.3。设 $\alpha = \max_{H' \subset H} {}^{\alpha}(H')$。则在 H 中存在唯一的密度为 α 的极大子图。

定义 4.4。加权超图H 的谱分解是一种划分

$$V = V_{\alpha_1} \dot\cup V_{\alpha_2} \dot\cup \ldots \dot\cup V_{\alpha_k}, \quad E = E_{\alpha_1} \dot\cup E_{\alpha_2} \dot\cup \ldots \dot\cup E_{\alpha_k},$$

归纳定义为：$(V_{\alpha_t}, E_{\alpha_t})$ 是商图 $(V \setminus \overset{t-1}{\underset{r=1}{\cup}} V_{\alpha_r}, E \setminus \overset{t-1}{\underset{r=1}{\cup}} E_{\alpha_r})$ 中密度最大的极大子图 α_t，其中 $t = 1, \ldots, k$。

显然，对于谱分解，$\alpha_1 > \alpha_2 > \ldots > \alpha_k > 0$。

以下定理是定理 2.6 和注记 3.5 的直接推论：

定理4.5。设 $(V = \overset{k}{\underset{r=1}{\cup}} V_{\alpha_r}, E = \overset{k}{\underset{r=1}{\cup}} E_{\alpha_r})$是加权超图 H的谱分解。那么存在一个支撑矩阵 A，使得对于任意顶点 $v_j \in V_{\alpha_r}$，列和 $s_j(A)$ 等于 α_r。这样的矩阵将具有分块分解，其分块对应于谱分解 $(V_{\alpha_1}, E_{\alpha_1})$、$(V_{\alpha_2}, E_{\alpha_2})$ ..... $(V_{\alpha_k}, E_{\alpha_k})$ 的因子。

现在我们给出定理 4.1 的证明。我们首先证明，对于任意子图 $H' = (V', E') \subset H$ 以及任意支撑矩阵 $A \in \mathcal{S}(H)\tilde{}\ \tilde{}$，不等式 $\alpha(H') \leq \|A^T A\|$ 成立。构造向量 $x \in \mathbb{R}^n$，其

$$x_j = \begin{cases} \sqrt{w_j}, & \text{if } v_j \in V', \\ 0, & \text{otherwise.} \end{cases}$$

于是$(x, x) = \mathrm{wt}(V')$。让我们计算向量 $U^{-\frac{1}{2}} A W^{\frac{1}{2}} x$。我们有

$$(W^{\frac{1}{2}} x)_j = \begin{cases} w_j, & \text{if } v_j \in V', \\ 0, & \text{otherwise.} \end{cases}$$

如果 $e_i \in E'$ 那么 $\mathrm{supp}(e_i) \subset V'$。回顾支撑矩阵的定义性质，可得

$$(AW^{\frac{1}{2}} x)_i = \begin{cases} u_i, & \text{if } e_i \in E', \\ \text{non-negative}, & \text{otherwise,} \end{cases}$$

and

$$(U^{-\frac{1}{2}} A W^{\frac{1}{2}} x)_i = \begin{cases} \sqrt{u_i}, & \text{if } e_i \in E', \\ \text{non-negative}, & \text{otherwise,} \end{cases}$$

~ ~由此可见，$(Ax, Ax) \geq \mathrm{wt}(E')$。因此

$$\alpha(H') \leq \frac{(\widetilde{A}x, \widetilde{A}x)}{(x,x)} \leq \|\widetilde{A}^T\widetilde{A}\|.$$

Since A here is an arbitrary support matrix, and H' is an arbitrary subgraph, we get that

$$\max_{H' \subset H} \alpha(H') \leq \min_{A \in \mathcal{S}(H)} \|\widetilde{A}^T\widetilde{A}\|.$$

下面证明事实上这里是等式：对于一个最大密度子图，存在一个使等式成立的支撑矩阵。

令 $H_{\alpha_1} = (V_{\alpha_1}, E_{\alpha_1})$ 为 H 中密度最大的极大子图 α_1。考虑 H 的谱分解。对于每个因子 $(V_{\alpha_r}, E_{\alpha_r})$，我们可以构造一个支撑矩阵，使其所有列和都等于 α_r。通过将每个因子的支撑矩阵排列成分块，并在这些分块之外填入零，构造 A，用于 H。那么对于 $v_j \in V_{\alpha_r}$，列和 $s_j(A)$ 等于 α_r。上面我们已经为 $\widetilde{A}^T\widetilde{A}$ 构造了特征向量 $x^{(1)}, \ldots, x^{(k)}$，其特征值为 $\alpha_1, \ldots, \alpha_k$。每个向量 $x^{(r)}$ 在对应于 V_{α_r} 的位置上具有正分量，其他位置上均为零。

Matrix $\widetilde{A}$ 是对称且半正定的。因此，其特征值均为实数且非负。对应于~ ~不同特征值的特征向量彼此正交。此外，A^TA 的条目均非负。根据 Perron-Frobenius 定理，A^TA 有一个所有分量均非负的主特征向量（[4], 定理 XIII.3）。该向量不可能与所有向量$x^{(1)}$、……、$x^{(k)}$的点积都为零。因此，主特征值是 α_1、……、α_k之一，这意味着 A 的主特征值恰好是 α_1，即 H中子图的最大密度。这就证明了定理 4.1 的结论。

定理 4.5 给出的最优支撑矩阵的分块分解还可以进一步细化。考虑 H 的谱分解中的一个因子 $H_\alpha = (V_\alpha, E_\alpha)$。根据定义，$H_\alpha$ 没有密度超过 α 的子图。因此，由引理 4.2 可知，H_α 中密度为 α 的子图集合构成一个分配格。考虑该格中的因子（密度为 α 的子图，或两个嵌套且密度为 α 的子图之商）。如果一个因子没有真子图的密度为 α，我们就称它为简单因子。我们得到与该格的简单因子相对应的 V_α 和 E_α 的划分，而最优支撑矩阵将具有分块结构，其块与这些简单因子相对应。

5. Duality

本节将研究超图 H 与其对偶 H^* 的谱分解之间的关系。

定理 5.1. 设 $\{(V_1, E_1) \ldots, (V_k, E_k)\}$ 是加权超图 H 的谱分解，其密度为 $\alpha_1 > \ldots > \alpha_k > 0$。那么，$\{(E_k, V_k) \ldots, (E_1, V_1)\}$ 是其对偶加权超图 H^* 的谱分解，密度为 $\alpha_k^{-1} > \ldots > \alpha_1^{-1} > 0$。

为了证明此定理，我们将首先说明，最优支撑矩阵决定了超图的谱分解。

命题 5.2. 设 $\alpha_1 > \alpha_2 > \ldots > \alpha_k$。令

$$V = V_{\alpha_1} \dot\cup V_{\alpha_2} \dot\cup \ldots \dot\cup V_{\alpha_k}, \quad E = E_{\alpha_1} \dot\cup E_{\alpha_2} \dot\cup \ldots \dot\cup E_{\alpha_k} \tag{4}$$

为加权超图 H 的顶点集和超边集的划分，使得对每个 $t = 1, \ldots, k$, $(\overset{t}{\underset{r=1}{\cup}} V_{\alpha_r}, \overset{t}{\underset{r=1}{\cup}} E_{\alpha_r})$ 都是 H 中的一个子图，且 $\mathrm{wt}(E_r)/\mathrm{wt}(V_r) = \alpha_r$。设 A 是 H 的一个支撑矩阵，使得对每个顶点 $v_j \in V_{\alpha_r}$，列和 $s_j(A)$ 等于 α_r。那么，(4) 是 H 的谱分解。

证明。我们通过对 k 进行归纳来证明该命题。首先，我们来证明$(V_{\alpha_1}, E_{\alpha_1})$是 H中具有最大密度的极大子图。该子图的密度为 α_1，由命题 2.5 可知，α_1 是 H 中子图的最大密度。根据同一个命题，商超图 $H'' = (V \backslash V_{\alpha_1}, E \backslash E_{\alpha_1})$不包含密度超过 α_2的子图。这意味着$(V_{\alpha_1}, E_{\alpha_1})$是具有最大密度的极大子图。根据归纳假设，$(\overset{k}{\underset{r=2}{\cup}} V_{\alpha_r}, \overset{k}{\underset{r=2}{\cup}} E_{\alpha_r})$ 是商图H''的谱分解。命题 5.2 的结论现已得到。□

为了证明对偶定理 5.1，我们展示如何构造一个最优支撑矩阵对于对偶超图 H^* 从最优支撑矩阵 A 对于 H。设 A 是由定理 4.5 给出的最优支撑矩阵。

让我们为对偶超图 H^* 构造一个支撑矩阵。令B 为一个 $n \times m$ 矩阵，其条目

$$b_{ji} = \alpha_r^{-1} u_i^{-1} a_{ij} w_j,$$

其中 $v_j \in V_{\alpha_r}$ 是 H的谱分解。显然，B 的各项均为非负数；当顶点 e_i 不属于 H^*中的超边 v_j 时，$a_{ij} = 0$，因此 $b_{ji} = 0$。下面计算加权

B中的行和:

$$\sum_{i=1}^{m} u_i b_{ji} = \alpha_r^{-1} w_j \sum_{i=1}^{m} a_{ij} = w_j.$$

因此，矩阵 B 是 H^*的支撑矩阵。我们来计算 B的列和。由于 B 具有分块结构，对于 i，其中 $e_i \in E_{\alpha_r}$ 的所有非零元素 b_{ji} 在 B 的第 i 列中仅出现在满足 $v_j \in V_{\alpha_r}$的行 j 中。假设 $e_i \in E_{\alpha_r}$。那么

$$\sum_{j=1}^{n} b_{ji} = \alpha_r^{-1} u_i^{-1} \sum_{j=1}^{n} w_j a_{ij} = \alpha_r^{-1}.$$

现在可以看出，矩阵 B 满足命题 5.2 对于 H^*的条件。我们有 $\alpha_k^{-1} > \alpha_{k-1}^{-1} > \ldots > \alpha_1^{-1}$。$H$ 的商图对应于 H^*中的子图。因此，对于每个 t，$(\bigcup_{r=t}^{k} E_{\alpha_r}, \bigcup_{r=t}^{k} V_{\alpha_r})$ 是 H^*中的一个子图。我们有 $\mathrm{wt}(V_{\alpha_r})/\mathrm{wt}(E_{\alpha_r}) = \alpha_r^{-1}$，并且对于每个 $e_i \in E_{\alpha_r}$，B 中的列和等于 α_r^{-1}。因此，根据命题 5.2，H 的谱分解的对偶是 H^*的谱分解。

6. Refinement of the Dulmage-Mendelsohn decomposition

本节将证明，超图的谱分解是 Dulmage-Mendelsohn 分解的一种细化。本节中我们假设所有权重 u_i 和 w_j 均等于 1。

我们来回顾超图 H [3]的 Dulmage-Mendelsohn 分解的构造。该构造基于 H的极小外覆盖这一概念。

定义 6.1. 若一对子集 (V, E)，$V \subset V$，$E \subset E$ 满足：对于任意一对 $v \in V, e \in E$，只要 $v \in \mathrm{supp}(e)$，就有 $v \in V$ 或 $e \in E$，则称其为 H 的一个外覆盖。若外覆盖 (V, E) 在所有 H 的外覆盖中使 $|V| + |E|$ 达到最小，则称其为极小外覆盖。

令 $\mathcal{M}$ 为 H 的所有极小外部覆盖的集合。设

$$V^+ = \bigcap_{(\widehat{V},\widehat{E})\in\mathcal{M}} \widehat{V}, \quad V^- = \bigcap_{(\widehat{V},\widehat{E})\in\mathcal{M}} V\backslash\widehat{V}, \quad V^0 = V\backslash(V^+ \cup V^-),$$

$$E^+ = \bigcap_{(\widehat{V},\widehat{E})\in\mathcal{M}} E\backslash\widehat{E}, \quad E^- = \bigcap_{(\widehat{V},\widehat{E})\in\mathcal{M}} \widehat{E}, \quad E^0 = E\backslash(E^+ \cup E^-).$$

Dulmage 和 Mendelsohn 证明了 (V^+, E^+) 和 $(V^+ \cup V^0, E^+ \cup E^0)$ 是 H [3] 中的子图。

定理 6.2。设 $(V^+ \dot\cup V^0 \dot\cup V^-, E^+ \dot\cup E^0 \dot\cup E^-)$ 是超图 H 的 Dulmage-Mendelsohn 分解，并设 $\{(V_{\alpha_r}, E_{\alpha_r}) \mid r = 1, \dots, k\}$ 是 H 的谱分解。那么

$$V^+ = \bigcup_{\alpha>1} V_\alpha, \quad V^0 = V_1, \quad V^- = \bigcup_{\alpha<1} V_\alpha,$$
$$E^+ = \bigcup_{\alpha>1} E_\alpha, \quad E^0 = E_1, \quad E^- = \bigcup_{\alpha<1} E_\alpha.$$

Proof. First of all we point out that both

$$\Big(\bigcup_{\alpha>1} V_\alpha, \bigcup_{\alpha\le 1} E_\alpha\Big) \text{ and } \Big(\bigcup_{\alpha\ge 1} V_\alpha, \bigcup_{\alpha<1} E_\alpha\Big) \tag{5}$$

are exterior covers for H.

Let $\widehat{H} = (\widehat{V}, \widehat{E})$ 是 H 的一个极小外部覆盖。将 $H_\alpha = \widehat{H}(V_\alpha, E_\alpha)$ 视为 H 的商超图的一个子图。令 $V_\alpha = V \cap V_\alpha$, $E_\alpha = E \cap E_\alpha$。那么 (V_α, E_α) 是 H_α 的一个外部覆盖。我们希望证明，对于 $\alpha > 1$，必有 $V_\alpha = V_\alpha$, $E_\alpha = \varnothing$；而对于 $\alpha < 1$，必有 $V_\alpha = \varnothing$, $E_\alpha = E_\alpha$。

Fix $\alpha \ge 1$. Let $\widehat{E}'_\alpha = E_\alpha \setminus \widehat{E}_\alpha$。根据外部覆盖的定义，$E'_\alpha$ 中的超边必须支撑于 V_α中。因此，(V_α, E'_α)是 H_α中的一个子图。由于 H_α 中的子图的密度不超过 α，我们有 $|E'_\alpha| \le \alpha |V_\alpha|$。然后

$$|V_\alpha| = \frac{1}{\alpha}|E_\alpha| = \frac{1}{\alpha}(|\widehat{E}_\alpha| + |\widehat{E}'_\alpha|) \le \frac{1}{\alpha}|\widehat{E}_\alpha| + |\widehat{V}_\alpha| \le |\widehat{E}_\alpha| + |\widehat{V}_\alpha|.$$

由对偶性可知，对于 $\alpha \le 1$ 我们有 $|E_\alpha| \le |E_\alpha| + |V_\alpha|$。取所有 α的和，我们得到 $|V| + |E|$ 大于或等于外部覆盖的大小 (5)。因此，(5) 中的两个外部覆盖都是极小的。

从上述计算我们看到，对于任何最小外部覆盖 (V, E) 的 H 我们必须有 $E_\alpha = \varnothing$, $V_\alpha = V_\alpha$ 对于 $\alpha > 1$ 并且，通过对偶，$V_\alpha = \varnothing$, $E_\alpha = E_\alpha$ 对于 $\alpha < 1$。回忆Dulmage-Mendelsohn分解的定义，我们得到定理的论断。 □

7. Algorithm for finding a subgraph of the maximumdensity.

我们的方法是找到一个支撑矩阵 A，使 $\mathcal{S}(H)$ 的值最小（更确切地说，是找到一个支撑矩阵 A，使 $s_{\max}(A)$接近最优值），然后从该支撑矩阵中提取最优子图。我们将运行一个迭代过程，计算逐步逼近最优支撑矩阵的连续近似值，并在每次迭代中降低 $s_{\max}(A)$ 的值。

让我们介绍算法的步骤，然后详细讨论每个步骤。

Step 0. Initialize a support matrix A.
Step 1. For each row of the support matrix perform the Row Equalization Operation.
步骤 2。迭代步骤1。预计迭代次数（推测）为 log(|E|)，其中 |E| 是超边数（ A 的行数）。
步骤3。使用所得的支撑矩阵提取一个最大密度子图，更一般地，对 H进行谱分解。

Now let us give the details.
步骤 0。这里介绍一种合理的支撑矩阵初始化方法。

步骤 0.1。若 $v_j \in$supp(e_i)，则设为 $a_{ij} = \begin{cases} 1; \text{ 否则设为 } 0。 \end{cases}$

步骤 0.2。对于 A 的每一列 j，计算该列中 A 的元素之和 d_j（这将作为对应顶点的度）。将列 j 除以 d_j。

$_n$步骤 0.3。对于 A 的每一行 i，计算加权和 $v_i = \sum_{j=1} w_j a_{ij}.$ 将行 i 乘以 u_i/v_i。

步骤 0.3 确保 A 的每一行都有正确的加权和。我们的最终目标是最小化 A 中的最大列和。步骤 0.2 的目的是避免与高阶顶点对应的列产生巨大的列和。

步骤 1. 行均衡操作。考虑矩阵 A 的一行 $(a_{i1}, a_{i2}, \ldots, a_{in})$。

矩阵 A 的第 i 行对应于 H 中的超边 e_i，并且根据支撑矩阵的定义，对于每一个使得 v_j 不属于超边 e_i 的 j，我们必须有 $a_{ij} = 0$ 。我们将忽略所有此类条目。为简化记号，我们将假设超边 e_i 包含顶点 $\{v_1, v_2, \ldots, v_n\}$。因此在本步骤的描述中， n 不是 H 中顶点的总数，而是 e_i 中顶点的数量。

由于行均衡操作处理的是 A 的单行，在本步骤中我们将该行中的条目简单地记为 $(a_1, a_2, \ldots, a_n)$。我们还需要知道矩阵 A 的列和 $(s_1, s_2, \ldots, s_n)$。

行均衡操作的目标是以这样一种方式计算该行（$a'_1, a'_2, \ldots, a'_n$）的新值，使得新的列和（$s'_1, s'_2, \ldots, s'_n$）具备以下两个性质：

(1) 存在一个值 $\tilde{s}$ 使得 $s'_j = \tilde{s}$ 每当 $a'_j \neq 0$。

(2) 如果 $a'_j = 0$ 那么 $s'_j \geq \tilde{s}$。

Of course, while doing this, we must also preserve the condition

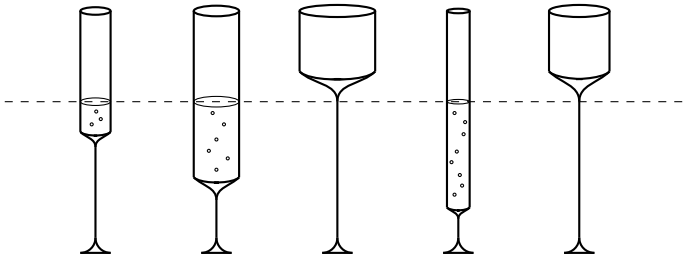
$$\sum_{j=1}^{n} w_j a_j = u_i.$$

注意，列和 s_j可写作 $s_j = a_j + b_j$，其中 b_j是所有行中列 j 的和，但不包括当前行。显然， b_j 的值不会因行均衡操作而改变，因此

$$b_j = s_j - a_j = s'_j - a'_j.$$

特别地，由(2)可知，我们将设定 $a'_j = 0$ 当且仅当$b_j \geq \tilde{s}$。确定 $\tilde{s}$的值是行均衡操作的关键部分。

让我们用三年级数学的风格重新表述行均衡问题。我们需要将香槟倒入形状各异的 n 个玻璃杯中。已知这些玻璃杯杯梗的长度为 b_1、.....、 b_n，以及横截面积为 w_1、.....、 w_n（我们假设所有玻璃杯都是圆柱形的）。我们需要将一瓶体积为 u_i 的香槟倒入这些玻璃杯，使所有装满的玻璃杯中的香槟液面 $\tilde{s}$ 保持相同，同时杯梗长度大于 $\tilde{s}$ 的玻璃杯保持空置。



$\tilde{s}$ 的值将通过逐步提高其下界的过程来确定。使用香槟模型的语言，我们将选择一个玻璃杯，并检验瓶中的香槟是否足够将各个玻璃杯装到所选玻璃杯底部的高度。如果成功，该高度将成为 b $\tilde{s}$的新下界。我们将记录已经倒出的香槟数量 u ，以及当前部分装满的玻璃杯的总横截面积 w 。我们用 B 表示目前尚未确定是否会被装满的玻璃杯集合。由于所有杯梗低于 b 的玻璃杯都将被装满，因此 B 中的所有元素都大于 b。步骤 1.0。用 $b \leftarrow 0$、 $u \leftarrow 0$、 $w \leftarrow 0$、 $B \leftarrow \{b_1, \ldots, b_n\}$进行初始化。

步骤1.1.选择一个元素 $b_j \in B$。将酒杯装满至液面 b_j 所需的香槟体积为

$$u' = u + w(b_j - b) + \sum_{\substack{b_k \in B \\ b_k \le b_j}} w_k(b_j - b_k).$$

其中， u 是已经倒入的体积， $w(b_j - b)$是将部分装满的玻璃杯中的液面从 b 提高到 b_j所需的额外量，最后一个总和表示将杯梗短于 b_j 的空玻璃杯装至液面 b_j所需的体积。

Step 1.2. If $u' \le u_i$ then set

$$u \leftarrow u', \quad b \leftarrow b_j, \quad w \leftarrow w + \sum_{\substack{b_k \in B \\ b_k \le b_j}} w_k$$

and remove from B all elements b_k with $b_k \le b_j$.

步骤 1.3。如果 $u' > u_i$ ，则从 B 中移除所有满足 $b_k \ge b_j$ 的元素 b_k 。

Step 1.4. Iterate Steps 1.1–1.3 until set B is empty.

步骤 1.5。现在我们知道，恰好满足$b_k \le b$的玻璃杯会被装满。现在，液面高度 $\tilde{s}$ 由剩余香槟的量决定：

$$\tilde{s} = b + (u_i - u)/w.$$

步骤 1.6。在支撑矩阵的当前行中设置新值：

$$a'_j = \begin{cases} \tilde{s} - b_j \text{ if } b_j \le b, \\ 0, \text{ otherwise.} \end{cases}$$

我们可以看到，步骤 1.1–1.3 的计算复杂度与 B的规模呈线性关系。平均而言，在步骤 1.2/1.3 中， B 的规模会呈指数级减小。因此，行均衡过程的计算复杂度平均而言与 n 呈线性关系，而在最坏情况下，复杂度为 n^2 （请注意，此处 n 是 e_i 支撑集的规模，而不是 V的规模）。

还可以给出一种更复杂的算法，使其在所有情况下都具有线性复杂度。该算法基于在集合B中寻找中位数，这可以通过应用中位数的中位数算法 [1]以线性复杂度完成。

记录每一行中对应于 b 的索引，以及 $\{b_1, \ldots, b_n\}$中的次高值，直到同一行的下一次行均衡操作，可能会很有用，因为在许多情况下这些索引不会发生变化；如果我们使用这两个索引开始步骤 1.1 的分析，就能加快计算速度。

作为示例，我们来介绍以下行均衡操作的例子。在此例中，我们令所有 $w_j = 1$ 和 $u_i = 1$。

j	1	2	3	4	5	6
a_j	0.3	0.1	0.3	0.15	0.1	0.05
s_j	1.8	1.5	1.3	0.9	1.0	1.2
b_j	1.5	1.4	1.0	0.75	0.9	1.15
a'_j	0	0	0.2	0.45	0.3	0.05
s'_j	1.5	1.4	1.2	1.2	1.2	1.2

Step 3. Let us describe how we can extract a subgraph from the computed support matrix A.

最优矩阵将具有这样的性质：最大列和 $s_{\max} = \alpha(H')$ 将出现在对应于最密子图 H' 的顶点的每一列中。实际上我们只会得到最优矩阵的一个近似，因此对应于 H' 的顶点的列中的列和将仅接近 $\alpha(H')$。因此我们将使用迭代过程构建顶点集 V 和超边集 E 用于子超图：

步骤3.0。以 $V_\alpha = \varnothing$, $E_\alpha = \varnothing$开始。确定 A 中具有最大列和的列，并将相应的顶点添加到集合V_α中。

步骤 3.1。对于上一步添加的每个新顶点 v_j，取 A 的对应列，并将每个带有 $a_{ij} \neq 0$ 的 i 所对应的超边 e_i 添加到 E_α 中（如果其中尚不存在该超边）。

步骤 3.2。对于上一步添加到 E_α 的每条新超边 e_i，如果这些顶点尚未在 V_α 中，则将其包含的所有顶点添加到 V_α 中。

迭代步骤 3.1-3.2，直到没有新的顶点添加到 V_α 中。

只需对最优支撑矩阵进行良好近似，即可得到最优子图。我们指出，任何支撑矩阵都能给出子图密度的上界，其中H

$$s_{\max}(A) \geq \alpha(H'),$$

因此，通过比较提取的子图的密度与计算出的支撑矩阵的最大列和，我们可以评估我们距离最优对有多远。在算法的迭代过程中，最大列和 $s_{\max}(A)$ 将减小，而子图的密度将增加，直到我们达到最优子图。

对于无权超图，如果我们达到 $s_{\max}(A) - \alpha(H') < 1/d|V|$，其中 d 是有理数 $\alpha(H')$ 的分母，这将证明 H' 是最优的。

上述算法可以通过将 A 的行子集分配给每个处理器来轻松并行化。这些处理器必须共享的唯一信息是列和向量。

总之，本文进一步发展了加权超图的密度理论，将密度最大化与支撑矩阵的谱问题和极小极大问题联系起来。尽管我们要解决的问题完全是组合性的，但所提出的解决方案基于连续优化。我们还提出了一种高效的迭代算法，用于在给定超图中寻找最大密度子图。我们算法的每次迭代都明显快于基于最大流的算法的一次迭代，从而使求解 DSP 扩展到更大的超图成为可能。

References

[1] M. Blum 等，《选择的时间界》，《计算机系统科学杂志》，7（1973），448-461。[2] M. Charikar，《在图中寻找稠密分量的贪心近似算法》，《组合优化中的近似算法：第三届国际研讨会，APPROX 2000》，德国萨尔布吕肯，2000年9月，会议论文集。Springer Berlin Heidelberg，2003年。[3] A.L. Dulmage，N. S. Mendelsohn，《二部图的覆盖》，《加拿大数学杂志》，10（1958），517-534。[4] F. R. Gantmacher，《矩阵理论》，第2卷，Chelsea Publishing Co.，纽约，1959年。[5] A.V.Goldberg，《寻找最大密度子图》，加利福尼亚大学伯克利分校，1984年。[6] A. V.Goldberg，S. Rao，《超越流分解障碍》，《美国计算机协会杂志》（JACM）45.5（1998）：783-797。[7] A.V. Goldberg，R. E. Tarjan，《高效的最大流算法》，《美国计算机协会通讯》，57（2014），82-89。[8] S.Hu，X.Wu 和T-H.H. Chan，《在演化超图中高效维护最稠密子集》，《2017年美国计算机协会信息与知识管理会议论文集》，2017年。[9] R. J. Vanderbei，《线性规划》，第2版，Kluwer Academic Publishers，波士顿，2001年。

School of Mathematics and Statistics, Carleton University, Ot-tawa, Canada电子邮件地址：billig@math.carleton.ca