通过成对观测实现近乎渐近最优的主动聚类

Rachel S. Y. Teo* P. N. Karthik† Ramya Korlakai Vinayak‡ Vincent Y. F. Tan*
*新加坡国立大学 † 印度理工学院海得拉巴分校
‡ 威斯康星大学麦迪逊分校 电子邮件：rachel.tsy@u.nus.edu pnkarthik@ai.iith.ac.in
ramya@ece.wisc.edu vtan@nus.edu.sg

arXiv:2602.05690v1 [cs.LG] 5 Feb 2026

摘要

我们提出了一种新的分析框架，利用带噪声且主动收集的响应，将 M 个项目聚类为数量未知的 K 个不同组。在每个时间步，智能体都可以查询项目对并观测老虎机式二元反馈。如果项目对属于同一簇（分别属于不同簇），则观测到的反馈为 1 的概率为$p > 1/2$ （分别为 $q < 1/2$）。借助普遍适用的测度变换技术，我们建立了一个基本下界，刻画达到目标聚类准确率所需查询次数的期望值，并将其表述为一个上确界—下确界优化问题。在这一理论基础上，我们设计了一种渐近最优算法，其停止准则是将内层下确界的经验版本——广义似然比统计量（GLR）——与阈值进行比较。我们开发了一个计算上可行的 GLR 统计量变体，并证明，其与下界之间的性能差距可以通过经验准确估计，且始终不超过下界的一个常数倍。

一、引言

聚类是无监督学习中的一项基本任务，旨在将一组项目划分为若干组，使同一组内的项目彼此之间比与不同组中的项目更加相似 [1]–[4]。传统方法，如 K-均值 [5], 、层次聚类 [6], 和谱方法 [7] ，通常依赖明确的相似度度量来发现簇结构。另一种模型——借助含噪预言机进行聚类——由 [8], 提出，利用众包来揭示数据中的潜在结构。

在该模型中，给定一组 M 未标记项目，需要将其聚为 K 个簇，其中 K 未知。有一个预言机，在收到对两个项目的查询时，会指出它们是否属于同一簇。不过，答案有一定概率出错。该模型的动机在于其实际应用价值，例如众包实体解析 [9]–[11], 、社会网络中的符号边预测 [12]–[15] 以及相关聚类 [16]–[18]。

带噪查询聚类与随机块模型（SBM）密切相关。随机块模型是一种广受认可的框架，用于分析图聚类和社区检测问题 [19]–[22]。在随机块模型中，一组 M 顶点（项目）被划分为 K 个隐藏分区，并根据这些顶点所属的分区生成随机图。如果两个顶点属于同一簇，则它们之间以概率 p 形成一条边；如果不属于同一簇，则以概率 q 形成一条边，其中 $q < p$。目标是从按照随机块模型采样得到的随机图中恢复底层簇。

然而，我们的框架在几个关键方面不同于随机块模型研究。首先，对于每一对项目，我们会获取多个自适应采样的查询；而随机块模型仅提供图连通关系的单个快照 ，反映项目之间的关系。其次，类似于固定置信度最佳臂识别 [23], ，我们研究真实聚类恢复错误概率趋近于零的渐近情形。这种情形适用于高保证度的实体解析任务，例如跨医院信息系统进行患者记录链接：在此， M 表示患者记录的（有限）集合，成对查询对应于成本高昂的人工核查或专家核查，而错误的合并（将两名患者混为一人）或拆分 （将一名患者拆分为多个身份）可能导致严重的后续临床后果。因此，系统要求通常会规定严格的目标错误概率水平，这使得错误概率消失的渐近分析比主动聚类中传统的令 $M \to \infty$ 趋于无穷更为相关。最后，随机块模型中的概率 $p = p_M$和 $q = q_M$是 M的函数，通常按 $\Theta\left(\frac{\log M}{M}\right)$的尺度变化。然而，在我们的情境中， p和 q 被视为固定但未知的常数。

贡献。 我们提出了一种具有理论依据的算法，在错误概率趋于零的情形下，最小化预期查询次数。这一设定与 [24] 密切相关。我们的主要贡献如下：

- 我们将含噪预言机聚类问题建模为一个具有赌博机反馈的问题，其中每个臂代表一个待查询的项目对。我们定义了该问题实例的一种表达形式，其特征由 M 个项目的唯一划分以及预言机反馈概率 $0 \le q < 1/2 < p \le 1$ 决定。

- 我们推导出了期望样本复杂度的渐近、依赖于实例的下界，并将其表示为加权二元KL散度上确界—下确界优化，从而刻画实例相对于其最接近替代方案的难度。利用问题的结构，我们大幅缩减了替代方案的搜索空间，从而能够设计出一种有原则的采样规则。

• 我们开发了一种渐近最优的算法，尽管其在统计意义上是最优的，但依赖于计算上不可行的停止规则。对此，我们提出了一种实用且计算可行的变体；尽管其性能并非最优，但其偏离最优性的程度（即渐近下界）具有乘法上界，同时还具有一种经验近似，并且该近似会收敛于理论上界。

相关工作。 Mazumdar 和 Saha [8] 开创了带噪查询聚类的研究，提出了一种计算上难以处理的算法，其样本复杂度接近他们的信息论下界。他们还提出了一种更高效但次优的算法，其样本复杂度相对于簇数量呈二次增长。在此基础上， [25] 提出了以固定错误概率恢复大小为 $\Omega(K \log M)$ 的簇的算法，而 [26] 改进了双簇情形的方法，[27] 则降低了对 $p-q$ 的依赖。最近， [28] 建立了新的下界，并通过与多臂老虎机建立联系，提出了一种多项式时间算法，其所达到的上界仅相差一个加性项，尽管采用了形成子簇的不同方法。我们的工作针对 p 和 q 未知的实际场景，并允许重复查询。与此相关， [24] 基于多臂老虎机的迭代对数律（迭代对数律）提出了一种计算高效的主动聚类算法，以接近最优的样本复杂度保证高概率恢复。以往采用被动策略的工作对小簇效果有限，转而关注所有项目对，或随机或精心选取的项目对子集 [29]–[32]。

我们的方法基于固定置信度纯探索中的技术，尤其是那些在错误概率趋于零时实现渐近最优性 [23], [33]的技术。与 [34],类似，我们将在线聚类任务表述为一个基于成对比较的纯探索问题，类似于对决老虎机 [35]。一些研究关注用于遗憾最小化的多臂老虎机算法中的聚类结构 [36]–[40], ，而我们着重于根据成对反馈揭示臂的实际划分。

II. 问题设定与预备知识

考虑 M 个项目，每个项目属于 $K \in [M]$ 个簇中的一个，其中 $[M] := \{1, \dots, M\}$。其簇分配由 $c_i \in [K]$ 指定，适用于 $i \in [M]$。对于 $\ell \in [K]$，我们定义 $g_\ell := \{i \in [M] : \ c_i = \ell\}$ 为簇 ℓ 中的项目集合。我们假设簇的数量 K 事先未知。令 $I = \{(i, \ j) \in [M] \times [M] : i < j\}$ 为一个指标集，表示所有可能的不同项目对。我们用 $\mathcal{P}_I$ 表示 I 上的分布集合。

在每个时刻 $t \in \mathbb{N}$，我们查询项目 $\mathbf{I}_t = i_t$ 和 $\mathbf{J}_t = j_t$ 是否属于同一个簇，并获得观测值 $\mathbf{Y}^{i_t j_t}$ ，其定义如下

$$\mathbf{Y}_{i_t j_t} \sim v_{i_t j_t} = \begin{cases} \mathrm{Ber}(p), & \text{if } c_{i_t} = c_{j_t}, \\ \mathrm{Ber}(q), & \text{if } c_{i_t} \neq c_{j_t}, \end{cases} \tag{1}$$

对于 $0 \le q < 1/2 < p \le 1$。也就是说，查询任意项目对所获得的观测值 $\mathbf{Y}^{i_t j_t}$ 服从参数为 p （分别为 q）的伯努利分布，前提是这两个项目属于同一个簇（分别属于不同的簇）。假设参数 p 和 q 未知。我们用 $N_{ij}(t)$ 表示截至时刻 t，针对项目对 $(i, \ j) \in I$ 的查询所获得的反馈观测值数量。

令 $\mathcal{S}^M = \{C \in \mathbb{R}^{M \times M} \ C^\top = C\}$ M M $\mathcal{C} = \{C \in \mathcal{S}^M \ \forall\, i\ j \in$: : 表示大小为 × 的对称矩阵集合，且 : (,) I ， $c_{ij} = \mathbb{1}\{c_i = c_j\}p + \mathbb{1}\{c_i \neq c_j\}q$，其中 $0 \le q < 1/2 < p \le 1\}$。该集合刻画了所有问题实例，由项目 M 的所有可能聚类以及 p 和 q 的不同取值组成。

对于上述众包聚类问题，我们的目标是根据观测值找出项目的正确聚类。我们研究固定置信度设定：给定置信水平 $\delta \in (0, \ 1)$，我们希望设计一种算法，使其（近乎）最小化识别正确聚类所需的期望时间步数，记为 $\mathbb{E}[\tau_\delta]$，并以概率 $\ge 1 - \delta$ （ δ-正确算法）成功。具体而言，我们旨在设计并分析渐近最优性几乎达到最优的算法，其含义将在下文中精确定义。

等价关系：设 $C_=$ 和 $C_{\neq}$ 为 $\{0, 1\}$值矩阵，使得 $(c_=)_{ij} = \mathbb{1}\{c_i = c_j\}$ 且 $(c_{\neq})_{ij} = \mathbb{1}\{c_i \neq c_j\}$。定义 $\sim$ 上的关系 $\mathcal{C}$ 如下：对于 C、$C' \in \mathcal{C}$，若 $C_= = C'_=$ 且 $C_{\neq} = C'_{\neq}$，则令 $C \sim C'$。容易验证， $\sim$ 是一个等价关系。该关系将所有可能实例的集合划分为 $\mathcal{C} = \bigcup_{K=1}^{M} \{C' \in \mathcal{C} : \ \exists C \in \mathcal{C}$ ，其中包含 K 个簇，以及 $C' \sim C\}$: $= \bigcup_{K \le M} \mathcal{C}^K$，其中 $\mathcal{C}^K$ 表示恰好含有 K 个簇的所有实例的集合。若 C 具有 K 个簇，即 C 表示恰好包含 K 个不同簇的簇分配，我们就称其具有 K 个簇。每个划分 $\mathcal{C}^K$ 由其所含各实例表示的簇数 K 确定。令 $\mathcal{C}/\sim$ 表示所有等价类的集合，并令 $[C] \in \mathcal{C}/\sim$ 表示与实例 C 相关联的等价类。定义 f: $\widetilde{\mathcal{C}} \to \mathscr{P}(\mathscr{P}(M))$ 为 $f([C]) = \{g_1, g_2, \dots, g_K\}$，其中 $C \in \mathcal{C}^K$，而 g_1、 g_2、.....、 g_K 表示 C 中的簇分配， $\mathscr{P}(M)$ 表示 $[M]$ 的幂集。注意，形如 $\{g_1, \dots, g_K\}$ 的簇分配是 $\mathscr{P}(\mathscr{P}(M))$ 的一个元素。为便于理解，我们在附录 A 中给出一个例子。

记号： $\mathcal{F}_t = \sigma(\{(\mathbf{I}_s, \ \mathbf{J}_s), \ \mathbf{Y}_{\mathbf{I}_s \mathbf{J}_s}\}_{s=1}^t)$ 表示由截至时间 t 的观测值生成的 σ-代数。设 τ_δ 是关于滤过 $\{\mathcal{F}_t\}_{t \in \mathbb{N}}$ 的一个停止时间。

III. 下界

我们为主动众包聚类问题中的期望样本复杂度 $\mathbb{E}_C[\tau_\delta]$ 建立了一个依赖于实例的下界。对于一个实例 $C \in \mathcal{C}$，令

$$\mathrm{Alt}(C) \coloneqq \bigcup_{[C']\in\widetilde{\mathcal{C}}:\ [C']\neq[C]} [C']$$

表示与 C 的聚类结果不同、构成错误聚类的备选实例集合。

定理 1. 对于置信水平 $\delta \in (0,\ 1)$ 和实例 $C \in \mathcal{C}$，任意 δ-正确算法都满足

$$\mathbb{E}_C[\tau_\delta]\ D^*(C)\ \geq\ \mathrm{KL}\big(\mathrm{Ber}(\delta), \mathrm{Ber}(1-\delta)\big),$$

其中

$$D^*(C) \coloneqq \sup_{\lambda\in\mathcal{P}_I} \inf_{C'\in\mathrm{Alt}(C)} \sum_{(i,j)\in I} \lambda_{ij}\mathrm{KL}(v_{ij}, v'_{ij}), \tag{2}$$

且 v'_{ij} 是分布 ，对应于 C'，与(1) 类似（参数相应为 $p' > 1/2 > q'$）。此外 *re,*

$$\liminf_{\delta\to 0} \frac{\mathbb{E}_C[\tau_\delta]}{\ln(1/\delta)} \geq D^*(C)^{-1}. \tag{3}$$

定理 1 的证明见附录 B-A。

注释 1。 我们通过一个简单示例，依据 $\Delta = p - q$ 和 K 计算期望样本复杂度。考虑这样一个实例：簇的数量为 $K = M/2$, 项目的数量（为偶数）为 $M > 4$，且每个簇恰好包含两个项目，即对所有 i， $|g_i| = 2$ 。显式计算 $D^*(C)$ 后，我们得到

$$\liminf_{\delta\to 0} \frac{\mathbb{E}_C[\tau_\delta]}{\ln(1/\delta)} \geq \underbrace{\left[\Delta \log \frac{1+\Delta}{1-\Delta}\right]^{-1}}_{\Theta(1/\Delta^2)\ as\ \Delta\to 0} \underbrace{\frac{K^4 - 5K^2 + 4K}{K^3 - K^2 - 2}}_{\Theta(K)\ as\ K\to\infty}. \tag{4}$$

我们强调，用于推导我们界的框架不同于文献中采用的框架。在这里，成对查询的期望数量 $\mathbb{E}_C[\tau_\delta]$ 按 $\ln(1/\delta)$归一化，且 $\delta \to 0$ 见式 (4)。这种方法类似于关于固定置信度最佳臂识别的研究 *[23]*。相比之下，其他研究（如 *[8]* 和 *[24]* ）在未经归一化的情况下推导 τ_δ 的高概率界，其结果通常按 $\tau_\delta = \widetilde{O}(MK/\Delta^2)$的量级增长。在这些分析中，失败概率也被设定为随 (M, K)呈多项式衰减。因此，尽管我们在式 (4) 中的界对 Δ 的依赖与以往研究一致，但对 K 的依赖似乎不那么一致。例如，在我们当前的例子中， $MK = \Theta(K^2)$。尽管如此，由于归一化因子 $\ln(1/\delta)$ 以及 $\delta \to 0$时的极限，无法进行直接比较。

(2) 中的量 $D^*(C)$ 刻画了识别 C 中聚类的“难度”，它是 C 所对应的伯努利分布（见 (1)）与其最易混淆的备选 C' 的伯努利分布之间 KL 散度的函数。根据定义，$D^*(C)$ 是一个上确界 - 下确界问题的解，而 $D^*(C)^{-1}$ 是依赖于实例的下界。求解 $D^*(C)$ 可得 $\Lambda^*(C)$，即所有对上的最优分配的集合。在每个时间步按照任意 $\lambda^* \in \Lambda^*(C)$ 进行采样，可提供信息量最大的待查询对选择，从而达到 (3) 中的样本复杂度下界，并有效地表明其紧性为 $\delta \to 0$。我们将使用 λ^* 来指导第 IV 节中算法的采样规则，以提供信息量最大的待查询对选择，从而达到 (3) 中的样本复杂度下界。

为此，我们需要求出式（2）中 $D^*(C)$ 表达式所涉及的 $\mathrm{Alt}(C)$ 上的下确界。然而，这在计算上代价高昂，因为 $\mathrm{Alt}(C)$ 中存在大量不同的等价类 $[C']$，每个等价类都由 M 个项目的一种独特聚类所决定。等价类的总数也称为贝尔数 [41], , 其随 M 超指数增长。此外，每个等价类中还存在由满足 $0 \leq q' < 1/2 < p' \leq 1$ 的 $(p',\ q')$ 参数化的不可数多个替代实例 C' 。这给求解 $\mathrm{Alt}(C)$ 上的下确界并求得 $\Lambda^*(C)$ 的闭式表达式带来了实际挑战。我们推导出 $D^*(C)$ 中目标函数的简化表达式，并证明 $\mathrm{Alt}(C)$ 上的下确界等于较小集合 $\mathrm{min}(C)$ 上的下确界。利用这些见解，我们求得 $\Lambda^*(C)$。

下面，我们将更详细地介绍我们的简化技术。对于固定的 C 和 $C' \in$Alt(C)，le t

$$\begin{aligned}
N_p &\coloneqq \{(i,j) \in I : c_i = c_j\} \subset I,\\
N_q &\coloneqq \{(i,j) \in I : c_i \neq c_j\} \subset I,\\
N_1 &\coloneqq \{(i,j) \in I : c_i = c_j \text{ and } c'_i \neq c'_j\} \subset N_p,\\
N_2 &\coloneqq \{(i,j) \in I : c_i \neq c_j \text{ and } c'_i = c'_j\} \subset N_q.
\end{aligned}$$

这里， N_p （分别为 N_q）表示在 C 中属于同一簇（分别属于不同簇）的项目对，而 N_1 （分别为N_2）表示在 C 中属于同一簇（分别属于不同簇），但在 C' 中属于不同簇（分别属于同一簇）的项目对。为简化记号，我们还引入映射 Λ_{N_1}， $\Lambda_{N_2}:\mathrm{Alt}(C) \to \mathscr{P}(I)$，使得 $\lambda_{N_1}(C') = N_1$且 $\lambda_{N_2}(C') = N_2$。注意

$$\begin{aligned}
&\mathrm{KL}(v_{ij}, v'_{ij}) = d(p,q') \quad \forall (i,j) \in N_1,\\
&\mathrm{KL}(v_{ij}, v'_{ij}) = d(q,p') \quad \forall (i,j) \in N_2,\\
&\mathrm{KL}(v_{ij}, v'_{ij}) = d(p,p') \quad \forall (i,j) \in N_p \setminus N_1,\\
&\mathrm{KL}(v_{ij}, v'_{ij}) = d(q,q') \quad \forall (i,j) \in N_q \setminus N_2,
\end{aligned}$$

其中， p' 和 q' 是 C'的伯努利参数。利用上述记号，我们有

$$D^*(C) = \sup_{\lambda \in \mathcal{P}_I} \inf_{C' \in \mathrm{Alt}(C)} h(\lambda, C'), \tag{5}$$

其中， $h(\lambda, C')$ 定义为总和

$$\begin{aligned}
h(\lambda, C') = {} & d(p,q') \sum_{(i,j) \in \Lambda_{N_1}(C')} \lambda_{ij} + d(q,q') \sum_{(i,j) \in N_q \setminus \Lambda_{N_2}(C')} \lambda_{ij}\\
& + d(q,p') \sum_{(i,j) \in \Lambda_{N_2}(C')} \lambda_{ij} + d(p,p') \sum_{(i,j) \in N_p \setminus \Lambda_{N_1}(C')} \lambda_{ij}.
\end{aligned}$$

定理 2. 固定 $M > K > 1$。对于 $C \in \mathcal{C}^K$，令 $\min(C) := \left(\mathcal{C}^{K-1} \cup \mathcal{C}^{K+1}\right) \cap \mathrm{Alt}(C)_1$，，其中 $\mathrm{Alt}(C)_1$ 是替代实例的集合，这些实例的聚类可由 C 的聚类通过以下方式之一得到：将 C 中的一个簇拆分为两个簇，或将 C 中的两个簇合并为一个簇。然后，我们可以将式 (5) 中的 $\mathrm{Alt}(C)$ 集合缩减为一个小得多的子集 $\min(C)$

$$D^*(C) = \sup_{\lambda \in \mathcal{P}_I} \inf_{C' \in \mathrm{Alt}(C)} h(\lambda, C') = \sup_{\lambda \in \mathcal{P}_I} \inf_{C' \in \min(C)} h(\lambda, C'). \tag{6}$$

定理 2 的证明见附录 B-B。

定理 2 表明，如果 C有 K个簇，那么与 C最接近的替代实例必定包含 $K+1$个或 $K-1$个簇；它由将 C 中的两个簇合并为一个，或将 C 中的一个簇拆分为两个而得到。固定 $\lambda \in \mathcal{P}_I$后，可以根据 λ 以及支配 C的参数 (p, q)，在 $\min(C)$ 的每个等价类中求得最优的 (p', q')。随后，对 $\min(C)$ 的有限个等价类取最小值，即可得到 $D^*(C)$。

尽管进行了上述简化，一般情况下， $\Lambda^*(C)$ 不存在闭式解，只有在特殊情形下例外：$K = 1$ （所有项目都属于同一个簇）或 $K = M$ （每个簇都只包含一个项目）。在这些特殊情形下，最优分配是唯一的（即 $\Lambda^*(C)$ 是单元素集合），并且等于在 I 上的均匀分布，正如下述结果所述。

命题 1. 当 $C \in \mathcal{C}^K$， $K = 1$ 或 $K = M$时，令

$$\Lambda^*(C) = \arg\max_{\lambda \in \mathcal{P}_I} \inf_{C' \in \mathrm{Alt}(C)} h(\lambda, C').$$

则， $\Lambda^*(C)$ 是 I 上的均匀分布。

命题 1 的证明见附录 B-C。

由命题 1 可知，当所有项目都属于同一个簇，或每个项目各自构成一个簇时，最优采样规则应当是在所有项目对上进行均匀采样。这是因为在这种情况下，每一对 $(i, j) \in I$ 都会产生参数为 p （分别为 q）的伯努利观测，因此每一对都具有相同的信息量。

IV. 在相差一个常数的意义下达到下界

(3) 中的下界表明，在每个时间步按照分布 $\lambda^* \in \Lambda^*(C)$ 对项目进行采样的任何算法，都能在 $\delta \to 0$ 时达到该下界。然而，由于事先不知道底层实例 C ，必须在过程中对其进行估计。正如下文所述，关键问题在于，估计得到的实例未必属于所有可行实例的集合 $\mathcal{C}$ 。我们精心设计了一种算法：在每个时间步，将估计得到的实例投影到 $\mathcal{C}$ 上，利用该投影对一对项目进行采样，并持续执行，直到对底层聚类有足够的把握。我们对该算法进行了改进，以解决计算瓶颈，并设计了一种计算高效变体，使其几乎达到渐近最优。

算法1 Proj$_{\mathcal{C}}(\hat{C}_t)$

输入： C_t1：对所有 $(i,\ j)$，设置 $(\hat{c}_{t_=})_{ij} = \mathbb{1}\{(\hat{c}_t)_{ij} \geq 0.5\}$ 和 $(\hat{c}_{t_{\neq}})_{ij} = \mathbb{1}\{(\hat{c}_t)_{ij} < 0.5\}$ 2：求解

$$[C_t] = \arg\min_{[C']\in\widetilde{C}} \sum_{(i,j)\in I} |(\hat{c}_{t_=})_{ij} - (c'_=)_{ij}| + |(\hat{c}_{t_{\neq}})_{ij} - (c'_{\neq})_{ij}|. \tag{7}$$

3：通过 p_t 和 q_t 获得

$$p_t = \frac{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_=})_{ij} = 1\}(\hat{c}_t)_{ij} \cdot N_{ij}(t)}{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_=})_{ij} = 1\}N_{ij}(t)}, \quad q_t = \frac{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_{\neq}})_{ij} = 1\}(\hat{c}_t)_{ij} \cdot N_{ij}(t)}{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_{\neq}})_{ij} = 1\}N_{ij}(t)}.$$

4: 设置 $C_t = p_t C_{t_=} + q_t C_{t_{\neq}} \in \mathcal{C}$

5: **输出：** C_t

A. 基于修正 *D-Tracking* 的采样规则

现在，我们设计一种采样规则，该规则利用对未知实例 C 的估计，来建议在每个时间步查询信息量最大的两个项目。令 C_t 表示时间 t 时的经验实例，其元素定义如下:

$$(\widehat{c}_t)_{ji} = (\widehat{c}_t)_{ij} = \frac{1}{N_{ij}(t)} \sum_{s\leq t} \mathbf{Y}_{i_s j_s} \mathbb{1}\{I_s = i, J_s = j\}.$$

一种简单的方法是使用 C_t 作为 C 的代理，以求解 $D^*(C)$并获得 $\Lambda^*(C_t)$，随后针对某个 $\lambda^* \in \Lambda^*(C_t)$采样$(i,j) \sim \lambda^*$。若 $C_t \notin \mathcal{C}$，这种方法就会失效。回想一下，任何 $C \in \mathcal{C}$ 都必须满足：对于所有属于同一簇的 i、j、$k \in [M]$，均有 $c_{ji} = c_{ki}$。然而，$(\widehat{c}_t)_{ji}$和$(\widehat{c}_t)_{ki}$可能并不相同。因此，直接将 C_t 代入(6)并不合适。为此，我们构造了一个子程序 Proj$_{\mathcal{C}}$，详见算法1，用于找出 $\mathcal{C}$ 中与 C_t“最接近”的元素，即将 C_t 投影到 $\mathcal{C}$上。令 $C_t :=$Proj$_{\mathcal{C}}(C_t)$。受 Russo 和 Proutiere [42], wor $\sigma > 0$的启发，我们令

$$\lambda^*(\sigma; C_t) := \arg\max_{\lambda\in\mathcal{P}_I} \min_{C'\in\min(C_t)} h(\lambda, C') - \frac{\sigma}{2}\|\lambda\|^2.$$

由于正则化项 $\|\cdot\|^2$ 的严格凸性，最优分配 $\lambda^*(\sigma; C_t)$ 是唯一的。我们建议从 $\lambda^*(\sigma;\ C_t)$ 与 I 上的均匀分布的凸混合中采样，以确保当 $t \to \infty$ 时，$N(t)/t$ 收敛到一个严格为正的分布。对于 $\epsilon \in (0,1)$，令我们的混合分布为

$$\lambda^*_\epsilon(\sigma; C_t) := (1-\epsilon)\lambda^*(\sigma; C_t) + \epsilon \cdot u, \tag{8}$$

其中 u 是 I 上的均匀分布，即对所有 $(i,j) \in I$，有 $u_{ij} = 1/|I|$。我们使用 [23] 中的原始 *D-Tracking* 规则进行采样。引入 $U_t = \{(i,j) \in I:\ N_{ij}(t) < (\sqrt{t} - \binom{M}{2}/2)_+\}$，我们将项目 $(\mathbf{I}_{t+1}, \mathbf{J}_{t+1})$ 采样为

$$(\mathbf{I}_{t+1}, \mathbf{J}_{t+1}) = \begin{cases} \arg\min_{(i,j)\in U_t} N_{ij}(t) & U_t \neq \emptyset \\ \arg\max_{(i,j)\in I} t\lambda^*_\epsilon(\sigma; C_t)_{ij} - N_{ij}(t) & U_t = \emptyset \end{cases}. \tag{9}$$

命题 2. 在采样规则 (9) 下以及实例 $C \in \mathcal{C}$ 下，$N(t)/t$ 几乎必然收敛到 $\lambda^*_\epsilon(\sigma; C)$，即

$$\Pr\left(\lim_{t\to\infty} \frac{N(t)}{t} = \lambda^*_\epsilon(\sigma; C)\right) = 1,$$

其中 $N(t) = [N_{ij}(t) : (i,\ j) \in I]$，$\epsilon \in (0,\ 1)$ 且 $\sigma > 0$。

命题 2 的证明见附录 C-A。

B. 基于广义似然比的停止规则

在上一节建立采样规则后，我们现在概述一个停止准则，以确定算法何时应停止，并输出一个错误概率不超过 δ 的项目聚类。一个自然的停止规则是采用许多现有纯探索问题研究所使用的广义似然比（ GLR ）检验 [33], [43], [44]。我们证明，该停止规则是 δ-正确的，并且在 趋于极限时渐近最优。

然而，该统计量难以计算，因此难以按原始形式实现这一停止规则。为了解决这一问题，我们在第 IV-C 节提出另一种计算高效的停止规则，该规则实现了近乎渐近最优性，并且与下界之间的差距得到明确控制。

现在，我们介绍用于检验 C_t 及其替代实例的广义似然比统计量，以及 [45]中建议的阈值。广义似然比统计量为

$$Z(t) \coloneqq \inf_{C' \in \mathrm{Alt}(C_t)} \sum_{(i,j)\in I} N_{ij}(t)\,\mathrm{KL}((\widehat{v}_t)_{ij}, v'_{ij}), \tag{10}$$

$\widehat{\ }$其中，$\widehat{v}_t$表示在 C_t下的伯努利分布（见（1）），相应的阈值为 β（t，δ）：$= 3\sum_{i\ j\ \in I}\ 1 +\ N_{ij}\ t\ + \binom{M}{2}\mathcal{C}_{\exp}\left(\ln(1/\delta)\right)$ _____ $\mathcal{C}_{\exp}$ · [45, (,) ln(ln ( )) $\binom{M}{2}$，其中 ( ) 的精确数学表达式见定理 7]。随后，广义似然比停止规则为

$$\tau_\delta \coloneqq \inf\{t \in \mathbb{N} : Z(t) > \beta(t,\delta)\}. \tag{11}$$

下面的结果表明，基于广义似然比规则的聚类不正确的概率不超过 δ。

命题3。（11）中的停止规则是 δ-正确的，即

$$\Pr\left(\tau_\delta < \infty,\ f([C_{\tau_\delta}]) \neq f([C])\right) \leq\ \delta.$$

命题3的证明见附录C-B。

在定理3中，我们证明，采用式 (9) 中的采样规则和式 (11) 中的停止规则，能够达到与定理 1 中给出的依赖于实例的下界 $D^*(C)^{-1}$ 相匹配的期望样本复杂度，其中参数为ϵ，$\sigma \to 0$，因此该策略几乎渐近最优。具体而言，我们证明，期望停止时间随 δ 至多按对数函数增长，且其归一化增长率的上界为 $D^*(C)^{-1}$。结合下界，这便确立了渐近最优性。

定理 3。对于任意实例 $C \in \mathcal{C}$ 以及固定参数 $\sigma > 0$ 和 $\epsilon \in (0,\ 1)$，采用式 (9) 中的采样规则和式 (11) 中的停止规则的策略满足

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_C[\tau_\delta]}{\ln(1/\delta)} \leq D^*_\epsilon(\sigma; C)^{-1},$$

其中，常数 $0 < D^*_\epsilon(\sigma;\ C)\ < \infty$ 由下式给出

$$\left[\inf_{C' \in \mathrm{Alt}(C)} \sum_{(i,j)\in I} \lambda^*_\epsilon(\sigma; C)\,\mathrm{KL}(v_{ij}, v'_{ij})\right] - \frac{\sigma}{2}\|\lambda^*_\epsilon(\sigma; C)\|^2.$$

根据最大值定理 *[46]*, ，可以证明

$$\lim_{\epsilon,\sigma\to 0} D^*_\epsilon(\sigma; C) = D^*(C). \tag{12}$$

定理3的证明见附录 C-C。

C. 一个计算上可行的算法：$\mathrm{A}^3\mathrm{CNP}$

尽管式(10)中的广义似然比统计量定义了一个几乎渐近最优的停止规则，但其在计算上不可行。关键瓶颈在于计算式(10)中的下确界，这需要同时对 $N_{ij}(t)$和$(\widehat{c}_t)_{ij}$进行优化，从而导致需要优化的大量组合。尽管前述问题看起来与式(5)中的上确界-下确界问题相似，但使得一个问题的解决方案无法适用于另一个问题的关键区别在于：在 $Z(t)$中，需要在所有$(\widehat{c}_t)_{ij}$ $i, j \in I$ 上进行优化的不同取值，相较于 $D^*(C)^{-1}$中仅有的两个取值 p 和 q。此外，在 C的经验估计 C_t 上使用$\widehat{\ }$，而非其投影 C_t ，对于证明命题3中的 δ-正确性性质至关重要。因此，仅在式(10)的目标函数中用$(c_t)_{ij}$ 替代$(\widehat{c}_t)_{ij}$ ，并不能解决当前问题。

我们提出了一种计算高效的检验统计量，它保持 δ-正确性这一性质，并实现近乎渐近最优性，误差仅相差一个可高效估计的乘法因子。从本质上说，我们通过“剥离”最小查询次数 $\min_{(i',j')\in I} N_{i'j'}(t)$，消除了广义似然比统计量关于 $N_{ij}(t)$ 的复杂性。令

$$\widehat{Z}(t) \coloneqq \left[\min_{(i',j')\in I} N_{i'j'}(t)\right]\left[\inf_{C' \in \mathrm{Alt}(C_t)} \sum_{(i,j)\in I} \mathrm{KL}((\widehat{v}_t)_{ij}, v'_{ij})\right], \tag{13}$$

并将修改后的停止规则定义为

$$\widehat{\tau}_\delta = \inf\{t \in \mathbb{N} : \widehat{Z}(t) > \beta(t,\delta)\}. \tag{14}$$

由于 $\widehat{Z}(t) \geq Z(t)$ 几乎必然成立，采用 $\widehat{Z}(t)$ 的停止规则继承了命题3的 δ-正确性。下面陈述这一结论，但不予证明。

推论 1. (14) 中的停止规则是 δ-正确的，即，

$$\Pr\left(\widehat{\tau}_\delta < \infty,\ f([C_{\widehat{\tau}_\delta}]) \neq f([C])\right) \leq \delta.$$

在附录 D 的命题 4 中，我们推导出 $\widehat{Z}(t)$ 的闭式表达式，并证明 (13) 中对 $\mathrm{Alt}(C_t)$ 的搜索可以简化为对有限多个等价类的搜索。因此，(13) 中新的停止统计量成为一种计算上可行且实用的替代方案，可取代基于广义似然比的统计量 $Z(t)$（由于缺少闭式表达式，其计算不可行）。然而，这种计算效率的提升可能带来一个缺点——可能要以牺牲渐近最优性为代价。为了刻画我们可能偏离最优性的程度，我们寻求一个期望采样复杂度的上界，如定理 3 中针对广义似然比统计量所得到的上界。

定理 4. 对于任意实例 $C \in \mathcal{C}$ 以及充分小的 $\sigma > 0$，在(9)中的采样规则和 (14)中的备选停止规则下，期望样本复杂度满足

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_C[\widehat{\tau}_\delta]}{\ln(1/\delta)} \leq \widetilde{D}^*_\epsilon(\sigma; C)^{-1}, \tag{15}$$

其中常数 $0 < D^*_\epsilon(\sigma; C) < \infty$ 由下式给出

$$\widetilde{D}^*_\epsilon(\sigma; C) = \left[\min_{(i',j') \in I} \lambda^*_\epsilon(\sigma; C)_{i'j'}\right]\left[\inf_{C' \in \mathrm{Alt}(C)} \sum_{(i,j)\in I} \mathrm{KL}(v_{ij}, v'_{ij})\right] - \frac{\sigma}{2}\|\lambda^*_\epsilon(\sigma; C)\|^2.$$

定理 4 的证明见附录 C-D。

将 (9) 中的 D-Tracking 采样规则与 (14) 中的实用停止规则结合起来，我们总结出算法 A^{3}CNP：基于含噪成对观测的近乎渐近最优主动聚类。A^{3}CNP 首先对每个项目对采样一次，随后按照 (9) 中的 D-Tracking 采样规则与 (14) 中的实用停止规则结合起来，我们总结出算法 A^{3}CNP：基于含噪成对观测的近乎渐近最优主动聚类。A^{3}CNP 首先对每个项目对采样一次，随后按照 (t) 进行采样。在时间步 t，先将估计的实例投影到 $\mathcal{C}$ 上，得到 $C_t = \mathrm{Proj}_{\mathcal{C}}(C_t)$。在求解 $\lambda^*(\sigma; C_t)$ 之后，采样分布 $\lambda^*_\epsilon(\sigma; C_t)$ 按 (8) 中的方式由 $\lambda^*(\sigma; C_t)$ 与均匀分布混合得到，并查询项目对 $(\mathbf{I}_{t+1}, \mathbf{J}_{t+1}) \sim \lambda^*_\epsilon(\sigma; C_t)$。收到查询的二元奖励后，计算统计量 $\widehat{Z}(t)$。若 $\widehat{Z}(t) \geq \beta(t,\delta)$，则 A^{3}CNP 停止，并在停止时间 $\widehat{\tau}_\delta = t$ 输出 $f(C_t)$，即 C_t 下的聚类。

D. 关于 *A3CNP* 的次优性差距

现在，我们研究用 A3CNP 中计算上可行的替代量取代式 (10) 中的广义似然比统计量所导致的、依赖于实例的乘法次优性差距；对于固定的 $\epsilon \in (0, 1)$ 和 $\sigma > 0$，这将得到式 (15)。具体而言，我们通过比值比较 $D^\star_\epsilon(\sigma; C)$ 和 $D^\star_\epsilon(\sigma; C)$

$$\mathrm{SG}_\epsilon(\sigma; C) \coloneqq \frac{D^\star_\epsilon(\sigma; C)}{\widetilde{D}^\star_\epsilon(\sigma; C)}.$$

我们强调，相对于式 (3) 下界中出现的信息论基准 $D^\star(C)$，实际差距为 $D^\star(C)/D^\star_\epsilon(\sigma; C)$。尽管如此，$\mathrm{SG}_\epsilon(\sigma; C)$ 是一种有理有据且方便的代理：根据式 (12)，当 $\epsilon,\ \sigma \to 0$ 时，分子 $D^\star_\epsilon(\sigma; C)$ 趋近于 $D^\star(C)$；因此，当 ϵ 和 σ 足够小时，预期比值 $D^\star(C)/D^\star_\epsilon(\sigma; C)$ 和 $D^\star_\epsilon(\sigma; C)/D^\star_\epsilon(\sigma; C)$ 会较为接近（这是启发式判断，并且我们的第五节实验结果也为此提供了实证支持）。

为了获得 $\mathrm{SG}_\epsilon(\sigma; C)$ 的显式依赖具体实例的上界，将采样分配 $\lambda^\star_\epsilon(\sigma; C)$ 的最大值和最小值分别定义为 $m \coloneqq \max_{(i,j)\in I}\left(\lambda^\star_\epsilon(\sigma; C)\right)_{ij}$ 和 $m \coloneqq \min_{(i,j)\in I}\left(\lambda^\star_\epsilon(\sigma; C)\right)_{ij}$，并引入依赖具体实例的分离常数

$$A(C) \coloneqq \inf_{C' \in \mathrm{Alt}(C)} \sum_{(i,j)\in I} \mathrm{KL}\big(v_{ij}, v'_{ij}\big).$$

利用基本界

$$\underline{m} \sum_{i,j} \mathrm{KL}(v_{ij}, v'_{ij}) \leq \sum_{i,j} (\lambda^\star_\epsilon)_{ij} \mathrm{KL}(v_{ij}, v'_{ij}) \leq \overline{m} \sum_{i,j} \mathrm{KL}(v_{ij}, v'_{ij})$$

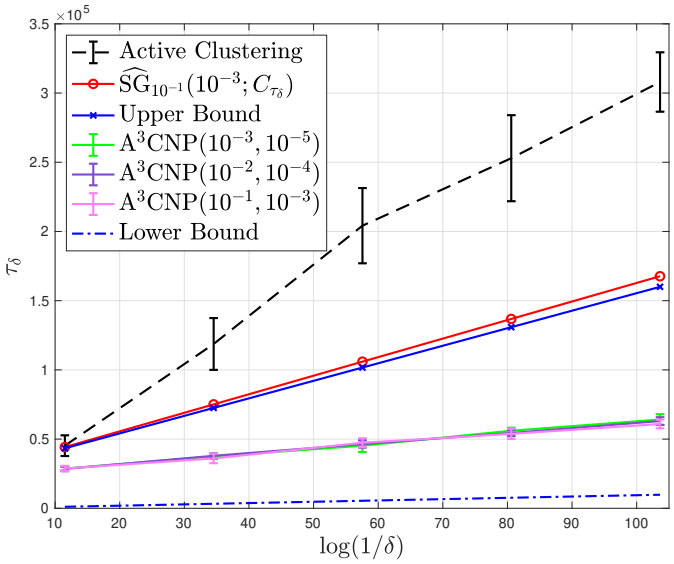

图 1. A³CNP 的渐近（$\delta \to 0$）样本复杂度。图中给出了不同 ϵ（第一个参数）和 σ（第二个参数）取值下，相对于 [24] 的主动聚类算法的结果。图中还包括理论下界 (3)、上界 (15)，以及在停止时间 τ_δ 处使用 $\epsilon = 10^{-1}$ 和 $\sigma = 10^{-3}$ 计算的 (19) 中的数据依赖代理 $\widehat{\mathrm{SG}}_\epsilon(\sigma;\, C_{\tau_\delta})$。

结合 $D^\star_\epsilon(\sigma;\, C)$ 和 $D^\star_\epsilon(\sigma;\, C)$ 的定义，我们得到

$$\mathrm{SG}_\epsilon(\sigma; C) \;\le\; \frac{\overline{m} A(C)}{\underline{m} A(C) - \frac{\sigma}{2}\left\|\lambda^\star_\epsilon(\sigma; C)\right\|_2^2}, \tag{16}$$

只要分母为正。

接下来需要控制 $\|\lambda^\star_\epsilon(\sigma;\, C)\|_2^2$。回顾 (8) 中 $\lambda^\star_\epsilon(\sigma;\, C)$ 的定义，并利用 $\|\lambda^\star(\sigma;\, C)\|_2^2 \le 1$、$\langle \lambda^\star(\sigma;\, C)\; u\rangle = 1/|I|$ 和 $\|u\|_2^2 = 1/|I|$，可得

$$\begin{aligned}\left\|\lambda^\star_\epsilon(\sigma; C)\right\|_2^2 &= \left\|(1-\epsilon)\lambda^\star(\sigma; C) + \epsilon u\right\|_2^2 \\ &\le (1-\epsilon)^2 + \frac{2\epsilon - \epsilon^2}{|I|}.\end{aligned} \tag{17}$$

将 (17) 代入 (16)，得到显式界

$$\mathrm{SG}_\epsilon(\sigma; C) \;\le\; \frac{\overline{m}/\underline{m}}{1 - \frac{\sigma}{2\underline{m} A(C)}\left((1-\epsilon)^2 + \frac{2\epsilon-\epsilon^2}{|I|}\right)}. \tag{18}$$

(18) 中的界阐明了差距的两个来源：抽样比例的离散度 $\overline{m}/\underline{m}$，以及正则化相对于 $\underline{m} A(C)$ 的大小。特别地，当 σ 足够小，使得分母中的修正项可以忽略时，该界主要由 $\overline{m}/\underline{m}$ 支配。

受 (18) 的启发（其中 ϵ 和 σ 较小），我们使用数据依赖代理：

$$\widehat{\mathrm{SG}}_\epsilon(\sigma; C_t) \;:=\; \frac{\max_{(i,j)\in I}\left(\lambda^\star_\epsilon(\sigma; C_t)\right)_{ij}}{\min_{(i,j)\in I}\left(\lambda^\star_\epsilon(\sigma; C_t)\right)_{ij}}. \tag{19}$$

当以足够小的 ϵ 和 σ 运行 A³CNP 时，我们观察到 $\widehat{\mathrm{SG}}_\epsilon(\sigma; C_t)$ 能够很好地跟踪 $\mathrm{SG}_\epsilon(\sigma; C)$，从而为占主导地位的离散度项 $\overline{m}/\underline{m}$ 提供一个简单的诊断指标。

V. 实验结果

为验证 A³CNP 的性能，我们进行了实验，并与 Chen等人 [24],提出的算法进行基准比较，该算法针对相同的问题设定。我们使用 $M = 6$、$K = 3$，选择概率彼此相对接近这一困难情形，其中 $p = 0.6$ 和 $q = 0.4$，且簇分配为 $\{\{1, 2\}, \{3, 4, 5\}, \{6\}\}$。此外，借鉴 [45]，我们在实验中采用了较小的阈值 $\beta(t,\ \delta) := 3R\log(1 + \log(t/R)) + R\,\mathcal{C}_{\exp}(\log((M-1)/\delta)/R)$。这里，$M$ 表示 M 个项目可能的聚类方式总数（底层簇的数量事先未知），而 $R = M - 1$ 表示问题的秩——描述任意给定实例等价类所需变量数的最坏情况值

；关于秩的讨论，请参见附录E。正如 [45, 命题 21], 所述，使用上述阈值的基于GLR的停止规则仍然是 δ-正确的。

在图1中，我们比较了 A^{3}CNP 与 [24], 的主动聚类算法在 10 次独立试验中的平均停止时间。此外，我们还绘出了定理 1 的下界，以及式 (18) 中的上界；后者是将式 (18) 右侧乘以 $D^*(C)^{-1}\ln(1/\delta)$ 得到的。在所有测试的 ϵ 和 σ取值下，A^{3}CNP 的斜率都与下界的斜率高度吻合，且明显比上界的斜率平缓，体现出近最优缩放。此外，A^{3}CNP 的停止时间随 ϵ 和 σ的变化仅有微小波动，彰显了其对这些 $\widehat{}$参数的强鲁棒性。式 (19) 中的数据依赖代理 $\mathrm{SG}_\epsilon(\sigma;\, C_{\tau_\delta})$（在停止时间 τ_δ处评估）始终能准确近似上界。总体而言，在这一实验设置下，A^{3}CNP 的表现持续显著优于 Chen等人 [24], 提出的算法，所需停止时间明显更短。

VI. 结论与未来研究

H

在本研究中，我们为固定置信度情形下基于含噪成对观测的主动聚类建立了一个有理有据的框架。通过将该问题表述为纯探索赌博机任务，我们推导出了期望查询复杂度的依赖于实例的下界，并展示了如何通过自适应采样和停止规则逼近该下界。我们的分析最终得出了 A^{3}CNP 算法，该算法将统计效率高的采样与计算上易于处理的停止准则相结合；随着目标错误概率趋于零，我们还刻画了其难度参数与下界对应难度参数之间的差距。

除具体的算法贡献之外，本文还从更广阔的视角强调了主动聚类作为信息获取问题的意义。最难备择实例可被刻画为对真实聚类进行简单的合并或拆分扰动，这表明，对细粒度结构的理解能够在设计高效自适应流程时发挥关键作用。这一观点为新一类聚类算法打开了大门：它们会动态地将精力集中于最具歧义的结构性决策，而不是均匀地细化所有成对关系。

展望未来，有几个方向显得尤为有前景。将框架扩展到更丰富、更贴近现实的反馈模型——例如异质噪声水平、上下文信息或边信息，或结构化查询代价——将使主动聚类更接近实际部署场景。在理论方面，开发计算高效且完全渐近最优的停止规则仍是一个开放挑战；在无法查询所有配对的大规模情形中理解最优权衡亦是如此。更广泛地说，我们相信本文引入的工具将在聚类之外发挥作用，为研究赌博机、信息论与无监督学习之交集上的自适应结构发现问题奠定基础。

附录 A 等价关系示例演示（第二节）

本质上，第二节中定义的等价关系只是在表示相同簇分配、但奖励分布具有不同伯努利参数的实例上形成等价类。例如，设 $C, C' \in \mathcal{C}$

$$C = \begin{pmatrix} p & p & q & q & q & q \\ p & p & q & q & q & q \\ q & q & p & q & q & q \\ q & q & q & p & p & p \\ q & q & q & p & p & p \\ q & q & q & p & p & p \end{pmatrix}; \qquad C' = \begin{pmatrix} p' & p' & q' & q' & q' & q' \\ p' & p' & q' & q' & q' & q' \\ q' & q' & p' & q' & q' & q' \\ q' & q' & q' & p' & p' & p' \\ q' & q' & q' & p' & p' & p' \\ q' & q' & q' & p' & p' & p' \end{pmatrix},$$

其中 $1 \geq p, p' > 0.5 > q, q \geq 0$ 以及 $p \neq p'$, $q \neq q'$。二元矩阵 $C_=$, $C'_=$, $C_\neq$ 和 $C'_\neq$ 如下

$$C_= = C'_= \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}; \qquad C_\neq = C'_\neq = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

将 $C_=$ 和 $C'_=$ 中的 p、p'（分别为 q、q'）全部替换为 1（分别为 0），并在 $C_\neq$ 和 $C'_\neq$ 中进行相反的替换，即可轻易看出这一点。然后，$C \sim C'$ 和这两个矩阵都表示簇分配 $\{\{1,2\},\{3\},\{4,5,6\}\}$。

附录 B 第三节结果的证明

A. 定理 *1* 的证明

证明。 我们首先假设$\mathbb{E}_C[\tau_\delta]$ 是有限的。根据附录 B-B 中上确界-下确界优化问题的简化形式，很容易看出 $D^*(C) > 0$。对于固定的 $C \in \mathcal{C}$ 和 $C' \in \mathrm{Alt}(C)$，我们可以将它们视为分别具有 $\binom{M}{2}$ 个臂的赌博机模型 $\{v_{ij}\}_{(i,j)\in I}$ 和 $\{v'_{ij}\}_{(i,j)\in I}$。然后，根据 [47, 引理 1], ，我们有

$$\sum_{(i,j)\in I} \mathbb{E}_C[N_{ij}(\tau_\delta)]\mathrm{KL}(v_{ij}, v'_{ij}) \geq d(\delta, 1-\delta).$$

由于$\sum_{(i,j)\in I} \mathbb{E}_C[N_{ij}(\tau_\delta)] = \mathbb{E}_C[\tau_\delta]$ 以及 $\{\frac{\mathbb{E}_C[N_{ij}}{\mathbb{E}_C[\tau_\delta]}^{\tau_\delta}\}_{(i,j)\in I})]$ 在 I上构成一个概率分布，

$$\begin{aligned}
d(\delta, 1-\delta) &\leq \inf_{C'\in\mathrm{Alt}(C)} \sum_{(i,j)\in I} \mathbb{E}_C[N_{ij}(\tau_\delta)]\mathrm{KL}(v_{ij}, v'_{ij}) \\
&= \mathbb{E}_C[\tau_\delta] \inf_{C'\in\mathrm{Alt}(C)} \sum_{(i,j)\in I} \frac{\mathbb{E}_C[N_{ij}(\tau_\delta)]}{\mathbb{E}_C[\tau_\delta]}\mathrm{KL}(v_{ij}, v'_{ij}) \\
&\leq \mathbb{E}_C[\tau_\delta] \sup_{\lambda\in\mathcal{P}_I} \inf_{C'\in\mathrm{Alt}(C)} \sum_{(i,j)\in I} \lambda_{ij}\mathrm{KL}(v_{ij}, v'_{ij}) \\
&= \mathbb{E}_C[\tau_\delta] D^*(C).
\end{aligned}$$

如果$\mathbb{E}_C[\tau_\delta]$ 是无穷大，则该不等式显然成立。于是，由于 $\lim_{\delta\to 0} \frac{d}{(\delta,1-\delta)\ \ln(1/\delta)} = 1$,

$$\liminf_{\delta\to 0} \frac{\mathbb{E}_C[\tau_\delta]}{\ln(1/\delta)} \geq D^*(C)^{-1}$$

如所需，结论成立。 □

B. 定理 *2* 的证明

证明。 利用第二节中定义的等价关系，我们可以将上确界-下确界优化问题写为

$$\sup_{\lambda\in\mathcal{P}_I} \inf_{C'\in\mathrm{Alt}(C)} h(\lambda, C') = \sup_{\lambda\in\mathcal{P}_I} \inf_{[C']\in\mathrm{Alt}(C)/\sim} \inf_{C''\in[C']} h(\lambda, C''),$$

其中，回顾一下，$\mathrm{Alt}(C)/\sim$ 表示所有等价类 $\mathrm{Alt}(C)$ 的集合。根据引理 1 ，

$$\begin{aligned}
&\sup_{\lambda\in\mathcal{P}_I} \inf_{[C']\in\mathrm{Alt}(C)/\sim} \inf_{C''\in[C']} h(\lambda, C'') \\
&= \sup_{\lambda\in\mathcal{P}_I} \inf_{[C']\in\mathrm{Alt}(C)/\sim} \alpha - b_1\log(1-q') - b_2\log(q') - b_3\log(1-p') - b_4\log(p') \\
&= \sup_{\lambda\in\mathcal{P}_I} \min_{[C']\in\mathrm{Alt}(C)/\sim} z(\lambda_{N_1}, \lambda_{N_2}),
\end{aligned}$$

其中， α、 b_1、 b_2、 b_3、 b_4、 p'和 q'的定义见引理；由于 $\mathrm{Alt}(C)/\sim$是有限集，inf 可以写成 min。上述目标函数等价于引理 2（如下）中定义的 $z(\lambda_{N_1},\lambda_{N_2})$，且在固定 λ和 $C \in \mathcal{C}$时，关于 λ_{N_1}、 λ_{N_2}递增，其中 $\lambda_{N_1} = \sum_{(i,j)\in\Lambda_{N_1}([C'])} \lambda_{ij}$ 和 $\lambda_{N_2} = \sum_{(i,j)\in\Lambda_{N_2}([C'])} \lambda_{ij}$。

接下来，我们分析内部最小值

$$\min_{[C']\in\mathrm{Alt}(C)/\sim} z(\lambda_{N_1}, \lambda_{N_2}),.$$

在这种情况下，当对任意的 $\lambda \in \mathcal{P}_I$ 在始终不相交的 $\Lambda_{N_1}([C'])$ $\Lambda_{N_2}([C'])$ 上最小化 $\lambda_{N_1} + \lambda_{N_2}$ 时，必有

$$\begin{aligned}
&\lambda_{N_1} + \lambda_{N_2} \leq \lambda_{N'_1} + \lambda_{N'_2} = \sum_{(i,j)\in\Lambda_{N_1}([C''])} \lambda_{ij} + \sum_{(i,j)\in\Lambda_{N_2}([C''])} \lambda_{ij} \\
&\overset{\text{(i)}}{\Longleftrightarrow} \Lambda_{N_1}([C']) \subset \Lambda_{N_1}([C'']) \text{ and } \Lambda_{N_2}([C']) \subset \Lambda_{N_2}([C'']) \\
&\overset{\text{(ii)}}{\Longleftrightarrow} \lambda_{N_1} \leq \lambda_{N'_1} \text{ and } \lambda_{N_2} \leq \lambda_{N'_2} \\
&\overset{\text{(iii)}}{\Rightarrow} z(\lambda_{N_1}, \lambda_{N_2}) \leq z(\lambda_{N'_1}, \lambda_{N'_2}).
\end{aligned}$$

从 $\Lambda_{N_1}([C'])$、 $\Lambda_{N_2}([C'])$ 的定义以及 z 关于 $\lambda_{N_1}, \lambda_{N_2}$ 递增这一事实，可以明显看出蕴含关系 (ii) 和 (iii)。由 (ii) 可知，(i) 中的 $\Leftarrow$ 显然成立。因此，我们只需证明 (i) 中的 $\Rightarrow$。假设 $\lambda_{N_1} + \lambda_{N_2} \leq \lambda_{N_1'} + \lambda_{N_2'}$，但 $\Lambda_{N_1}([C']) \not\subset \Lambda_{N_1}([C''])$ 或 $\Lambda_{N_2}([C']) \not\subset \Lambda_{N_2}([C''])$。不失一般性，假设 $\Lambda_{N_1}([C']) \not\subset \Lambda_{N_1}([C''])$，这样我们就能找到 $(i,\ j) \in \Lambda_{N_1}([C']) \setminus (\Lambda_{N_1}([C']) \cap \Lambda_{N_1}([C''])$。于是，存在 $\lambda' \in \mathcal{P}_I$ 和足够小的 $\epsilon, \gamma > 0$，使得

$$\lambda'_{N_1'} + \lambda'_{N_2'} = \epsilon > \lambda'_{\Lambda_{N_1}([C'])\cap\Lambda_{N_1}([C''])} + \lambda'_{\Lambda_{N_2}([C'])\cap\Lambda_{N_2}([C''])} \geq 0,$$
$$\lambda'_{N_p\setminus(\Lambda_{N_1}([C'])\cup\Lambda_{N_1}([C'']))} + \lambda'_{N_q\setminus(\Lambda_{N_2}([C'])\cup\Lambda_{N_2}([C'']))} = \gamma > 0$$

and

$$\lambda'_{N_1} + \lambda'_{N_2} = 1 - \gamma - \epsilon + \lambda'_{\Lambda_{N_1}([C'])\cap\Lambda_{N_1}([C''])} + \lambda'_{\Lambda_{N_2}([C'])\cap\Lambda_{N_2}([C''])} > \epsilon.$$

因此，产生矛盾。

根据上述蕴含关系 1、2 和 3，显然，对于任意固定的 $\lambda \in \mathcal{P}_I$ 和 $C \in \mathcal{C}^K$，最小化$\lambda_{N_1} + \lambda_{N_2}$ 会导致最小化 $z(\lambda_{N_1}, \lambda_{N_2})$。我们将在引理 3 中反复利用这一事实，从而进一步简化内部最小值，

$$\sup_{\lambda\in\mathcal{P}_I} \min_{[C']\in \mathrm{Alt}(C)/\sim} z(\lambda_{N_1}, \lambda_{N_2}) = \sup_{\lambda\in\mathcal{P}_I} \min_{[C']\in(\mathrm{Alt}(C)\cap(\mathcal{C}^{K-1}\cup\mathcal{C}^{K+1}))/\sim} z(\lambda_{N_1}, \lambda_{N_2}).$$

最后，为了完成证明，我们需要证明：对于 $[C'] \in (\mathrm{Alt}(C) \cap (\mathcal{C}^{K-1} \cup \mathcal{C}^{K+1}))/\sim$，若 $[C'] \notin \mathrm{Alt}(C)_1/\sim$，则存在一个 $[C] \in \mathrm{Alt}(C)_1/\sim$，使得 $\lambda_{\bar{N}_1} + \lambda_{\bar{N}_2} = \sum_{(i,j)\in\Lambda_{N_1}([\bar{C}])} \lambda_{ij} + \sum_{(i,j)\in\Lambda_{N_2}([\bar{C}])} \lambda_{ij} \leq \lambda_{N_1} + \lambda_{N_2}$。按照引理 3 的方法，我们将考虑两种情形：第一种，$[C']$ 有 $K+1$ 个簇；第二种，$[C']$ 有 $K-1$ 个簇。

情况1：$[C'] \in (\mathrm{Alt}(C) \cap \mathcal{C}^{K+1})/\sim$如果 $[C'] \notin \mathrm{Alt}(C)_1/\sim$，那么

$$f([C']) \neq \{g_1, \cdots, g_{i-1}, g'_{i_1}, g'_{i_2}, g_{i+1}, \cdots, g_K\},$$

其中 $g'_{i_1} \cup g'_{i_2} = g_i$，并且对所有 $k \in [K]$，至少存在一个 $g'_j \not\subset g_k$。[1] 这意味着存在项目 $m_1 > m_2 \in g'_j$，使得 m_1 和 m_2 属于不同的 g_k。因此，$(m_2, m_1) \in \Lambda_{N_2}([C'])$。此外，由于 $|f([C'])| = K+1$，对于所有 $g_i \in f(C)$，我们也不可能有 $g_i \subset g'_k$，否则 $f([C'])$ 中的集合将少于 $K+1$ 个。因此，必定至少存在一个 $g_i \in f(C)$，使得存在 $g'_{k_1}, g'_{k_2} \in f([C'])$，其中 $|g'_{k_1} \cap g_i|, |g'_{k_2} \cap g_i| > 0$。接着，令 $g_{k_1,i} := g'_{k_1} \cap g_i$，且 $I_{k_1,i} := \{(i,j) \in g_{k_1,i} \times (g_i \setminus g_{k_1,i}) \cup (g_i \setminus g_{k_1,i}) \times g_{k_1,i} :\ j > i\} \subset \lambda_{N_1}(C')$。由于 $[\bar{C}] := f^{-1}(\{g_1, \cdots, g_{i-1}, g_{k_1,i}, g_i \setminus g_{k_1,i}, g_{i+1}, \cdots, g_K\}) \in \mathrm{Alt}(C)_1$，且 $\lambda_{\bar{N}_1} + \lambda_{\bar{N}_2} = \sum_{(i,j)\in I_{k_1,i}} \lambda_{ij}$，我们必有 $\lambda_{\bar{N}_1} + \lambda_{\bar{N}_2} \leq \lambda_{N_1} + \lambda_{N_2}$。

情况 2：$[C'] \in (\mathrm{Alt}(C) \cap \mathcal{C}^{K-1})/\sim$如果 $[C'] \notin \mathrm{Alt}(C)_1/\sim$，那么

$$f([C']) \neq \{g_1, \cdots, g_{i_1-1}, g_{i_1+1}, \cdots, g_{i_2-1}, g_{i_2+1}, \cdots, g_K, g'_i\},$$

其中 $g'_i = g_{i_1} \cup g_{i_2}$ 并且对于所有 $k \in [K-1]$，至少存在一个 $g_j \not\subset g'_k$。[2] 这意味着存在某个$g'_k \in f([C'])$，使得 $|g_j| > |g_j \cap g'_k| > 0$；令 $g_{k,\ j} := g'_k \cap g_j$，以及 $I_{k,\ j} := \{(i,\ j) \in g_k,\ _j \times (g_j \setminus g_k,\ _j) \cup (g_j \setminus g_k,\ _j) \times g_k,\ _j :\ j > i\} \subset \Lambda_{N_1}([C'])$。那么，对于 $[\bar{C}] := f^{-1}(\{g_1, \cdots, g_{j-1}, g_{k,j}, g_j \setminus g_{k,j}, g_{j+1}, \cdots, g_K\}) \in \mathrm{Alt}(C)_1$，有$\lambda_{\bar{N}_1} + \lambda_{\bar{N}_2} = \sum_{(i,\ j)\in I_k,\ _j} \lambda_{ij} \leq \lambda_{N_1} + \lambda_{N_2}$。 □

引理 1. 对于固定的 $\lambda \in \mathcal{P}_I$、$C \in \mathcal{C}$ 以及任意 $[C']$，求解

$$\inf_{C''\in[C']} h(\lambda, C'') = \inf_{1\geq p''>0.5>q''\geq 0} h(\lambda, C'')$$

得到

$$p'' = \max\left\{0.5, \frac{\lambda'_{N_1}}{\lambda'_{N_1} + \lambda_{N_2}} p + \frac{\lambda_{N_2}}{\lambda'_{N_1} + \lambda_{N_2}} q\right\};\quad q'' = \min\left\{0.5, \frac{\lambda_{N_1}}{\lambda_{N_1} + \lambda'_{N_2}} p + \frac{\lambda'_{N_2}}{\lambda_{N_1} + \lambda'_{N_2}} q\right\},$$

其中，略微滥用记号，$\lambda_{N_1} = \sum_{(i,j)\in\Lambda_{N_1}([C'])} \lambda_{ij}$, $\lambda'_{N_1} = \sum_{(i,j)\in N_p} \lambda_{ij} - \lambda_{N_1}$, $\lambda_{N_2} = \sum_{(i,j)\in\Lambda_{N_2}([C'])} \lambda_{ij}$ 以及$\lambda'_{N_2} = \sum_{(i,j)\in N_q} \lambda_{ij} - \lambda_{N_2}$。

[1] If $g'_j \subset g_k$ 对于所有 $j \in [K+1]$，则不能有超过两个 g'_j 是某个 g_k的真子集，从而迫使 $[C'] \in \mathrm{Alt}(C)_1$.

[2] 如果对所有 $g_j \in f(C)$，$g_j \subset g'_k$，那么所有 g_j 均未被拆分。为了使 $|f([C'])| = K-1$，必须恰好有两个 g_{i_1}, $g_{i_2} \in f(C)$ 合并形成 g'_i。因此，$[C'] \in \mathrm{Alt}(C)_1/\sim$。

证明。回顾式 (5) 中 $h(\lambda,\ C)$ 的定义。由于对所有 $C'',\ C''' \in [C'],\ \Lambda_{N_1}(C'') = \Lambda_{N_1}(C''')$ 且 $\Lambda_{N_2}(C'') = \Lambda_{N_2}(C''')$，$\lambda_{N_1}$，$\lambda_{N_2}$，$\lambda'_{N_1}$ 和 λ'_{N_2} 都是常数。因此，引理中的优化问题只需通过求解最优的 p'' 和 q'' 即可。展开 $h(\lambda,\ C'')$，得到

$$
\begin{aligned}
h(\lambda, C'') &= \lambda_{N_1} d(p, q'') + \lambda'_{N_2} d(q, q'') + \lambda_{N_2} d(q, p'') + \lambda'_{N_1} d(p, p'') \\
&= \lambda_{N_1}\left[(1-p)\log\left(\frac{1-p}{1-q''}\right) + p\log\left(\frac{p}{q''}\right)\right] + \lambda'_{N_2}\left[(1-q)\log\left(\frac{1-q}{1-q''}\right) + q\log\left(\frac{q}{q''}\right)\right] \\
&\quad + \lambda_{N_2}\left[(1-q)\log\left(\frac{1-q}{1-p''}\right) + q\log\left(\frac{q}{p''}\right)\right] + \lambda'_{N_1}\left[(1-p)\log\left(\frac{1-p}{1-p''}\right) + p\log\left(\frac{p}{p''}\right)\right] \\
&= \lambda_{N_1}(1-p)\log(1-p) - \lambda_{N_1}(1-p)\log(1-q''') + \lambda_{N_1} p\log(p) - \lambda_{N_1} p\log(q''') \\
&\quad + \lambda'_{N_2}(1-q)\log(1-q) - \lambda'_{N_2}(1-q)\log(1-q''') + \lambda'_{N_2} q\log(q) - \lambda'_{N_2} q\log(q''') \\
&\quad + \lambda_{N_2}(1-q)\log(1-q) - \lambda_{N_2}(1-q)\log(1-p'') + \lambda_{N_2} q\log(q) - \lambda_{N_2} q\log(p'') \\
&\quad + \lambda'_{N_1}(1-p)\log(1-p) - \lambda'_{N_1}(1-p)\log(1-p'') + \lambda'_{N_1} p\log(p) - \lambda'_{N_1} p\log(p'') \\
&= (\lambda_{N_1} + \lambda'_{N_1})[(1-p)\log(1-p) + p\log(p)] + (\lambda_{N_2} + \lambda'_{N_2})[(1-q)\log(1-q) + q\log(q)] \\
&\quad - [\lambda_{N_1}(1-p) + \lambda'_{N_2}(1-q)]\log(1-q'') - [\lambda_{N_1} p + \lambda'_{N_2} q]\log(q'') \\
&\quad - [\lambda'_{N_1}(1-p) + \lambda_{N_2}(1-q)]\log(1-p'') - [\lambda'_{N_1} p + \lambda_{N_2} q]\log(p'') \\
&= \alpha - b_1\log(1-q'') - b_2\log(q'') - b_3\log(1-p'') - b_4\log(p''),
\end{aligned}
$$

其中，$\alpha := (\lambda_{N_1}+\lambda'_{N_1})[(1-p)\log(1-p)+p\log(p)] + (\lambda_{N_2}+\lambda'_{N_2})[(1-q)\log(1-q)+q\log(q)]$, $b_1 := \lambda_{N_1}(1-p)+\lambda'_{N_2}(1-q)$, $b_2 := \lambda_{N_1} p + \lambda'_{N_2} q$, $b_3 := \lambda'_{N_1}(1-p) + \lambda_{N_2}(1-q)$，且 $b_4 := \lambda'_{N_1} p + \lambda_{N_2} q$ 均为常数。

定义 $g_2(q'') := -b_1\log(1-q'') - b_2\log(q'')$ 和 $g_4(p'') := -b_3\log(1-p'') - b_4\log(p'')$，并将 p'' 重新标记为 y，将 q'' 重新标记为 x，以便阅读。对 x 和 y 求 h 的导数，可得

$$
\frac{dh}{dx} = \frac{dg_2}{dx} = \frac{b_1}{1-x} - \frac{b_2}{x} = (b_1+b_2)\frac{x - \frac{b_2}{b_1+b_2}}{x(1-x)} \tag{20}
$$

$$
\frac{dh}{dy} = \frac{dg_4}{dy} = \frac{b_3}{1-y} - \frac{b_4}{y} = (b_3+b_4)\frac{y - \frac{b_4}{b_3+b_4}}{y(1-y)}. \tag{21}
$$

由于 $b_1, b_2, b_3, b_4 \geq 0$ 以及 $1/2 > x \geq 0$ 和 $1 \geq y > 1/2$:

考察 (20):

Case 1: $\frac{b_2}{b_1+b_2} \geq 1/2$

则 $\frac{dh}{dx} < 0,\ \ \forall x \in (0, 1/2)$，且下确界在 $x = 1/2$ 处取得

Case 2: $0 < \frac{b_2}{b_1+b_2} < 1/2$

then, $\frac{dh}{dx} < 0, \forall x \in (0, \frac{b_2}{b_1+b_2})$ and $\frac{dh}{dx} > 0, \forall x \in (\frac{b_2}{b_1+b_2}, 1/2)$ so, the infimum occurs at $x = \frac{b_2}{b_1+b_2}$

Case 3: $b_2 = 0, b_1 > 0 \implies \frac{b_2}{b_1+b_2} = 0$

则 $\frac{dh}{dx} = b_1 \frac{1}{1-x} > 0,\ \ \forall x \in (0, 1/2)$，故下确界在 $x = 0 = \frac{b_2}{b_1+b_2}$ 处取得。

Case 4: $b_2 > 0, b_1 = 0 \implies \frac{b_2}{b_1+b_2} = 1$

则 $\frac{dh}{dx} = b_2 \frac{-1}{x} < 0,\ \ \forall x \in (0, 1/2)$，故下确界在 $x = 0.5$ 处取得。

Case 5: $b_2 = b_1 = 0 \implies g_2(x) = 0.$

考察 (21):

Case 1: $0 < \frac{b_4}{b_4+b_3} \leq 1/2$

那么, $\frac{dh}{dy} > 0,\ \ \forall y \in$ （1/2, 1），且下确界在 $y = 1/2$ 处取得

情况 2: $\frac{b_4}{b_4+b_3} > 1/2$

那么, $\frac{dh}{dy} < 0, \forall y \in (1/2, \frac{b_4}{b_3+b_4})$ 和 $\frac{dh}{dy} > 0, \forall y \in (\frac{b_4}{b_3+b_4}, 1]$ 因此, 下确界在 $y = \frac{b_4}{b_3+b_4}$ 处取得

Case 3: $b_4 = 0, b_3 > 0 \implies \frac{b_4}{b_3+b_4} = 0$

那么, $\frac{dh}{dy} = b_3 \frac{1}{1-y} > 0, \forall y \in (1/2, 1)$, 因此下确界在 $y = 0.5$ 处取得。

Case 4: $b_4 > 0, b_3 = 0 \implies \frac{b_4}{b_3+b_4} = 1$

那么, $\frac{dh}{dx} = b_4 \frac{-1}{y} < 0, \forall y \in (1/2, 1)$, 因此下确界在 $y = 1 = \frac{b_4}{b_3+b_4}$ 处取得。

Case 5: $b_3 = b_4 = 0 \implies g_4(y) = 0.$

综合所有情形, 我们有

$$q'' = x = \min\left\{0.5, \frac{b_2}{b_1+b_2}\right\} = \min\left\{0.5, \frac{\lambda_{N_1}}{\lambda_{N_1}+\lambda'_{N_2}}p + \frac{\lambda'_{N_2}}{\lambda_{N_1}+\lambda'_{N_2}}q\right\}$$
$$p'' = y = \max\left\{0.5, \frac{b_4}{b_3+b_4}\right\} = \max\left\{0.5, \frac{\lambda'_{N_1}}{\lambda'_{N_1}+\lambda_{N_2}}p + \frac{\lambda_{N_2}}{\lambda'_{N_1}+\lambda_{N_2}}q\right\},$$

其中约定: 若 $b_1 = b_2 = 0$ $\frac{b_2}{b_1+b_2} = 0$ $b_3 = b_4 = 0$ $b_1 = b_2 =$, 则....., 类似地, 对于.....。不可能出现这种情况, 因为 $\sum_{ij \in I} \lambda_{ij} = 1$。 □

引理 2. 对于固定的 $\lambda \in \mathcal{P}_I$、 $C \in \mathcal{C}$, 我们可以定义一个函数 z: $\mathbb{R}^2 \to \mathbb{R}$ 其中

$$\begin{aligned} z(\lambda_{N_1}, \lambda_{N_2}) = \alpha &- b_1(\lambda_{N_1}, \lambda_{N_2}) \log(1 - q'(\lambda_{N_1}, \lambda_{N_2})) \\ &- b_2(\lambda_{N_1}, \lambda_{N_2}) \log(q'(\lambda_{N_1}, \lambda_{N_2})) \\ &- b_3(\lambda_{N_1}, \lambda_{N_2}) \log(1 - p'(\lambda_{N_1}, \lambda_{N_2})) \\ &- b_4(\lambda_{N_1}, \lambda_{N_2}) \log(p'(\lambda_{N_1}, \lambda_{N_2})). \end{aligned}$$

其中

$$p' = \max\{0.5, \frac{b_4}{b_3+b_4}\}; \quad q' = \min\{0.5, \frac{b_2}{b_1+b_2}\}$$
$$b_1 = \lambda_{N_1}(1-p) + \lambda'_{N_2}(1-q); \quad b_2 = \lambda_{N_1}p + \lambda'_{N_2}q$$
$$b_3 = \lambda'_{N_1}(1-p) + \lambda_{N_2}(1-q); \quad b_4 = \lambda'_{N_1}p + \lambda_{N_2}q$$

且 $\lambda'_{N_1} = \sum_{ij \in N_p} \lambda_{ij} - \lambda_{N_1}$ $\lambda'_{N_2} = \sum_{ij \in N_q} \lambda_{ij} - \lambda_{N_2}$ z $D = \{\lambda_{N_1}\ \lambda_{N_2} \in \mathbb{R}^2 \quad p' > 0.5 \quad q' <$ (,), (,)。0.5}

在(,) 上递增: 或。

证明。 由于 $\frac{d\lambda'_{N_2}}{d\lambda_{N_2}} = -1$ 和 $\frac{b_2}{b_1+b_2} = \frac{\lambda_{N_1}}{\lambda_{N_1}+\lambda'_{N_2}}p + \frac{\lambda'_{N_2}}{\lambda_{N_1}+\lambda'_{N_2}}q$, 若 $b_2 < b_1$,

$$\begin{aligned} \frac{dq'}{d\lambda_{N_1}} &= \frac{p}{\lambda_{N_1}+\lambda'_{N_2}} - \frac{\lambda_{N_1}p}{(\lambda_{N_1}+\lambda'_{N_2})^2} - \frac{\lambda'_{N_2}q}{(\lambda_{N_1}+\lambda'_{N_2})^2} \\ &= \frac{\lambda_{N_1}p + \lambda'_{N_2}p - \lambda_{N_1}p - \lambda'_{N_2}q}{(\lambda_{N_1}+\lambda'_{N_2})^2} \\ &= \frac{\lambda'_{N_2}(p-q)}{(\lambda_{N_1}+\lambda'_{N_2})^2} \\ \frac{dq'}{d\lambda_{N_2}} &= \frac{-q}{\lambda_{N_1}+\lambda'_{N_2}} + \frac{\lambda'_{N_2}q}{(\lambda_{N_1}+\lambda'_{N_2})^2} + \frac{\lambda_{N_1}p}{(\lambda_{N_1}+\lambda'_{N_2})^2} \\ &= \frac{-\lambda_{N_1}q - \lambda'_{N_2}q + \lambda'_{N_2}q + \lambda_{N_1}p}{(\lambda_{N_1}+\lambda'_{N_2})^2} \\ &= \frac{\lambda_{N_1}(p-q)}{(\lambda_{N_1}+\lambda'_{N_2})^2} \end{aligned}$$

若 $b_2 \geq b_1$, 则 $\frac{dq'}{d\lambda_{N_1}} = \frac{dq'}{d\lambda_{N_2}} = 0$。

由于 $\frac{d\lambda'_{N_1}}{d\lambda_{N_1}} = -1$ 和 $\frac{b_4}{b_3+b_4} = \frac{\lambda'_{N_1}}{\lambda'_{N_1}+\lambda_{N_2}} p + \frac{\lambda_{N_2}}{\lambda'_{N_1}+\lambda_{N_2}} q$, 若 $b_4 > b_3$,

$$\begin{aligned}
\frac{dp'}{d\lambda_{N_1}} &= \frac{-p}{\lambda'_{N_1}+\lambda_{N_2}} + \frac{\lambda'_{N_1} p}{(\lambda'_{N_1}+\lambda_{N_2})^2} + \frac{\lambda_{N_2} q}{(\lambda'_{N_1}+\lambda_{N_2})^2} \\
&= \frac{-\lambda'_{N_1} p - \lambda_{N_2} p + \lambda'_{N_1} p + \lambda_{N_2} q}{(\lambda'_{N_1}+\lambda_{N_2})^2} \\
&= \frac{\lambda_{N_2}(q-p)}{(\lambda'_{N_1}+\lambda_{N_2})^2} \\
\frac{dp'}{d\lambda_{N_2}} &= \frac{q}{\lambda'_{N_1}+\lambda_{N_2}} - \frac{\lambda_{N_2} q}{(\lambda'_{N_1}+\lambda_{N_2})^2} - \frac{\lambda'_{N_1} p}{(\lambda'_{N_1}+\lambda_{N_2})^2} \\
&= \frac{\lambda'_{N_1} q + \lambda_{N_2} q - \lambda_{N_2} q - \lambda'_{N_1} p}{(\lambda'_{N_1}+\lambda_{N_2})^2} \\
&= \frac{\lambda'_{N_1}(q-p)}{(\lambda'_{N_1}+\lambda_{N_2})^2}
\end{aligned}$$

若 $b_4 \leq b_3$, $\frac{dp'}{d\lambda_{N_1}} = \frac{dp'}{d\lambda_{N_2}} = 0$。

那么,

$$\frac{dq'}{d\lambda_{N_1}} = \begin{cases} \frac{\lambda'_{N_2}(p-q)}{(\lambda_{N_1}+\lambda'_{N_2})^2} \geq 0, & \text{if } b_1 - b_2 > 0 \\ 0, & \text{otherwise} \end{cases}; \qquad \frac{dq'}{d\lambda_{N_2}} = \begin{cases} \frac{\lambda_{N_1}(p-q)}{(\lambda_{N_1}+\lambda'_{N_2})^2} \geq 0, & \text{if } b_1 - b_2 > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\frac{dp'}{d\lambda_{N_1}} = \begin{cases} \frac{\lambda_{N_2}(q-p)}{(\lambda'_{N_1}+\lambda_{N_2})^2} \leq 0, & \text{if } b_3 - b_4 < 0 \\ 0, & \text{otherwise} \end{cases}; \qquad \frac{dp'}{d\lambda_{N_2}} = \begin{cases} \frac{\lambda'_{N_1}(q-p)}{(\lambda'_{N_1}+\lambda_{N_2})^2} \leq 0, & \text{if } b_3 - b_4 < 0 \\ 0, & \text{otherwise} \end{cases}$$

由于

$$\begin{aligned}
&\frac{db_1}{d\lambda_{N_1}} = (1-p); \quad \frac{db_1}{d\lambda_{N_2}} = -(1-q) \\
&\frac{db_2}{d\lambda_{N_1}} = p; \quad \frac{db_2}{d\lambda_{N_2}} = -q \\
&\frac{db_3}{d\lambda_{N_1}} = -(1-p); \quad \frac{db_3}{d\lambda_{N_2}} = (1-q) \\
&\frac{db_4}{d\lambda_{N_1}} = -p; \quad \frac{db_4}{d\lambda_{N_2}} = q,
\end{aligned}$$

我们有

$$\begin{aligned}
\frac{dz(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_1}} =& -\frac{db_1}{d\lambda_{N_1}} \log\{1-q'(\lambda_{N_1},\lambda_{N_2})\} - b_1 \frac{d\log\{1-q'(\lambda_{N_1},\lambda_{N_2})\}}{d\lambda_{N_1}} \\
&- \frac{db_2}{d\lambda_{N_1}} \log\{q'(\lambda_{N_1},\lambda_{N_2})\} - b_2 \frac{d\log\{q'(\lambda_{N_1},\lambda_{N_2})\}}{d\lambda_{N_1}} \\
&- \frac{db_3}{d\lambda_{N_1}} \log\{1-p'(\lambda_{N_1},\lambda_{N_2})\} - b_3 \frac{d\log\{1-p'(\lambda_{N_1},\lambda_{N_2})\}}{d\lambda_{N_1}} \\
&- \frac{db_4}{d\lambda_{N_1}} \log\{p'(\lambda_{N_1},\lambda_{N_2})\} - b_4 \frac{d\log\{p'(\lambda_{N_1},\lambda_{N_2})\}}{d\lambda_{N_1}} \\
=& -(1-p)\log\{1-q'(\lambda_{N_1},\lambda_{N_2})\} + b_1(1-q'(\lambda_{N_1},\lambda_{N_2}))^{-1} \frac{dq'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_1}} \\
&- p\log\{q'(\lambda_{N_1},\lambda_{N_2})\} - b_2(q'(\lambda_{N_1},\lambda_{N_2}))^{-1} \frac{dq'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_1}} \\
&+ (1-p)\log\{1-p'(\lambda_{N_1},\lambda_{N_2})\} + b_3(1-p'(\lambda_{N_1},\lambda_{N_2}))^{-1} \frac{dp'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_1}} \\
&+ p\log\{p'(\lambda_{N_1},\lambda_{N_2})\} - b_4(p'(\lambda_{N_1},\lambda_{N_2}))^{-1} \frac{dp'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_1}}
\end{aligned}$$

$$=(1-p)\log\{\frac{1-p'(\lambda_{N_1},\lambda_{N_2})}{1-q'(\lambda_{N_1},\lambda_{N_2})}\}+p\log\{\frac{p'(\lambda_{N_1},\lambda_{N_2})}{q'(\lambda_{N_1},\lambda_{N_2})}\} \tag{22}$$

$$+\left(\frac{b_1}{1-q'(\lambda_{N_1},\lambda_{N_2})}-\frac{b_2}{q'(\lambda_{N_1},\lambda_{N_2})}\right)\frac{dq'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_1}} \tag{23}$$

$$+\left(\frac{b_3}{1-p'(\lambda_{N_1},\lambda_{N_2})}-\frac{b_4}{p'(\lambda_{N_1},\lambda_{N_2})}\right)\frac{dp'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_1}}. \tag{24}$$

那么，当 $b_1>b_2$, $q'(\lambda_{N_1},\lambda_{N_2})=\frac{b_2}{b_1+b_2}$, $1-q'(\lambda_{N_1},\lambda_{N_2})=\frac{b_1}{b_1+b_2}$ 时，我们有

$$\frac{b_1}{1-q'(\lambda_{N_1},\lambda_{N_2})}-\frac{b_2}{q'(\lambda_{N_1},\lambda_{N_2})}=\frac{b_1}{\frac{b_1}{b_1+b_2}}-\frac{b_2}{\frac{b_2}{b_1+b_2}}=0.$$

另一方面，当 $b_1\le b_2$, $\frac{dq'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_1}}=0$。对最后一项进行类似的计算可知，行（23）和（24）为 0。

接下来，我们分析第（22）行中的各项，

$$\begin{aligned}&(1-p)\log\{\frac{1-p'(\lambda_{N_1},\lambda_{N_2})}{1-q'(\lambda_{N_1},\lambda_{N_2})}\}+p\log\{\frac{p'(\lambda_{N_1},\lambda_{N_2})}{q'(\lambda_{N_1},\lambda_{N_2})}\}\\&=\log\underbrace{\left(\frac{1-p'(\lambda_{N_1},\lambda_{N_2})}{1-q'(\lambda_{N_1},\lambda_{N_2})}\right)^{(1-p)}}_{<1}+\log\underbrace{\left(\frac{p'(\lambda_{N_1},\lambda_{N_2})}{q'(\lambda_{N_1},\lambda_{N_2})}\right)^{p}}_{>1}\end{aligned}$$

注意，$p'\in[0.5,p]$, $q'\in[q,0.5]$, $1-p'\in[1-p,0.5]$ 和 $1-q'\in[0.5,1-q]$。为简化表达式，暂且忽略其对 λ_{N_1} 和 λ_{N_2} 的依赖。由于 log 是递增函数，要证明该表达式为正，我们需要

$$\begin{aligned}\log\left(\frac{p'}{q'}\right)^p\ge-\log\left(\frac{1-p'}{1-q'}\right)^{(1-p)}&\iff\log\left(\frac{p'}{q'}\right)^p\ge\log\left(\frac{1-q'}{1-p'}\right)^{(1-p)}\\&\iff\left(\frac{p'}{q'}\right)^p\ge\left(\frac{1-q'}{1-p'}\right)^{(1-p)}\\&\iff(1-p')^{(1-p)}(p')^p\ge(1-q')^{(1-p)}(q')^p\end{aligned}$$

由于对于 $g(x)=(1-x)^{(1-p)}x^p$, $\forall x\in[0,0.5]$, $g(x)\in[0,0.5]$, $0.5\ge(1-q')^{(1-p)}(q')^p$。接着，考虑 $f(x)=\log\{(1-x)^{(1-p)}(x)^p\}=(1-p)\log(1-x)+p\log x$，其中

$$f'(x)=\frac{-(1-p)}{1-x}+\frac{p}{x}.$$

对于 $p\ge x\ge0.5$,

$$\frac{p}{1-p}\ge\frac{x}{1-p}\ge\frac{x}{1-x}\implies p(1-x)\ge x(1-p)\implies\frac{p}{x}\ge\frac{1-p}{1-x}\implies f'(x)\ge0.$$

于是，我们有

$$f(x)\ge f(0.5)=\log0.5\implies(1-x)^{(1-p)}x^p\ge0.5$$

因此，

$$(1-p')^{(1-p)}(p')^p\ge0.5\ge(1-q')^{(1-p)}(q')^p$$

and

$$\frac{dz(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_1}}\ge0.$$

注意，对于 $p'>0.5$ 或 $q'<0.5$，上述等式严格成立。

对 λ_{N_2} 重复上述过程，

$$\begin{aligned}\frac{dz(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_2}}=&-\frac{db_1}{d\lambda_{N_2}}\log\{1-q'(\lambda_{N_1},\lambda_{N_2})\}-b_1\frac{d\log\{1-q'(\lambda_{N_1},\lambda_{N_2})\}}{d\lambda_{N_2}}\\&-\frac{db_2}{d\lambda_{N_2}}\log\{q'(\lambda_{N_1},\lambda_{N_2})\}-b_2\frac{d\log\{q'(\lambda_{N_1},\lambda_{N_2})\}}{d\lambda_{N_2}}\end{aligned}$$

$$
\begin{aligned}
&- \frac{db_3}{d\lambda_{N_2}} \log\{1-p'(\lambda_{N_1},\lambda_{N_2})\} - b_3 \frac{d\log\{1-p'(\lambda_{N_1},\lambda_{N_2})\}}{d\lambda_{N_2}} \\
&- \frac{db_4}{d\lambda_{N_2}} \log\{p'(\lambda_{N_1},\lambda_{N_2})\} - b_4 \frac{d\log\{p'(\lambda_{N_1},\lambda_{N_2})\}}{d\lambda_{N_2}} \\
=&(1-q)\log\{1-q'(\lambda_{N_1},\lambda_{N_2})\} + b_1(1-q'(\lambda_{N_1},\lambda_{N_2}))^{-1}\frac{dq'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_2}} \\
&+ q\log\{q'(\lambda_{N_1},\lambda_{N_2})\} - b_2(q'(\lambda_{N_1},\lambda_{N_2}))^{-1}\frac{dq'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_2}} \\
&- (1-q)\log\{1-p'(\lambda_{N_1},\lambda_{N_2})\} + b_3(1-p'(\lambda_{N_1},\lambda_{N_2}))^{-1}\frac{dp'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_2}} \\
&- q\log\{p'(\lambda_{N_1},\lambda_{N_2})\} - b_4(p'(\lambda_{N_1},\lambda_{N_2}))^{-1}\frac{dp'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_2}} \\
=&(1-q)\log\{\frac{1-q'(\lambda_{N_1},\lambda_{N_2})}{1-p'(\lambda_{N_1},\lambda_{N_2})}\} + q\log\{\frac{q'(\lambda_{N_1},\lambda_{N_2})}{p'(\lambda_{N_1},\lambda_{N_2})}\} \qquad (25)\\
&+ \left(\frac{b_1}{1-q'(\lambda_{N_1},\lambda_{N_2})} - \frac{b_2}{q'(\lambda_{N_1},\lambda_{N_2})}\right)\frac{dq'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_2}} \qquad (26)\\
&+ \left(\frac{b_3}{1-p'(\lambda_{N_1},\lambda_{N_2})} - \frac{b_4}{p'(\lambda_{N_1},\lambda_{N_2})}\right)\frac{dp'(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_2}}. \qquad (27)
\end{aligned}
$$

类似地，式 (25) 和 (26) 为 0，我们像之前一样分析式 (27),

$$
\begin{aligned}
&(1-q)\log\{\frac{1-q'(\lambda_{N_1},\lambda_{N_2})}{1-p'(\lambda_{N_1},\lambda_{N_2})}\} + q\log\{\frac{q'(\lambda_{N_1},\lambda_{N_2})}{p'(\lambda_{N_1},\lambda_{N_2})}\} \\
=&\log\underbrace{\left(\frac{1-q'(\lambda_{N_1},\lambda_{N_2})}{1-p'(\lambda_{N_1},\lambda_{N_2})}\right)^{(1-q)}}_{>1} + \log\underbrace{\left(\frac{q'(\lambda_{N_1},\lambda_{N_2})}{p'(\lambda_{N_1},\lambda_{N_2})}\right)^{q}}_{<1} \geq 0 \\
\iff& \left(\frac{1-q'}{1-p'}\right)^{(1-q)} \geq \left(\frac{p'}{q'}\right)^{q} \iff (1-q')^{(1-q)}(q')^q \geq (p')^q(1-p')^{(1-q)}.
\end{aligned}
$$

于是，对于 $g(x) = (1-x)^{(1-q)}x^q, \forall x \in [0.5,1]$, $g(x) \in [0,0.5]$, $0.5 \geq (p')^q(1-p')^{(1-q)}$.
再次考虑 $f(x) = \log\{(1-x)^{(1-q)}(x)^q\} = (1-q)\log(1-x) + q\log x$, 其中

$$f'(x) = \frac{-(1-q)}{1-x} + \frac{q}{x}$$

对于 $0.5 \geq x \geq q$,

$$\frac{1-q}{q} \geq \frac{1-x}{q} \geq \frac{1-x}{x} \implies x(1-q) \geq q(1-x) \implies \frac{1-q}{1-x} \geq \frac{q}{x} \implies f'(x) \leq 0$$

于是，我们有

$$f(x) \geq f(0.5) = \log 0.5 \implies (1-x)^{(1-q)}(x)^q \geq 0.5$$

因此,

$$(1-q')^{(1-q)}(q')^q \geq 0.5 \geq (p')^q(1-p')^{(1-q)}$$

and

$$\frac{dz(\lambda_{N_1},\lambda_{N_2})}{d\lambda_{N_2}} \geq 0.$$

类似地，我们注意到 上述结论在 $p' > 0.5$ 或 $q' < 0.5$ 的情况下严格成立。因此，证明至此结束 □

.引理 3. 对于固定的 $\lambda \in \mathcal{P}_I$、 $C \in \mathcal{C}^K$,

$$\min_{[C']\in \mathrm{Alt}(C)/\sim} z(\lambda_{N_1},\lambda_{N_2}) = \begin{cases} \min_{[C']\in(\mathrm{Alt}(C)\cap(\mathcal{C}^{K-1}\cup\mathcal{C}^{K+1}))/\sim} z(\lambda_{N_1},\lambda_{N_2}), & M > K > 1 \\ \min_{[C']\in(\mathrm{Alt}(C)\cap\mathcal{C}^{K+1})/\sim} z(\lambda_{N_1},\lambda_{N_2}), & K = 1 \\ \min_{[C']\in(\mathrm{Alt}(C)\cap\mathcal{C}^{K-1})/\sim} z(\lambda_{N_1},\lambda_{N_2}), & K = M, \end{cases}$$

其中 $\lambda_{N_1} = \sum_{(i,j)\in\Lambda_{N_1}([C'])} \lambda_{ij}$ 和 $\lambda_{N_2} = \sum_{(i,j)\in\Lambda_{N_2}([C'])} \lambda_{ij}$。

证明。我们将考虑 $[C'] \in \mathsf{Alt}(C)/\sim$ 的 3 种情形，并证明我们总能找到某个 $[C] \in (\mathsf{Alt}(C) \cap (\mathcal{C}^{K-1} \cup \mathcal{C}^{K+1}))/\sim$，使得 $\lambda_{\bar{N}_1} + \lambda_{\bar{N}_2} = \sum_{(i,j)\in\Lambda_{N_1}([\bar{C}])} \lambda_{ij} + \sum_{(i,j)\in\Lambda_{N_2}([\bar{C}])} \lambda_{ij} \le \lambda_{N_1} + \lambda_{N_2}$。

情况1：考虑 $M \ge K \ge 1$ 和 $[C'] \in (\mathsf{Alt}(C) \cap \mathcal{C}^K)/\sim$，这是一个与 C 具有相同簇数的替代聚类，K。那么，存在一个 g_k，使得对于所有 $g'_{k'}$，$g_k \not\subset g'_{k'}$。（因为如果对于每个 g_k，都存在一个 $g'_{k'}$，使得$g_k \subset g'_{k'}$，并且由于 $[C'] \in \mathsf{Alt}(C)/\sim$，必定存在一个 g_j，使得对于所有 $g'_{k'}$，$g_j \neq g'_{k'}$，那么项目数必定超过M，这就构成了矛盾。）于是，我们可以选择一个 $g'_{k'}$，使得 $g_k \cap g'_{k'} \neq \emptyset$ 且 $g_k \setminus (g_k \cap g'_{k'}) \neq \emptyset$。令 $N'_1 = \{(i,j) : i \in g_k \cap g'_{k'}, j \in g_k \setminus (g_k \cap g'_{k'})\ j \ge i\} \cup \{(j,i) : i \in g_k \cap g'_{k'}, j \in g_k \setminus (g_k \cap g'_{k'})\ i \ge j\} \subset \Lambda_{N_1}([C']) \cup \Lambda_{N_2}([C'])$。现在，我们可以找到一个 $[\bar{C}] \in (\mathsf{Alt}(C) \cap \mathcal{C}^{K+1})/\sim$，其中 $\bar{g}_j = g_j$、$\forall j \neq k$、$\bar{g}_k = g_k \setminus (g_k \cap g'_{k'})$ 且 $\bar{g}_{K+1} = g_k \cap g'_{k'}$。因此，我们有 $\Lambda_{N_1}([C]) \cup \Lambda_{N_2}([C]) = N'_1 \subset \Lambda_{N_1}([C']) \cup \Lambda_{N_2}([C'])$ 且 $\lambda_{\bar{N}_1} + \lambda_{\bar{N}_2} \le \lambda_{N_1} + \lambda_{N_2}$。

情况 2：考虑 $M \ge K > 2$ 和 $[C'] \in (\mathsf{Alt}(C) \cap \mathcal{C}^{K-2})/\sim$，这是一个与 C 不同、包含 $K-2$ 个簇的替代聚类。于是，存在一个 g_k，使得 $g_k \subset \bigcup_{k'\in J} g'_{k'}$对某个指标集 $J \subset [K-2]$ 以及某个 $k' \in J$, $|g'_{k'}| > 1$成立。（因为每个 g_k都是足够多个 $g'_{k'}$的子集，并且由于 $K-2 < M$，至少有一个 $g'_{k'}$必须包含多个项目。）现在，我们可以找到一个 $[\bar{C}] \in (\mathsf{Alt}(C) \cap \mathcal{C}^{K-1})/\sim$，其中 $\bar{g}_{K-1} = g_k$ 且 $\bar{g}_{k'} = g'_{k'} \setminus (g'_{k'} \cap g_k) \neq \emptyset$。然后，我们也有 $\Lambda_{N_1}([C]) \cup \Lambda_{N_2}([C]) \subset \Lambda_{N_1}([C']) \cup \Lambda_{N_2}([C'])$，并且 $\lambda_{\bar{N}_1} + \lambda_{\bar{N}_2} \le \lambda_{N_1} + \lambda_{N_2}$。

情况3：考虑 $M-1 > K \ge 1$ 和 $[C'] \in (\mathsf{Alt}(C) \cap \mathcal{C}^{K+2})/\sim$，这是一个相对于 C 的替代聚类，其中包含 $K+2$ 个簇。那么，存在一个 g_k，使得对于所有 $g'_{k'}$，$g_k \not\subset g'_{k'}$均成立。（因为如果对于所有 g_k，都存在一个 $g'_{k'}$ 使得 $g_k \subset g'_{k'}$成立，那么项目数必定超过 M，或者簇数至多为 K，这就产生了矛盾。）现在，类似于情况1，我们可以找到一个包含 $K+1$ 个簇的 $[C] \in \mathsf{Alt}(C)/\sim$，使得 $\lambda_{\bar{N}_1} + \lambda_{\bar{N}_2} \le \lambda_{N_1} + \lambda_{N_2}$。

对于包含多于 $K+2$ 个且少于 $K-2$ 个簇的 $[C'] \in \mathsf{Alt}(C)/\sim$，也可以得到类似的结果。于是，由于 $\lambda_{\bar{N}_1} + \lambda_{\bar{N}_2} \le \lambda_{N_1} + \lambda_{N_2} \implies z(\lambda_{\bar{N}_1}, \lambda_{\bar{N}_2}) \le z(\lambda_{N_1}, \lambda_{N_2})$，根据这3种情形，如果 $M > K > 1$，我们只需考虑 $[C'] \in (\mathsf{Alt}(C) \cap (\mathcal{C}^{K-1} \cup \mathcal{C}^{K+1}))/\sim$来求解最小值。此外，对于 $K = M$，我们只需考虑$[C'] \in (\mathsf{Alt}(C) \cap \mathcal{C}^{M-1})/\sim$；对于 $K = 1$，我们只需考虑 $[C'] \in (\mathsf{Alt}(C) \cap \mathcal{C}^2)/\sim$。 □

C. 命题 1 的证明

证明。我们首先考虑所有 M 项目都属于同一个簇的情况。根据引理 1、3以及定理 2证明中的情况1，我们知道

$$\max_{\lambda\in\mathcal{P}_I} \inf_{C'\in\mathsf{Alt}_1(C)} h(\lambda, C') = \max_{\lambda\in\mathcal{P}_I} \min_{[C']\in(\mathsf{Alt}_1(C)\cap\mathcal{C}^2)/\sim} z(\lambda_{N_1}, \lambda_{N_2}), \tag{28}$$

其中 $\lambda_{N_1} = \sum_{(i,j)\in\Lambda_{N_1}([C'])} \lambda_{ij}$ 且 $\lambda_{N_2} = \sum_{(i,j)\in\Lambda_{N_2}([C'])} \lambda_{ij}$。由于 $\Lambda_{N_2}([C']) = \emptyset$，对于所有 $[C'] \in (\mathsf{Alt}_1(C) \cap \mathcal{C}^2)/\sim$，其中 $N_q = \emptyset$，我们只需考虑 λ_{N_1}；根据引理 2，$z(\lambda_{N_1}, \lambda_{N_2})$ 在 λ_{N_1} 上递增。因此，我们只需在 λ_{N_1} 上求解 (28)，而非在 z 上求解。

为使表述清晰，我们省略 C' 和 $\mathsf{Alt}_1(C) \cap \mathcal{C}^2$ 的等价类记号，并用每个 C' 所属较小簇的大小及其中的项目来表示它。然后，我们可以将 $\mathsf{Alt}_1(C) \cap \mathcal{C}^2$ 划分为较小簇中有 i 个项目的集合，其中 $i \in [\lfloor M/2 \rfloor]$。对于 $C' \in \mathsf{Alt}_1(C) \cap \mathcal{C}^2$，重新标记 $g_1 = \arg\min\{|g'_1|, |g'_2|\}$ 和 $g_2 = \arg\max\{|g'_1|, |g'_2|\}$。我们沿用 f 的记号，将每个 C' 映射到相应的 g_1，即 $f : \mathsf{Alt}_1(C) \cap \mathcal{C}^2 \to \bigcup_{C'\in\mathsf{Alt}_1(C)\cap\mathcal{C}^2} \{g_1\}$；对于所有此类 g_1，我们只需考虑 λ_{N_1}。注意，g_1 取决于 C'，因为 g_1 表示 C' 中的簇分配。令 $\mathcal{G}_1 = \bigcup_{C'\in\mathsf{Alt}_1(C)\cap\mathcal{C}^2} \{g_1\}$。然后，我们可以将 $\mathcal{G}_1$ 划分为由恰好包含 i 个项目的 g_1 构成的集合，$\mathcal{G}_1 = \bigcup_{i\in[\lfloor M/2 \rfloor]} \{g_1 \in \mathcal{G}_1 : |g_1| = i\} := \bigcup_{i\in[\lfloor M/2 \rfloor]} \mathcal{G}_1^i$。

对于 $g_1 \in \mathcal{G}_1^i$, $|\Lambda_{N_1}(f^{-1}(g_1))| = i(M-i)$。[3]根据双射映射 f，我们还可以得到一个双射（略微滥用记号），$\lambda_{N_1} : \mathcal{G}_1 \to \{\sum_{(i,j)\in\Lambda_{N_1}(f^{-1}(g_1))} \lambda_{ij}\}$。如果 λ 是均匀的，那么 $\min_{g_1\in\mathcal{G}_1} \lambda_{N_1}(g_1) = (M-1)/\binom{M}{2} \in \lambda_{N_1}(\mathcal{G}_1^1)$，因为 $\min_{i\in[\lfloor M/2 \rfloor]} i(M-i) = M-1$。这里，$\lambda_{N_1}$ 将 $g_1 \mapsto \sum_{(i,j)\in\Lambda_{N_1}(f^{-1}(g_1))} \lambda_{ij}$映射到。

反设最优的 λ不均匀，则会导出矛盾。于是，$\min_{g_1\in\mathcal{G}_1} \lambda_{N_1}(g_1) > (M-1)/\binom{M}{2}$，并令 $\epsilon' = \min_{g_1\in\mathcal{G}_1} \lambda_{N_1}(g_1) - (M-1)/\binom{M}{2} > 0$。此外，令 $i \ge 2$ 为使最小值在某个 $g_1 \in \mathcal{G}_1^i$ 处取得的指标。不失一般性，令 $\lambda_{ij} = \frac{M-1}{i(M-i)\binom{M}{2}} + \frac{\epsilon'}{i(M-i)}$ 对所有 $(i,j) \in \Lambda_{N_1}(f^{-1}(g_1))$成立。

在接下来的2段中，为便于记法，我们将采用如下约定：(i,j) 索引项目 $i \in g_1$ 和项目$j \in g_2$，而不必担心 $j > i$。我们可以这样做，因为对于每个满足 $j < i$的(i,j)，$(j,i) \in I$，并且我们可以假定改为引用该索引。

[3]将每个表示 g_1 的矩阵 C' 视为最简单分块矩阵的一个排列，其中项目 $1, \cdots, i$ 位于 g_1

考虑 $g_1 \supset g_1' = \{i_1\} \in \mathcal{G}_1^1$和 $\lambda_{N_1}(g_1')$。然后，对于 $j \in g_2 \subset g_2'$的所有(i_1, j)，$(i_1, j) \in \Lambda_{N_1}(f^{-1}(g_1'))$，并且我们可以划分
$\Lambda_{N_1}\ f^{-1}\ g_1 \ = \{\ i_1\ j \quad j \in g_2\} \cup \{\ i_1\ j \quad j \in [M] \setminus \{i_1\} \cup g_2\ \}$ $\lambda_{N_1}\ g_1' \ = \ {}_{i}^{M-1}{}_{\frac{M}{2}} + {}_{i}^{\epsilon'} + (\ (\ ))\ (,) : (,) : (\ )$。因此，$(\ )$
$\sum_{\{\ i_1\ j \quad j \in [M] \setminus \{i_1\} \cup g_2\ \}} \lambda_{i_1\ j}$ $g_1 \in \mathcal{G}_1\ \lambda_{N_1}\ g_1 \ = \lambda_{N_1}\ g_1 \ = \ ^{M-1} + \epsilon'$ $\underline{\lambda_{N_1}\ g_1} \ \leq \lambda_{N_1}\ g_1'$
$(\ )\ (,):(),$且$\min(\ )\ (\ )\ (\ ^{M}_{2})$。由于$(\ )\ (\ )$，

$$\sum_{\{(i_1,j):j\in[M]\setminus(\{i_1\}\cup g_2)\}} \lambda_{(i_1,j)} \geq \frac{(i-1)(M-1)}{i\binom{M}{2}} + \frac{(i-1)\epsilon'}{i}.$$

对 $i_k = i_1, i_2, \cdots, i_i \in g_1$这样做，我们有

$$\sum_{\{(i_k,j):j\in[M]\setminus(\{i_k\}\cup g_2)\}} \lambda_{i_k,j} \geq \frac{(i-1)(M-1)}{i\binom{M}{2}} + \frac{(i-1)\epsilon'}{i},$$

其中，我们在此类求和中将 g_1 中的每个项目对计数了两次。

接下来，考虑$g_2 \supset g_1'' = \{j_1\} \in \mathcal{G}_1^1$。类似地，我们可以划分$N_1'' = \{(i, j_1) :\ i \in g_1\} \cup \{(i, j_1) :\ i \in [M] \setminus (\{j_1\} \cup g_1)\}$和 $\lambda_{N_1}(g_1'') = \ ^{M-1}{}_{(M-i)(\frac{M}{2})} + {}_{M-i}^{\epsilon'} + \sum_{(i,j_1):i\in[M]\setminus(\{j_1\}\cup g_1)} \lambda_{(i,j_1)}$。由于 $\lambda_{N_1}(g_1) \leq \lambda_{N_1}(g_1'')$，

$$\sum_{(i,j_1):i\in[M]\setminus(\{j_1\}\cup g_1)} \lambda_{i,j_1} \geq \frac{(M-1)(M-i-1)}{(M-i)\binom{M}{2}} + \frac{(M-i-1)\epsilon'}{M-i}.$$

同样，对 $j_k = j_1, j_2, \cdots, j_{M-i} \in g_2$这样做，我们有

$$\sum_{(i,j_k):i\in[M]\setminus(\{j_k\}\cup g_1)} \lambda_{i,j_k} \geq \frac{(M-1)(M-i-1)}{(M-i)\binom{M}{2}} + \frac{(M-i-1)\epsilon'}{M-i}.$$

也就是说，在这样的求和中，我们会将 g_2 中的每一对项目都计数两次。

然后，我们注意到 $N_1' := \Lambda_{N_1}(f^{-1}(g_1)) \subset g_1 \times g_1$，且 $N_k'' := \{(i, j_k) :\ i \in [M] \setminus (\{j_k\} \cup g_1)\} \subset g_2 \times g_2$。因此，它们不相交，且

$$\sum_{(i,j)\in N_1\cup\bigcup_{k\in[i]}\tilde{N}_k'\cup\bigcup_{k\in[M-i]}\tilde{N}_k''} \lambda_{(i,j)} \geq \frac{M-1}{\binom{M}{2}} + \epsilon' + \frac{(i-1)(M-1)}{\binom{M}{2}} + (i-1)\epsilon' + \frac{(M-1)(M-i-1)}{\binom{M}{2}} + (M-i-1)\epsilon',$$

其中 $\bigcup_{k\in[i]} \tilde{N}_k' = \{\ i\ j \ \in g_1 \times g_1 \quad j \neq i\}$ $\bigcup_{k\in[M-i]} \tilde{N}_k'' = \{\ i\ j \ \in g_2 \times g_2 \quad j \neq i\}$ $\tilde{g}_1^2 = \{\ i\ j \ \in (,)$和$(,) :$。
$g_1 \times g_1 \quad j > i\}$ $\tilde{g}_2^2 = \{\ i\ j \ \in g_2 \times g_2 \quad j > i\}$
然后，令 $(,) :$和 $(,) :$，使得我们有

$$\begin{aligned}
&\sum_{(i,j)\in N_1\cup\tilde{g}_1^2\cup\tilde{g}_2^2} \lambda_{ij} \\
&\geq \frac{M-1}{\binom{M}{2}} + \epsilon' + \frac{(i-1)(M-1)}{2\binom{M}{2}} + \frac{(i-1)\epsilon'}{2} + \frac{(M-1)(M-i-1)}{2\binom{M}{2}} + \frac{(M-i-1)\epsilon'}{2} \\
&= \frac{M-1}{\binom{M}{2}} + \epsilon' + \frac{(M-2)(M-1)}{2\binom{M}{2}} + \frac{(M-2)\epsilon'}{2} \\
&= \frac{(M-1)M/2}{\binom{M}{2}} + \epsilon' + \frac{(M-2)\epsilon'}{2} \\
&= 1 + \epsilon' + \frac{(M-2)\epsilon'}{2} > 1.
\end{aligned}$$

由于 $\sum_{(i,j)\in N_1\cup\tilde{g}_1^2\cup\tilde{g}_2^2} \lambda_{i,j} = \sum_{(i,j)\in I} \lambda_{i,j} = 1$，我们得出矛盾。

然后，假设存在 $i = 1$，使得对于某个 $g_1 \in \mathcal{G}_1^1$，$\min_{g_1\in\mathcal{G}_1} \lambda_{N_1}(g_1) = \lambda_{N_1}(g_1)$。不失一般性，令
$g_1 = \{1\}$ $g_1 \in \mathcal{G}_1\ \lambda_{N_1}\ g_1 \ = \ ^{M-1}_{\frac{M}{2}} + \epsilon'$ $\epsilon' > 0$ $|\mathcal{G}_1^1| = M$ $\mathcal{G}_1^1 = \{\{k\} \quad k \in$，并且对于
$[M]\}$ $N_1^k = \Lambda_{N_1}\ f^{-1}\ \{k\}$ $i\ j \ \in I$ N_1^i N_1^j
某个，$\min(\ )\ (\ )$。我们注意到，由于$:$。对于$: (\ (\ ))$，因为每一对 $(,)$ 恰好出现两次，分别在 和 中，我们必有$\sum_{k\in[M]} \lambda_{N_1}(\{k\})$
$= 2$。根据 $\min_{g_1\in\mathcal{G}_1} \lambda_{N_1}(g_1)$ 的定义，对于所有 $k \in [M] \setminus \{1\}$，$\lambda_{N_1}(\{k\}) \geq \lambda_{N_1}(\{1\}) = \ ^{M-1}(\ ^{M}_{2}) + \epsilon'$。因此，$\sum_{k\in[M]} \lambda_{N_1}(\{k\})$
$\geq \ ^{M}(M-1)(\ ^{M}_{2}) + M\epsilon' > 2$，从而导致矛盾。

接下来，考虑所有 M 项目都属于不同簇的情况，即 $K = M$。那么，与之前类似，我们知道

$$\max_{\lambda\in\mathcal{P}_I} \inf_{C'\in \mathrm{Alt}_1(C)} h(\lambda, C') = \max_{\lambda\in\mathcal{P}_I} \min_{[C']\in(\mathrm{Alt}_1(C)\cap\mathcal{C}^{M-1})/\sim} z(\lambda_{N_1}, \lambda_{N_2}),$$

并且我们只需考虑对 λ_{N_2} 进行最小化。容易看出，对于所有 $[C'] \in (\mathrm{Alt}_1(C) \cap \mathcal{C}^{M-1})/\sim$, $\Lambda_{N_1}([C']) = \emptyset$以及 $\Lambda_{N_2}([C']) = \{(i,\ j)\}$, 其中项目 i和 j合并形成一个簇，且 $\bigcup_{[C'] \in (\mathrm{Alt}_1(C) \cap \mathcal{C}^{M-1})/\sim} \Lambda_{N_2}([C']) = I$。如果

λ^* $i\ j\ \in I$ $\lambda_{ij} < 1/\binom{M}{2}$ $[C'] \in \quad {}_1\ C \cap \mathcal{C}^{M-1}\ /\sim \lambda_{N_2} <$不是均匀分 $1/\binom{M}{2}$ $[C'] \in \quad {}_1\ C \cap \mathcal{C}^{M-1}\ /\sim \lambda_{N_2} = 1/\binom{M}{2}$

布，则存在一个 (,)，使得 且 $\min(\mathrm{Alt}())$。由于均匀分布可达到 $\min(\mathrm{Alt}())$，因此它应当是最优分布。 □

附录 C 第 IV 节结果的证明

A. 命题 2 的证明

证明。根据强大数定律，事件 $\mathcal{E} = \{\omega \in \Omega : \lim_{t\to\infty} \hat{C}_t(\omega) = C\}$ 的概率为 1。对于 $\omega \in \mathcal{E}, \forall \mu > 0, \exists t_\mu$，使得 $\forall t \geq t_\mu, |\hat{c}_t(\omega)_{ij} - c_{ij}| < \mu, \forall (i,j) \in I$。令 $\mu = \min\{p - 0.5, 0.5 - q\}$，那么若 $c_{ij} = p$，则有 $\hat{c}_t(\omega)_{ij} - 0.5 > -\mu + c_{ij} - 0.5 = -\mu + p - 0.5 \geq 0$；若 $c_{ij} = q$，则有 $\hat{c}_t(\omega)_{ij} - 0.5 < \mu + c_{ij} - 0.5 = \mu + q - 0.5 \leq 0$。然后，由于在算法1中，我们设置 $(\hat{c}_{t_=})_{ij} = \mathbb{1}\{(\hat{c}_t)_{ij} \geq 0.5\}$ 和 $(\hat{c}_{t_\neq})_{ij} = \mathbb{1}\{(\hat{c}_t)_{ij} < 0.5\}$，求解 (7) 时，我们精确地恢复 $c_{t_=}$ 为 $c_=$，并恢复 $c_{t_\neq}$ 为 $c_\neq$，且 $[C_t(\omega)] = [C]$。此外，

$$
\begin{aligned}
|p_t - p| &= \left| \frac{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_=})_{ij} = 1\}(\hat{c}_t)_{ij} \cdot N_{ij}(t)}{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_=})_{ij} = 1\} N_{ij}(t)} - p \right| \\
&= \left| \frac{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_=})_{ij} = 1\}(\hat{c}_t)_{ij} \cdot N_{ij}(t)}{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_=})_{ij} = 1\} N_{ij}(t)} - \frac{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_=})_{ij} = 1\} N_{ij}(t)}{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_=})_{ij} = 1\} N_{ij}(t)} p \right| \\
&= \left| \frac{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_=})_{ij} = 1\} N_{ij}(t) \cdot ((\hat{c}_t)_{ij} - p)}{\sum_{(i,j)\in I} \mathbb{1}\{(c_{t_=})_{ij} = 1\} N_{ij}(t)} \right| \\
&< \mu.
\end{aligned}
$$

对 q_t 进行类似的计算，也可得到 $|q_t - q| < \mu$。

接下来，对于 $\omega \in \mathcal{E}$ 和 $\mu > 0$，由于根据引理 4，$\lambda^*(\sigma; C) = \bar{g}_\sigma(p,q)$ 在 D 上关于 $x = (p,q)$ 连续，若 $[C_t(\omega)] = [C]$，则存在一个 $\min\{p - 0.5, 0.5 - q\} > \eta > 0$，使得对于所有 $t \geq t_\eta$ 和 $y_t = (p_t, q_t)$，都有 $\|y_t - x\| < \eta$，进而有 $\|\lambda^*(\sigma;\ C_t(\omega)) - \lambda^*(\sigma; C))\| < \mu/3(\binom{M}{2} - 1)(1-\epsilon)$。由此可得

$$
\begin{aligned}
\|\lambda^*_\epsilon(\sigma; C_t(\omega)) - \lambda^*_\epsilon(\sigma; C))\| &= \left\| (1-\epsilon)\lambda^*(\sigma; C_t(\omega)) + \epsilon\frac{1}{|I|} - (1-\epsilon)\lambda^*(\sigma; C)) - \epsilon\frac{1}{|I|} \right\| \\
&= (1-\epsilon)\|\lambda^*(\sigma; C_t(\omega)) - \lambda^*(\sigma; C))\| \\
&< \frac{\mu}{3\left(\binom{M}{2} - 1\right)}.
\end{aligned}
$$

为简化记号，我们省略 ω 项，因为它已由 C_t 隐式包含。根据 [23, 引理 17],，存在一个 $t_\mu \geq t_\eta$，使得对于所有 $t \geq t_\mu$，都有 $\| {}^{N(t)}_{t} — \lambda^*_\epsilon(\sigma;\ C))\| < \mu$。

□

引理 4. 对于固定的 $[C]$，令 $C \in [C]$ 和定义域 $D = (0.5, 1] \times [0, 0.5)$。定义 $\bar{g}:\ \ D \to \mathscr{P}(\mathcal{P}_I)$，使得

$$
\bar{g}_\sigma(p,q) := \arg\max_{\lambda \in \mathcal{P}_I} \min_{C' \in \min(C)/\sim} h(\lambda, C'; p, q) - \sigma R(\lambda),
$$

其中 $R(\lambda) = \frac{1}{2}\|\lambda\|^2$ 是严格凸的，且 $\sigma > 0$。那么，$\bar{g}_\sigma$ 在 D 上连续。

证明。 根据 $h(\lambda, C')$的定义，

$$
\begin{aligned}
h(\lambda, C') = {} & d\left(p, \min\left\{0.5, \frac{\lambda_{N_1}}{\lambda_{N_1} + \lambda'_{N_2}} p + \frac{\lambda'_{N_2}}{\lambda_{N_1} + \lambda'_{N_2}} q\right\}\right) \lambda_{N_1} \\
& + d\left(q, \min\left\{0.5, \frac{\lambda_{N_1}}{\lambda_{N_1} + \lambda'_{N_2}} p + \frac{\lambda'_{N_2}}{\lambda_{N_1} + \lambda'_{N_2}} q\right\}\right) \lambda'_{N_2} \\
& + d\left(q, \max\left\{0.5, \frac{\lambda'_{N_1}}{\lambda'_{N_1} + \lambda_{N_2}} p + \frac{\lambda_{N_2}}{\lambda'_{N_1} + \lambda_{N_2}} q\right\}\right) \lambda_{N_2} \\
& + d\left(p, \max\left\{0.5, \frac{\lambda'_{N_1}}{\lambda'_{N_1} + \lambda_{N_2}} p + \frac{\lambda_{N_2}}{\lambda'_{N_1} + \lambda_{N_2}} q\right\}\right) \lambda'_{N_1},
\end{aligned}
$$

其中$\lambda_{N_1} = \sum_{(i,j)\in\Lambda_{N_1}(C')} \lambda_{ij}$，$\lambda_{N_2} = \sum_{(i,j)\in\Lambda_{N_2}(C')} \lambda_{ij}$，$\lambda'_{N_1} = \sum_{(i,j)\in N_p} \lambda_{ij} - \lambda_{N_1}$以及$\lambda'_{N_2} = \sum_{(i,j)\in N_q} \lambda_{ij} - \lambda_{N_2}$，显然$(\lambda, p, q) \to h(\lambda, C';\ p, q)$在固定 C'时于 $\mathcal{P}_I \times D$上连续。然后，由于我们在$\min(C)/\sim$上对 $h(\lambda, C';\ p, q)$进行有限最小值运算，且 $R(\lambda)$连续，故$\min_{C'\in\min(C)/\sim} h(\lambda, C';\ p, q) - \sigma R(\lambda)$同样连续。此外，由于 $\mathcal{P}_I$ 不依赖于 p 或 q 且为紧集，我们可以应用贝尔热最大值定理 [48] 来证明 $\bar{g}_\sigma(p, q)$是上半连续的，又因其为单值映射，故连续。□

B. 命题*3*的证明

证明。 我们只需对错误概率作如下界定，

$$\begin{aligned}
\Pr(\tau_\delta < \infty, C_{\tau_\delta} \not\sim C) &\le \Pr(\exists t \in \mathbb{N} : C_t \not\sim C, Z(t) > \beta(t,\delta)) \\
&= \Pr(\exists t \in \mathbb{N} : C \in \mathrm{Alt}(C_t), Z(t) > \beta(t,\delta)) \\
&\le \Pr\left(\exists t \in \mathbb{N} : \sum_{(i,j)\in I} N_{ij}(t) d((\hat{c}_t)_{ij}, c_{ij}) > \beta(t,\delta)\right) \\
&\le \delta,
\end{aligned}$$

其中，最后一个不等式由 [45, 定理 7], 和 $x = \ln(1/\delta)$ 得出。□

C. 定理*3*的证明

证明。 首先，根据命题 2，在 D-Tracking 规则下，对所有 $(i, j) \in I$, $N_{ij}(t) \ge (\sqrt{t} - \binom{M}{2}/2)_+ - 1$，并且根据 [23,引理 17], $\forall \mu > 0$, $\exists t'_\mu \ge t_\mu$，使得

$$\sup_{t\ge t_\mu} \max_{(i,j)\in I} |\lambda^*_\epsilon(\sigma; C_t)_{ij} - \lambda^*_\epsilon(\sigma; C)_{ij}| < \mu \implies \sup_{t\ge t'_\mu} \max_{(i,j)\in I} \left|\frac{N_{ij}(t)}{t} - \lambda^*_\epsilon(\sigma; C)_{ij}\right| \le 3\left(\binom{M}{2} - 1\right)\mu.$$

令 $0 < \mu < \min\{p - 0.5, 0.5 - q\}$, $R(\lambda) := \|\lambda\|^2/2$，并定义事件

$$\mathcal{E}_T(\mu) := \bigcap_{t=h(T)}^{\infty} \left\{\omega \in \Omega : \max_{(i,j)\in I} |\hat{c}_t(\omega)_{ij} - c_{ij}| < \mu\right\}$$

和函数

$$\begin{aligned}
g(\tilde{\lambda}, \tilde{C}) &:= \inf_{C'\in\mathrm{Alt}(C)} \sum_{(i,j)\in I} \tilde{\lambda}_{ij} d(\tilde{c}_{ij}, c'_{ij}) - \sigma R(\tilde{\lambda}) \\
\hat{M}(t) &:= \inf_{C'\in\mathrm{Alt}(C)} \sum_{(i,j)\in I} \frac{N_{ij}(t)}{t} d((\hat{c}_t)_{ij}, c'_{ij}) - \sigma R\left(\frac{N_{ij}(t)}{t}\right) = g\left(\frac{N(t)}{t}, \hat{C}_t\right) \\
C_\mu(C) &:= \inf_{\substack{\tilde{C}:\max_{(i,j)\in I} |\tilde{c}_{ij} - c_{ij}| < \mu \\ \tilde{\lambda}:\max_{(i,j)\in I} |\tilde{\lambda}_{ij} - \lambda^*_\epsilon(\sigma;C)_{ij}| < 3\left(\binom{M}{2}-1\right)\mu}} g(\tilde{\lambda}, \tilde{C}),
\end{aligned}$$

对于某个递增函数 $h(t)$。我们在此注意到，由于所有已定义函数中的$\mathrm{Alt}(C)$**并不**依赖于输入，因此并非必须 $C \in \mathcal{C}$，我们可以直接使用 C_t。此外， g 是连续函数。

在事件 $\mathcal{E}_T(\mu)$上，我们有 $\forall t \ge h(T)$,

(i) $[C_t] = [C]$(ii) $\mathrm{Alt}(C_t) = \mathrm{Alt}(C)$ (iii) $Z(t) = \inf_{C'\in\mathrm{Alt}(C)} \sum_{(i,j)\in I} N_{ij}(t) d((\hat{c}_t)_{ij}, c'_{ij}) \le \beta(t,\delta) \implies tM(t) \le \beta(t,\delta)$ (iv) $\exists t'_\mu \ge h(T)$使得 $\forall t \ge t'_\mu$, $\sup^{t\ge t'_\mu} \max_{(i,j)\in I} |\frac{N_{ij}(t)}{t} - \lambda^*_\epsilon(\sigma; C)_{ij}| < 3(\binom{M}{2} - 1)\mu$ (v) $\forall t \ge t'_\mu$, $\hat{M}(t) \ge C_\mu(C)$。

第（iii）部分成立，因为 $Z(t) = t\hat{M}(t) + t\sigma R(\frac{N(t)}{t})$ 和 $t\sigma R(\frac{N(t)}{t})$ 为正。第（v）部分由第（iv）部分以及我们在 $C_\mu(C)$ 中对 $\tilde{\ }$ 和 λ 及 C 的集合所采用的定义推出。

现在，我们可以选择 T_1和 $h(T) > T$，使得 $h(T_1) \ge \binom{M}{2}^2$，并选择 T使得 $\sqrt{T} \ge t'_\mu$。那么，在 $\mathcal{E}_T(\mu)$ 上，

$$\min(\tau_\delta, T) \le \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{1}\{\tau_\delta > t\} \overset{\text{(iii)}}{\le} \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{1}\{t\hat{M}(t) \le \beta(t,\delta)\}$$

$$\overset{(v)}{\leq} \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{1}\{tC_\mu(C) \leq \beta(T,\delta)\} \leq \sqrt{T} + \frac{\beta(T,\delta)}{C_\mu(C)}.$$

令 $T_2(\delta) = \inf\{T \in \mathbb{N}: \sqrt{T} + \frac{\beta(T,\delta)}{C_\mu(C)} \leq T\}$，并令 $T_3 = \max\{T_2(\delta), (t'_\mu)^2\}$，则 $\tau_\delta \leq T_3$。($T - \sqrt{T}$是一个递增函数，且 $T_2(\delta)$ 是满足在 $\mathcal{E}_T(\mu)$ 上 $T \geq \tau_\delta$ 的最小 T) 因此，我们有 $\forall T \geq T_3, \mathcal{E}_T(\mu) \subset \{\tau_\delta \leq T\}$，且 $\Pr(\mathcal{E}_T^c(\mu)) \geq \Pr(\tau_\delta > T)$。

那么，截至目前，我们有

$$\begin{aligned}
\mathbb{E}_C[\tau_\delta] &= \sum_{t=1}^{T_3-1} \Pr(\tau_\delta > t) + \sum_{t=T_3}^{\infty} \Pr(\tau_\delta > t) \\
&\leq T_3 + \sum_{t=T_3}^{\infty} \Pr(\mathcal{E}_t^c(\mu)) \\
&\leq T_2(\delta) + (t'_\mu)^2 + \sum_{t=T_3}^{\infty} \Pr(\mathcal{E}_t^c(\mu))
\end{aligned}$$

现在，我们将依次处理每一项，首先给出 $T_2(\delta)$ 的上界。令 $\xi > 0$ and $C(\xi) = \inf\{T \in \mathbb{N}: \ T - \sqrt{T} \geq T/(1+\xi)\}$。然后，利用引理 5 中关于 $\beta(t,\delta)$ 的上界，我们有

$$\begin{aligned}
T_2(\delta) &\leq C(\xi) + \inf\left\{T \in \mathbb{N}: \frac{\beta(T,\delta)}{C_\mu(C)} \leq \frac{T}{1+\xi}\right\} \\
&\leq C + C(\xi) + \inf\left\{T \in \mathbb{N}: \frac{\ln(\frac{DT^\alpha}{\delta^\gamma})}{C_\mu(C)} \leq \frac{T}{1+\xi}\right\} \\
&= C + C(\xi) + \inf\left\{T \in \mathbb{N}: \frac{\ln(\frac{D^{1/\alpha}T}{\delta^{\gamma/\alpha}})}{C_\mu(C)} \leq \frac{T}{\alpha(1+\xi)}\right\}
\end{aligned}$$

由于

$$h\left(\frac{TC_\mu(C)}{\alpha(1+\xi)}\right) = \frac{TC_\mu(C)}{\alpha(1+\xi)} - \ln\left(\frac{TC_\mu(C)}{\alpha(1+\xi)}\right),$$

对于满足 $\ln\left(\frac{D^{1/\alpha}T}{\delta^{\gamma/\alpha}}\right) \leq \frac{T}{\alpha(1+\xi)}$ 的 T，我们有

$$\begin{aligned}
&\ln\left(\frac{D^{1/\alpha}T}{\delta^{\gamma/\alpha}}\right) - \ln\left(\frac{C_\mu(C)}{\alpha(1+\xi)}\right) \leq \frac{TC_\mu(C)}{\alpha(1+\xi)} - \ln\left(\frac{C_\mu(C)}{\alpha(1+\xi)}\right) \\
\iff &\ln\left(\frac{D^{1/\alpha}}{\delta^{\gamma/\alpha}}\right) - \ln\left(\frac{C_\mu(C)}{\alpha(1+\xi)}\right) \leq \frac{TC_\mu(C)}{\alpha(1+\xi)} - \ln\left(\frac{TC_\mu(C)}{\alpha(1+\xi)}\right) \\
\iff &\ln\left(\frac{\alpha(1+\xi)D^{1/\alpha}}{C_\mu(C)\delta^{\gamma/\alpha}}\right) \leq h\left(\frac{TC_\mu(C)}{\alpha(1+\xi)}\right).
\end{aligned}$$

那么，由于 h 是递增函数，

$$h^{-1}\left(\ln(\frac{\alpha(1+\xi)D^{1/\alpha}}{C_\mu(C)\delta^{\gamma/\alpha}})\right) \leq \frac{TC_\mu(C)}{\alpha(1+\xi)} \quad \iff \quad \frac{\alpha(1+\xi)}{C_\mu(C)} h^{-1}\left(\ln\left(\frac{\alpha(1+\xi)D^{1/\alpha}}{C_\mu(C)\delta^{\gamma/\alpha}}\right)\right) \leq T.$$

将该不等式代入并在 [45], 中使用命题8，我们得到

$$\begin{aligned}
T_2(\delta) &\leq C + C(\xi) + \inf\left\{T \in \mathbb{N}: \frac{\ln(\frac{D^{1/\alpha}T}{\delta^{\gamma/\alpha}})}{C_\mu(C)} \leq \frac{T}{\alpha(1+\xi)}\right\} \\
&= C + C(\xi) + \inf\left\{T \in \mathbb{N}: \frac{\alpha(1+\xi)}{C_\mu(C)} h^{-1}\left(\ln\left(\frac{\alpha(1+\xi)D^{1/\alpha}}{C_\mu(C)\delta^{\gamma/\alpha}}\right)\right) \leq T\right\} \\
&\leq C + C(\xi) + \frac{\alpha(1+\xi)}{C_\mu(C)} h^{-1}\left(\ln\left(\frac{\alpha(1+\xi)D^{1/\alpha}}{C_\mu(C)\delta^{\gamma/\alpha}}\right)\right) + 1 \\
&= C + C(\xi) + \frac{\alpha(1+\xi)}{C_\mu(C)} h^{-1}\left(\kappa + \ln\left(\frac{1}{\delta^{\gamma/\alpha}}\right)\right) + 1 \\
&\leq C + C(\xi) + 1
\end{aligned}$$

$$+\frac{\alpha(1+\xi)}{C_\mu(C)}\left[\kappa+\ln\left(\frac{1}{\delta^{\gamma/\alpha}}\right)+\ln\left(\kappa+\ln\left(\frac{1}{\delta^{\gamma/\alpha}}\right)+\sqrt{2\left(\kappa+\ln(\frac{1}{\delta^{\gamma/\alpha}})-1\right)}\right)\right],$$

其中 $\kappa := \ln(\frac{\alpha(1+\xi)D^{1/\alpha}}{C_\mu C}())$。

接下来，我们给出 $\sum_{t=T_3}^{\infty}\Pr(\mathcal{E}_t^c(\mu))$ 的上界。对于 $T \geq T_3$, $h(T) \geq \binom{M}{2}^2$，根据 Chernoff–Hoeffding 定理 [49],

$$\begin{aligned}
\Pr(\mathcal{E}_T^c(\mu)) &= \Pr\left(\bigcup_{t=h(T)}^{\infty}\left\{\max_{(i,j)\in I}|(\hat{c}_t)_{ij}-c_{ij}|\geq\mu\right\}\right)\\
&\leq \sum_{t=h(T)}^{\infty}\Pr\left(\max_{(i,j)\in I}|(\hat{c}_t)_{ij}-c_{ij}|\geq\mu\right)\\
&\leq \sum_{t=h(T)}^{\infty}\sum_{(i,j)\in I}\Pr(|(\hat{c}_t)_{ij}-c_{ij}|\geq\mu)\\
&= \sum_{t=h(T)}^{\infty}\sum_{(i,j)\in I}\Pr((\hat{c}_t)_{ij}-c_{ij}\geq\mu)+\Pr(c_{ij}-(\hat{c}_t)_{ij}\geq\mu)\\
&\leq \sum_{t=h(T)}^{\infty}\sum_{(i,j)\in I}e^{-d(c_{ij}+\mu,c_{ij})N_{ij}(t)}+e^{-d(c_{ij}-\mu,c_{ij})N_{ij}(t)}
\end{aligned}$$

根据 D-tracking 规则以及以下事实：$t \geq \binom{M}{2}^2$, $N_{ij}(t) \geq (\sqrt{t}-\binom{M}{2}/2)_+ - 1 \geq \sqrt{t}-\binom{M}{2}$ 对于 $M \geq 3$。令$d_\mu = \min\{d(p+\mu,\ p)\ d(p-\mu,\ p)\ d(q+\mu,\ q)\ d(q-\mu,\ q)\}$。那么，

$$\begin{aligned}
\Pr(\mathcal{E}_T^c(\mu)) &\leq \sum_{t=h(T)}^{\infty}\sum_{(i,j)\in I}e^{-d(c_{ij}+\mu,c_{ij})N_{ij}(t)}+e^{-d(c_{ij}-\mu,c_{ij})N_{ij}(t)}\\
&\leq \sum_{t=h(T)}^{\infty}2\binom{M}{2}e^{-d_\mu(\sqrt{t}-\binom{M}{2})}\\
&= 2\binom{M}{2}e^{d_\mu\binom{M}{2}}\sum_{t=h(T)}^{\infty}e^{-d_\mu\sqrt{t}}\\
&\leq 2\binom{M}{2}e^{d_\mu\binom{M}{2}}\int_{t=h(T)-1}^{\infty}e^{-d_\mu\sqrt{t}}dt\\
&= 4\binom{M}{2}e^{d_\mu\binom{M}{2}}\int_{t=\sqrt{h(T)-1}}^{\infty}te^{-d_\mu t}dt\\
&= 4\binom{M}{2}e^{d_\mu\binom{M}{2}}\left(\frac{\sqrt{h(T)-1}}{d_\mu}+\frac{1}{d_\mu^2}\right)e^{-d_\mu\sqrt{h(T)-1}}
\end{aligned}$$

那么，由于 $h(t)\geq t$ 是一个递增函数，我们可以将 $h(t)$ 视为 t 的一个子序列，并且有

$$\begin{aligned}
\sum_{t=T_3}^{\infty}\Pr(\mathcal{E}_t^c(\mu)) &\leq 4\binom{M}{2}e^{d_\mu\binom{M}{2}}\sum_{t=T_3}^{\infty}\left(\frac{\sqrt{h(t)-1}}{d_\mu}+\frac{1}{d_\mu^2}\right)e^{-d_\mu\sqrt{h(t)-1}}\\
&\leq \frac{4\binom{M}{2}e^{d_\mu\binom{M}{2}}}{d_\mu}\sum_{t=T_3}^{\infty}\left(\sqrt{t-1}+\frac{1}{d_\mu}\right)e^{-d_\mu\sqrt{t-1}}.
\end{aligned}$$

由于右侧的无穷级数收敛，因此 $\sum_{t=T_3}^{\infty}\Pr(\mathcal{E}_t^c(\mu))$ 也收敛。最后，我们得到

$$\begin{aligned}
\limsup_{\delta\to0}\frac{\mathbb{E}_C[\tau_\delta]}{\ln(1/\delta)} &\leq \limsup_{\delta\to0}\frac{T_2(\delta)}{\ln(1/\delta)}\\
&\leq \frac{\alpha(1+\xi)}{C_\mu(C)}\limsup_{\delta\to0}\frac{\kappa+\frac{\gamma}{\alpha}\ln(1/\delta)+\ln\left(\kappa+\frac{\gamma}{\alpha}\ln(1/\delta)+\sqrt{2\left(\kappa+\frac{\gamma}{\alpha}\ln(1/\delta)-1\right)}\right)}{\ln(1/\delta)}\\
&= \frac{\gamma(1+\xi)}{C_\mu(C)}
\end{aligned}$$

令 μ 和 ξ 趋于零，并利用 g 的连续性，我们得到上界

$$\limsup_{\delta\to 0}\frac{\mathbb{E}_C[\tau_\delta]}{\ln(1/\delta)} \le \gamma D_\epsilon^*(\sigma;C)^{-1},$$

其中，$\gamma>1$ 是一个可以任意接近 1 的常数。

由于对所有 $\sigma>0$，$\inf_{C'\in \mathrm{Alt}(C)}\sum_{(i,j)\in I}\lambda_\epsilon^*(\sigma;C)_{ij}d(c_{ij},c'_{ij})>0$，因此必定存在足够小的 σ，使得 $D_\epsilon^*(\sigma;C)>0$。 □

下面定义完整描述 $\beta(t,\delta)$ 所需的函数，

$$h(u)=u-\ln(u),\quad u\ge 1$$

$$\text{for some } z\in[1,e],\ \tilde{h}_z(x)=\begin{cases}e^{1/h^{-1}(x)}h^{-1}(x), \text{ if } x\ge h(1/\ln(z))\\ z(x-\ln\ln(z)), \text{ otherwise}\end{cases}$$

$$\zeta(s)=\sum_{n=1}^{\infty}n^{-s}$$

$$\mathcal{C}_{\exp}(x)=2\tilde{h}_{3/2}\left(\frac{h^{-1}(1+x)+\ln(2\zeta(2))}{2}\right)$$

$$\beta(t,\delta)=3\sum_{(i,j)\in I}\ln(1+\ln N_{ij}(t))+\binom{M}{2}\mathcal{C}_{\exp}\left(\frac{\ln(1/\delta)}{\binom{M}{2}}\right).$$

显然，$\beta(t,\delta)$ 关于 t 递增。

引理 5. 存在常数 α、C、$D>0$ 和 $\gamma>1$，使得 $\forall t\ge C$、$\delta\in(0,1)$，

$$\beta(t,\delta)\le \ln\left(\frac{Dt^\alpha}{\delta^\gamma}\right).$$

证明。我们先注意到 h 和 h^{-1} 的若干性质。首先，由于 h 定义在 $u\ge 1$ 上，因此 $h^{-1}(u)\ge 1$。然后，根据 [45], 中的引理 29，对于 $x\ge 0$、$z\in[1,e]$,

$$\tilde{h}_z(x)=\min_{y\in[1,z]}y(x-\ln\ln(y))$$

并且根据 $h^{-1}(x)$ 的变分表示，对于 $\gamma_2>1$

$$h^{-1}(x)=\inf_{z\ge 1}z\left(x-1+\ln(\frac{z}{z-1})\right)\le\gamma_2\left(x-1+\ln(\frac{\gamma_2}{\gamma_2-1})\right).$$

结合这些结果，对于 $\gamma_1\in(1,3/2]$，我们有

$$\begin{aligned}
\mathcal{C}_{\exp}\left(\frac{\ln(1/\delta)}{\binom{M}{2}}\right)&=2\tilde{h}_{3/2}\left(\frac{h^{-1}(1+\frac{\ln(1/\delta)}{\binom{M}{2}})+\ln(2\zeta(2))}{2}\right)\\
&=2\min_{y\in[1,3/2]}y\left(\frac{h^{-1}(1+\frac{\ln(1/\delta)}{\binom{M}{2}})+\ln(2\zeta(2))}{2}-\ln\ln(y)\right)\\
&\le 2\cdot\gamma_1\left(\frac{1}{2}h^{-1}(1+\frac{\ln(1/\delta)}{\binom{M}{2}})+\frac{1}{2}\ln(2\zeta(2))-\ln\ln(\gamma_1)\right)\\
&=\gamma_1\left(h^{-1}(1+\frac{\ln(1/\delta)}{\binom{M}{2}})+\ln(2\zeta(2))-2\ln\ln(\gamma_1)\right)\\
&\le\gamma_1\left(\gamma_2\left(\frac{\ln(1/\delta)}{\binom{M}{2}}+\ln(\frac{\gamma_2}{\gamma_2-1})\right)+\ln(2\zeta(2))-2\ln\ln(\gamma_1)\right)\\
&=\frac{\gamma_1\gamma_2\ln(1/\delta)}{\binom{M}{2}}+\gamma_1\gamma_2\ln(\frac{\gamma_2}{\gamma_2-1})+\gamma_1\ln(2\zeta(2))-2\gamma_1\ln\ln(\gamma_1)
\end{aligned}$$

and

$$\binom{M}{2}\mathcal{C}_{\exp}\left(\frac{\ln(1/\delta)}{\binom{M}{2}}\right) \le \gamma_1\gamma_2\ln(1/\delta) + \underbrace{\binom{M}{2}\left(\gamma_1\gamma_2\ln(\frac{\gamma_2}{\gamma_2-1}) + \gamma_1\ln(2\zeta(2)) - 2\gamma_1\ln\ln(\gamma_1)\right)}_{:=a>0}.$$

接着，对于 $\forall\delta\in(0,1)$，令 $D_1 > e^a$ 满足

$$\binom{M}{2}\mathcal{C}_{\exp}(\frac{\ln(1/\delta)}{\binom{M}{2}}) \le \gamma_1\gamma_2\ln(1/\delta) + \ln(D_1)$$

接下来，我们要为 $3\sum_{(i,j)\in I}\ln(1+\ln\ N_{ij}(t))$ 求一个上界，并注意到 $\ln(1+\ln\ x)$ 在 $x \ge e^{-2}$ 上是凹函数。于是，当 $\min_{(i,j)\in I} N_{ij}(t) \ge 1$ 时，根据詹森不等式 [50]，我们有：

$$\begin{aligned}\sum_{(i,j)\in I}\frac{1}{|I|}\ln(1+\ln N_{ij}(t)) &\le \ln(1+\ln\sum_{(i,j)\in I}\frac{1}{|I|}N_{ij}(t))\\ &= \ln(1+\ln t/|I|)\\ &\le 1+\ln t/|I|.\end{aligned}$$

接着，由于

$$3\sum_{(i,j)\in I}\ln(1+\ln N_{ij}(t)) \le 3|I|(1+\ln t/|I|) \le 3|I| - 3|I|\ln(I) + 3|I|\ln(t),$$

我们可以选择 $D_2 > e^{3|I|-3|I|\ln(I)}$ 和 C，使得 $\forall t\ge C$，$\min_{(i,j)\in I} N_{ij}(t)\ge 1$，从而得到

$$3\sum_{(i,j)\in I}\ln(1+\ln N_{ij}(t)) \le \ln(D_2) + 3|I|\ln(t).$$

综合这些不等式，我们有 $\forall\delta\in(0,\ 1)$ $\forall t\ge C$ 存在一个 $D = D_1D_2 > 0$，使得对于 $\gamma_1\in(1,\ 3/2],\ \gamma_2 > 1$,

$$\begin{aligned}\beta(t,\delta) &= 3\sum_{(i,j)\in I}\ln(1+\ln N_{ij}(t)) + \binom{M}{2}\mathcal{C}_{\exp}\left(\frac{\ln(1/\delta)}{\binom{M}{2}}\right)\\ &\le \ln(D_2) + 3|I|\ln(t) + \gamma_1\gamma_2\ln(1/\delta) + \ln(D_1)\\ &= \ln\left(\frac{Dt^{3\binom{M}{2}}}{\delta^{\gamma_1\gamma_2}}\right) := \ln\left(\frac{Dt^{\alpha}}{\delta^{\gamma}}\right)\end{aligned}$$

□

D. 定理 *4* 的证明

证明。 首先，根据 D-tracking 规则，对于所有 $(i,\ j)\in I$, $N_{ij}(t) \ge (\sqrt{t-\binom{M}{2}/2})_+ - 1$，并根据 [23, 引理 17], $\forall\mu>0$,$\exists t'_\mu \ge t_\mu$，使得

$$\sup_{t\ge t_\mu}\max_{(i,j)\in I}|\lambda^*_\epsilon(\sigma;C_t)_{ij} - \lambda^*_\epsilon(\sigma;C)_{ij}| < \mu \implies \sup_{t\ge t'_\mu}\max_{(i,j)\in I}\left|\frac{N_{ij}(t)}{t} - \lambda^*_\epsilon(\sigma;C)_{ij}\right| \le 3\left(\binom{M}{2}-1\right)\mu.$$

令 $0<\mu<\min\{p-0.5, 0.5-q\}$ and $R(\lambda) = \frac{1}{2}\|\lambda\|^2$。定义事件

$$\mathcal{E}_T(\mu) := \bigcap_{t=h(T)}^{\infty}\left\{\omega\in\Omega : \max_{(i,j)\in I}|\hat{c}_t(\omega)_{ij} - c_{ij}| < \mu\right\}$$

和函数

$$\begin{aligned}g(\tilde{\lambda},\tilde{C}) &:= \left[\min_{(i',j')\in I}\tilde{\lambda}_{i'j'}\right]\cdot\inf_{C'\in\mathrm{Alt}(C)}\sum_{(i,j)\in I}d(\tilde{c}_{ij},c'_{ij}) - \sigma R(\tilde{\lambda})\\ \hat{M}(t) &:= \left[\min_{(i',j')\in I}\frac{N_{i'j'}(t)}{t}\right]\cdot\inf_{C'\in\mathrm{Alt}(C)}\sum_{(i,j)\in I}d((\hat{c}_t)_{ij},c'_{ij}) - \sigma R\left(\frac{N(t)}{t}\right) = g\left(\frac{N(t)}{t},\hat{C}_t\right)\\ C_\mu(C) &:= \inf_{\substack{\tilde{C}:\max_{(i,j)\in I}|\tilde{c}_{ij}-c_{ij}|<\mu\\ \tilde{\lambda}:\max_{(i,j)\in I}|\tilde{\lambda}_{ij}-\lambda^*_\epsilon(\sigma;C)_{ij}|<3(\binom{M}{2}-1)\mu}} g(\tilde{\lambda},\tilde{C}),\end{aligned}$$

对于某个递增函数 $h(t)$。我们观察到， g 仍然是一个连续函数。

在事件 $\mathcal{E}_T(\mu)$ 上，我们有 $\forall t \geq h(T)$，

1) $[C_t] = [C]$2) $\mathrm{Alt}(C_t) = \mathrm{Alt}(C)$ 3) $Z(t) = \left[\min_{(i',j')\in I} N_{i'j'}(t)\right] \cdot \inf_{C'\in \mathrm{Alt}(C)} \sum_{(i,j)\in I} d((\hat{c}_t)_{ij},$ $c'_{ij}) \leq \beta(t,\delta) \implies t\hat{} M(t) \leq \beta(t,\delta)$ 4) $\exists t'_\mu \geq h(T)$ 使得 $\forall t \geq t'_\mu, \max_{(i,j)\in I} |^{N_{ij}}(t)_t - \lambda^*_\epsilon(\sigma;\ C)$ $_{ij}| \leq 3(\binom{M}{2} - 1)\mu\hat{}$5) $\forall t \geq t'_\mu,\ M(t) \geq C_\mu(C)$.

3. 成立是因为 $Z\ t\ = tM\ t\ + t\sigma R\ ^{N\ t}_{\ t}\quad t\sigma R\ ^{N\ t}_{\ t}\quad 5\quad 4\quad \tilde{\lambda}(\,)$ $\hat{}(\,)\,(\ _{()})$ 且 $(\,)\tilde{}$ 为正。. 由 . 以及在 $C_\mu(C)$ 中进行最小化时所涉及的和 C 的集合定义可得。

现在，我们可以选择 T_1 和 $h(T) > T$，使得 $h(T_1) \geq \binom{M}{2}^2$，并选择 T 使得 $\sqrt{T} \geq t'_{\mu^\circ}$ 那么，在 $\mathcal{E}_T(\mu)$ 上，

$$\begin{aligned}\min(\tilde{\tau}_\delta, T) &\leq \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{1}\{\tilde{\tau}_\delta > t\} \overset{3.}{\leq} \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{1}\{t\hat{M}(t) \leq \beta(t,\delta)\} \\ &\overset{5.}{\leq} \sqrt{T} + \sum_{t=\sqrt{T}}^{T} \mathbb{1}\{tC_\mu(C) \leq \beta(T,\delta)\} \leq \sqrt{T} + \frac{\beta(T,\delta)}{C_\mu(C)}.\end{aligned}$$

令 $T_2(\delta) = \inf\{T \in \mathbb{N}\colon\ \sqrt{T} + \frac{\beta}{(T,\delta)C_\mu(C)} \leq T\}$，且 $T_3 = \max\{T_2(\delta), (t_\mu)^2\}$，则 $\hat{}\tau_\delta \leq T_3$。（$T - \sqrt{T}$ 是一个递增函数）因此，我们有 $\forall T \geq T_3, \mathcal{E}_T(\mu) \subset \{\hat{}\tau_\delta \leq T\}$，且 $\Pr(\mathcal{E}^c_T(\mu)) \geq \Pr(\hat{}\tau_\delta > T)$。

那么，到目前为止我们有

$$\begin{aligned}\mathbb{E}_C[\tilde{\tau}_\delta] &= \sum_{t=1}^{T_3-1} \Pr(\tilde{\tau}_\delta > t) + \sum_{t=T_3}^{\infty} \Pr(\tilde{\tau}_\delta > t) \\ &\leq T_3 + \sum_{t=T_3}^{\infty} \Pr(\mathcal{E}^c_t(\mu)) \\ &\leq T_2(\delta) + (t'_\mu)^2 + \sum_{t=T_3}^{\infty} \Pr(\mathcal{E}^c_t(\mu)).\end{aligned}$$

我们可以按照附录 C-C 中的相同步骤，对 $T_2(\delta)$ 和 $\sum_{t=T_3}^{\infty} \Pr(\mathcal{E}^c_t(\mu))$ 给出上界。于是，我们有

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_C[\tilde{\tau}_\delta]}{\ln(1/\delta)} \leq \frac{\gamma(1+\xi)}{C_\mu(C)}.$$

由于

$$C_\mu(C) = \inf_{\substack{\tilde{C}:\max_{(i,j)\in I} |\tilde{c}_{ij} - c_{ij}| < \mu \\ \tilde{\lambda}:\max_{(i,j)\in I} |\tilde{\lambda}_{ij} - \lambda^*(\sigma;C)_{ij}| < 3(\binom{M}{2}-1)\mu}} g(\tilde{\lambda}, \tilde{C})$$

且 g 是连续函数，当 $\mu \to 0$ 时，

$$C_\mu(C) \to g(\lambda^*_\epsilon(\sigma;C), C) = \left[\min_{(i',j')\in I} \lambda^*_\epsilon(\sigma;C)_{i'j'}\right] \cdot \inf_{C'\in \mathrm{Alt}(C)} \sum_{(i,j)\in I} d(c_{ij}, c'_{ij}) - \sigma R(\lambda^*_\epsilon(\sigma;C))$$

$$:= \tilde{D}^*_\epsilon(\sigma;C).$$

由于 $\left[\min_{(i',j')\in I} \lambda^*_\epsilon(\sigma;\ C)_{i'j'}\right] \cdot \left[\inf_{C'\in \mathrm{Alt}(C)} \sum_{(i,j)\in I} d(c_{ij}, c'_{ij})\right] \geq \frac{\epsilon}{|I|}\left[\inf_{C'\in \mathrm{Alt}(C)} \sum_{(i,j)\in I} d(c_{ij}, c'_{ij})\right] > 0$ 对所有 $\sigma > 0$ 都成立，因此必定存在一个足够小的 σ，使得 $D^*_\epsilon(\sigma;\ C) > 0$。

□

APPENDIX D
ON THE COMPUTATIONALLY FEASIBLE STOPPING RULE

下面我们推导出 $Z(t)$ 的闭式表达式，从而使其成为基于广义似然比的统计量 $Z(t)$ 的一种计算上可行且实用的替代方案（由于缺少闭式表达式，后者无法计算）。我们证明，式 (13) 中对 $\mathrm{Alt}(C_t)$ 的搜索可简化为对有限多个等价类的搜索。此外，我们还证明，无需完整计算式 (13)，只需计算 $\hat{}p_n$ 的最大值和最小值即可，从而显著提高效率。

命题 4. 令 $(i_t^*, j_t^*) \in \arg\min_{(i,j)\in I} N_{ij}(t)$，对于固定的备选实例 $C' \in \mathrm{Alt}(C_t)$，定义 $N_p' = N_p'(C') = \{(i,j) \in I : c_i = c_j'\}$ $N_q' = N_q'(C') = \{(i,j) \in I : c_i \neq c_j'\}$ $n = |N_p'|$ $n' = \binom{M}{2} - n$ $\hat{C}_{\mathrm{tot}} = \sum_{(i,j)\in I} (\hat{c}_t)_{ij}$。此外，令

$$\hat{p}_n = \frac{1}{n} \sum_{(i,j)\in N_p'} (\hat{c}_t)_{ij}, \qquad \hat{p}_n^{\max} = \max_{\substack{C'\in \mathrm{Alt}(C_t)/\sim: \\ |N_p'|=n}} \hat{p}_n, \qquad \hat{p}_n^{\min} = \min_{\substack{C'\in \mathrm{Alt}(C_t)/\sim: \\ |N_p'|=n, \hat{C}_{\mathrm{tot}} - n\hat{p}_n < 0.5n'}} \hat{p}_n,$$

其中 $\hat{C}_{\mathrm{tot}} = \sum_{(i,j)\in I} (\hat{c}_t)_{ij}$。令 $s_n(p) = nH(p) + n'H\left(\frac{\hat{C}_{\mathrm{tot}} - np}{n'}\right)$，其中 $H(p) = -p \log p - (1-p)\log(1-p)$ 是 $\mathrm{Ber}(p)$ 的二元熵。于是，可计算的停止统计量 $Z(t)$ 可简化为

$$\widehat{Z}(t) = N_{i_t^* j_t^*}(t) \cdot \begin{cases} \min\limits_{n\in[\binom{M}{2}]} s_n(\hat{p}_n^{\max}) - \sum\limits_{(i,j)\in I} H((\hat{c}_t)_{ij}), & \text{for } \hat{C}_{\mathrm{tot}} > \frac{\binom{M}{2}}{2} \\ \min\limits_{n\in[\binom{M}{2}]} \min\{s_n(\hat{p}_n^{\max}), s_n(\hat{p}_n^{\min})\} - \sum\limits_{(i,j)\in I} H((\hat{c}_t)_{ij}), & \text{for } \hat{C}_{\mathrm{tot}} \leq \frac{\binom{M}{2}}{2} \end{cases}.$$

证明。和往常一样，我们首先定义集合，以便在备选停止规则中对目标函数中的各项进行分组

$$N_p' = N_p'(C') = \{(i,j) \in I : c_i' = c_j'\}; \qquad N_q' = N_q'(C') = \{(i,j) \in I : c_i' \neq c_j'\}.$$

接下来，我们通过展开 KL 散度来简化目标函数

$$\begin{aligned}
\widehat{Z}(t) &= \inf_{C'\in \mathrm{Alt}(C_t)} \Big[\min_{(i',j')\in I} N_{i'j'}(t) \Big] \sum_{(i,j)\in I} d((\hat{c}_t)_{ij}, c_{ij}') \\
&= N_{i_t^*, j_t^*}(t) \inf_{C'\in \mathrm{Alt}(C_t)} \sum_{(i,j)\in I} d((\hat{c}_t)_{ij}, c_{ij}') \\
&= N_{i_t^*, j_t^*}(t) \inf_{C'\in \mathrm{Alt}(C_t)} \sum_{(i,j)\in N_p'} d((\hat{c}_t)_{ij}, p') + \sum_{(i,j)\in N_q'} d((\hat{c}_t)_{ij}, q') \\
&= N_{i_t^*, j_t^*}(t) \inf_{C'\in \mathrm{Alt}(C_t)} \sum_{(i,j)\in N_p'} (\hat{c}_t)_{ij} \log \frac{(\hat{c}_t)_{ij}}{p'} + (1 - (\hat{c}_t)_{ij}) \log \frac{1 - (\hat{c}_t)_{ij}}{1 - p'} \\
&\qquad\qquad + \sum_{(i,j)\in N_q'} (\hat{c}_t)_{ij} \log \frac{(\hat{c}_t)_{ij}}{q'} + (1 - (\hat{c}_t)_{ij}) \log \frac{1 - (\hat{c}_t)_{ij}}{1 - q'} \\
&= N_{i_t^*, j_t^*}(t) \inf_{C'\in \mathrm{Alt}(C_t)} - \sum_{(i,j)\in N_p'} (\hat{c}_t)_{ij} \log p' - (1 - (\hat{c}_t)_{ij}) \log(1 - p') \\
&\qquad\qquad - \sum_{(i,j)\in N_q'} (\hat{c}_t)_{ij} \log q' - (1 - (\hat{c}_t)_{ij}) \log(1 - q') - \sum_{(i,j)\in I} H((\hat{c}_t)_{ij}) \\
&= N_{i_t^*, j_t^*}(t) \inf_{C'\in \mathrm{Alt}(C_t)} s(C') - \sum_{(i,j)\in I} H((\hat{c}_t)_{ij}),
\end{aligned} \tag{29}$$

其中 p' q' C' $s\ C' = -\sum_{i\ j\ \in N_p'} c_{t\ ij}$ $p' - 1 - c_{t\ ij}$ $1 - p'$ $-\sum_{i\ j\ \in N_q'} c_{t\ ij}$ $q' - 1 - c_{t\ ij}$ $1 - q'$ N_p' N_q' p' q' 和 分别是与实例 及 ( ) 相关的概率 $_{(,)}(\hat{\ }) \log(\hat{\ }) \log(\ )$ $_{(,)}(\hat{\ }) \log(\hat{\ }) \log(\ )$。为方便记号，我们省略 、 和 上的撇号。在等价类 $\mathrm{Alt}(C_t)/\sim$ 中，最优的 p和 q应分别是 N_p 和 N_q 中 $(\hat{c}_t)_{ij}$ 的经验平均，这一点应当不难理解。为完整起见，我们在下文求解它们。对 s 关于 p 求导，得到

$$\frac{ds}{dp} = |N_p| \left(-\frac{\sum_{(i,j)\in N_p} (\hat{c}_t)_{ij}}{|N_p|} \frac{1}{p} + \left(1 - \frac{\sum_{(i,j)\in N_p} (\hat{c}_t)_{ij}}{|N_p|}\right) \frac{1}{1-p} \right).$$

情况1：$\frac{\sum_{(i,j)\in N_p} c_{t\ ij}}{|N_p|} < 0.5 < p^{(\ )}$

于是，我们有

$$\frac{\sum_{(i,j)\in N_p} (\hat{c}_t)_{ij}}{|N_p|} \frac{1}{p} < 1, \qquad \left(1 - \frac{\sum_{(i,j)\in N_p} (\hat{c}_t)_{ij}}{|N_p|}\right) \frac{1}{1-p} > 1$$

并且 $\frac{ds}{dp} > 0$ 对于 $0.5 < p < 1$。因此， $p = 0.5$。

情况 2：$\sum_{(i,j)\in N_p} \frac{c_{t\ ij}}{|N_p|} > 0.5$（^）如果 $\sum_{(i,j)\in N_p} \frac{c_{t\ ij}}{|N_p|} > p > 0.5$（^），则，

$$\frac{\sum_{(i,j)\in N_p}(\hat{c}_t)_{ij}}{|N_p|}\frac{1}{p} > 1, \qquad \left(1 - \frac{\sum_{(i,j)\in N_p}(\hat{c}_t)_{ij}}{|N_p|}\right)\frac{1}{1-p} < 1$$

以及 $\frac{ds}{dp} < 0$。

如果 $p > \frac{\sum_{(i,j)\in N_p} c_{t\ ij}}{|N_p|} > 0.5$（^），那么情况1成立，且 $\frac{ds}{dp} \geq 0$。因此，$p = \frac{\sum_{(i,j)\in N_p} c_{t\ ij}\ (\hat{})_\circ}{|N_p|}$

综合这两种情形，我们有 $p = \max\{0.5, \frac{\sum_{(i,j)\in N_p} c_{t\ ij}}{|N_p|}\}(\hat{})_\circ$

对于 q，同理，

$$\frac{ds}{dq} = |N_q|\left(-\frac{\sum_{(i,j)\in N_q}(\hat{c}_t)_{ij}}{|N_q|}\frac{1}{q} + \left(1 - \frac{\sum_{(i,j)\in N_q}(\hat{c}_t)_{ij}}{|N_q|}\right)\frac{1}{1-q}\right).$$

情况1：$\sum_{(i,j)\in N_q} \frac{c_{t\ ij}}{|N_q|} > 0.5 > q$（^）那么，我们有

$$\frac{\sum_{(i,j)\in N_q}(\hat{c}_t)_{ij}}{|N_q|}\frac{1}{q} > 1, \qquad \left(1 - \frac{\sum_{(i,j)\in N_q}(\hat{c}_t)_{ij}}{|N_q|}\right)\frac{1}{1-q} < 1$$

且对于 $0.5 > q > 0$，$\frac{ds}{dq} < 0$。因此，$q = 0.5$。

情况 2：$\sum_{(i,j)\in N_q} \frac{c_{t\ ij}}{|N_q|} < 0.5$（^）如果 $\sum_{(i,j)\in N_q} \frac{c_{t\ ij}}{|N_q|} < q < 0.5$（^），那么，

$$\frac{\sum_{(i,j)\in N_q}(\hat{c}_t)_{ij}}{|N_q|}\frac{1}{q} < 1, \qquad \left(1 - \frac{\sum_{(i,j)\in N_q}(\hat{c}_t)_{ij}}{|N_q|}\right)\frac{1}{1-q} > 1$$

且 $\frac{ds}{dq} > 0$。

如果 $0.5 > \frac{\sum_{(i,j)\in N_q} c_{t\ ij}}{|N_q|} > q$（^），那么，我们有情况1，且 $\frac{ds}{dq} < 0$。因此，$q = \frac{\sum_{(i,j)\in N_q} c_{t\ ij}\ (\hat{})_\circ}{|N_q|}$

再次结合这两种情形，我们得到 $q = \min\{0.5, \frac{\sum_{(i,j)\in N_q}(\hat{c}_t)_{ij}}{|N_p|}\}$。

接下来，将我们的最优 p 和 q 代入 s 后，进行简化。固定 $|N_p|$，并令 $n := |N_p|$, $n' := |N_q| = \binom{M}{2} - n$, $\hat{}p := \sum_{(i,j)\in N_p}(\hat{c}_t)_{ij}$, $\hat{}C_{\text{tot}} := \sum_{(i,j)\in I}(\hat{c}_t)_{ij}$, $\hat{}q := \sum_{(i,j)\in N_q}(\hat{c}_t)_{ij} = \hat{}C_{\text{tot}} - \hat{}p$。于是，若 $\sum_{(i,j)\in N_p} \frac{c_{t\ ij}}{|N_p|} \geq 0.5$（^）且 $\sum_{(i,j)\in N_q} \frac{c_{t\ ij}}{|N_q|} \leq 0.5$（^），则 $p = \frac{\hat{}p}{n}$ 且 $q = \frac{\hat{}C_{\text{tot}} - \hat{}p}{n'}$。由于我们固定了 n，唯一的相关变量是 $\hat{}p$，因此可以将 s 写为

$$\begin{aligned}
-s_n(\hat{p}) &= \sum_{(i,j)\in N_p}(\hat{c}_t)_{ij}\log p + (1-(\hat{c}_t)_{ij})\log(1-p) + \sum_{(i,j)\in N_q}(\hat{c}_t)_{ij}\log q + (1-(\hat{c}_t)_{ij})\log(1-q) \\
&= \hat{p}\log\frac{\hat{p}}{n} + (n-\hat{p})\log\left(1-\frac{\hat{p}}{n}\right) + (\hat{C}_{\text{tot}}-\hat{p})\log\frac{\hat{C}_{\text{tot}}-\hat{p}}{n'} \\
&\quad + (n'-\hat{C}_{\text{tot}}+\hat{p})\log\left(1-\frac{\hat{C}_{\text{tot}}-\hat{p}}{n'}\right) \\
&= \hat{p}\log\frac{\hat{p}}{n} + (n-\hat{p})\log\left(\frac{n-\hat{p}}{n}\right) + (\hat{C}_{\text{tot}}-\hat{p})\log\frac{\hat{C}_{\text{tot}}-\hat{p}}{n'} \\
&\quad + (n'-\hat{C}_{\text{tot}}+\hat{p})\log\left(\frac{n'-\hat{C}_{\text{tot}}+\hat{p}}{n'}\right).
\end{aligned}$$

然后计算 $\frac{-ds_n}{d\hat{p}}$，——

$$\begin{aligned}
-\frac{ds_n}{d\hat{p}} &= \log\frac{\hat{p}}{n} + \hat{p}\cdot\frac{n}{\hat{p}}\cdot\frac{1}{n} - \log\left(\frac{n-\hat{p}}{n}\right) - (n-\hat{p})\cdot\frac{n}{n-\hat{p}}\cdot\frac{1}{n} \\
&\quad - \log\frac{\hat{C}_{\text{tot}}-\hat{p}}{n'} - (\hat{C}_{\text{tot}}-\hat{p})\cdot\frac{n'}{\hat{C}_{\text{tot}}-\hat{p}}\cdot\frac{1}{n'} + \log\left(\frac{n'-\hat{C}_{\text{tot}}+\hat{p}}{n'}\right) + (n'-\hat{C}_{\text{tot}}+\hat{p})\cdot\frac{n'}{n'-\hat{C}_{\text{tot}}+\hat{p}}\cdot\frac{1}{n'}
\end{aligned}$$

$$
\begin{aligned}
&= \log\frac{\hat{p}}{n} + 1 - \log\left(\frac{n-\hat{p}}{n}\right) - 1 - \log\frac{\hat{C}_{\text{tot}}-\hat{p}}{n'} - 1 + \log\left(\frac{n'-\hat{C}_{\text{tot}}+\hat{p}}{n'}\right) + 1 \\
&= \log\frac{\hat{p}}{n-\hat{p}} + \log\left(\frac{n'-\hat{C}_{\text{tot}}+\hat{p}}{\hat{C}_{\text{tot}}-\hat{p}}\right) \\
&= \log\left(\frac{\hat{p}\{n'-\hat{C}_{\text{tot}}+\hat{p}\}}{(n-\hat{p})\{\hat{C}_{\text{tot}}-\hat{p}\}}\right).
\end{aligned}
$$

然后，对于 $\hat{p} \le nC_{\text{tot}}\binom{M}{2}$，$\frac{ds_n}{d\hat{p}} > 0$；对于 $\hat{p} \ge nC_{\text{tot}}\binom{M}{2}$，$\frac{ds_n}{d\hat{p}} \le 0$。由于 $s_n(\hat{p})$ 是一个先递增后递减、并在 $\hat{p} = n\hat{C}_{\text{tot}}\binom{M}{2}$ 处具有驻点的函数，因此下确界将在 $1 > \frac{\hat{p}}{n} > 0.5$ 的最大值或最小值处取得。现在，我们可以将式（29）中的下确界简化如下，

$$
\begin{aligned}
\inf_{C'\in \text{Alt}(C_t)} s(C') &= \inf_{n\in[\binom{M}{2}]} \inf_{\hat{p}:|N_p|=n} s_n(\hat{p}) \\
&= \begin{cases} \inf_{n\in[\binom{M}{2}]} s_n(\max_{|N_p|=n}\hat{p}), & \text{for } \hat{C}_{\text{tot}} > \frac{\binom{M}{2}}{2} \\ \inf_{n\in[\binom{M}{2}]} \min\{s_n(\max_{|N_p|=n}\hat{p}), s_n(\min_{|N_p|=n,\hat{C}_{\text{tot}}-\hat{p}<0.5n'}\hat{p})\}, & \text{for } \hat{C}_{\text{tot}} \le \frac{\binom{M}{2}}{2} \end{cases}.
\end{aligned} \tag{30}
$$

在详细讨论每种情况之前，我们先确立目标函数的一个性质。由于 $q = \min\{0.5, \frac{\sum_{(i,j)\in N_q} c_{t\ ij}}{|N_q|}\}$，如果 $\frac{\sum_{(i,j)\in N_q} c_{t\ ij}}{|N_q|} = \frac{\hat{C}_{\text{tot}} - \hat{p}}{n'} > 0.5$，那么

$$
\begin{aligned}
-s_n(\hat{p}) &= \hat{p}\log\frac{\hat{p}}{n} + (n-\hat{p})\log\left(\frac{n-\hat{p}}{n}\right) + (\hat{C}_{\text{tot}}-\hat{p})\log\frac{\hat{C}_{\text{tot}}-\hat{p}}{n'} \\
&\quad + (n'-\hat{C}_{\text{tot}}+\hat{p})\log\left(\frac{n'-\hat{C}_{\text{tot}}+\hat{p}}{n'}\right) \\
&= \hat{p}\log\frac{\hat{p}}{n} + (n-\hat{p})\log\left(\frac{n-\hat{p}}{n}\right) + 0.5n'\log 0.5 + (n'-0.5n')\log(1-0.5) \\
&= \hat{p}\log\frac{\hat{p}}{n} + (n-\hat{p})\log\left(\frac{n-\hat{p}}{n}\right) + n'\log 0.5.
\end{aligned} \tag{31}
$$

对于 $\hat{q} = \hat{C}_{\text{tot}} - \hat{p} \le 0.5n'$，由于 $x\log x + (1-x)\log(1-x) > \log 0.5$ 对于 $x \in [0,1]$，

$$
\begin{aligned}
&(\hat{C}_{\text{tot}}-\hat{p})\log\frac{\hat{C}_{\text{tot}}-\hat{p}}{n'} + (n'-\hat{C}_{\text{tot}}+\hat{p})\log\left(\frac{n'-\hat{C}_{\text{tot}}+\hat{p}}{n'}\right) \\
&= n'\frac{\hat{q}}{n'}\log\frac{\hat{q}}{n'} + (n'-n'\frac{\hat{q}}{n'})\log\left(\frac{n'-\hat{q}}{n'}\right) \\
&= n'\frac{\hat{q}}{n'}\log\frac{\hat{q}}{n'} + n'(1-\frac{\hat{q}}{n'})\log\left(1-\frac{\hat{q}}{n'}\right) \\
&= n'\left(\frac{\hat{q}}{n'}\log\frac{\hat{q}}{n'} + (1-\frac{\hat{q}}{n'})\log\left(1-\frac{\hat{q}}{n'}\right)\right) \\
&\ge n'\log 0.5.
\end{aligned} \tag{32}
$$

结合不等式 (31) 和 (32)，对于满足 $\hat{C}_{\text{tot}} - \hat{p}' > 0.5n'$ 的 $\hat{p}'$，以及满足 $\hat{C}_{\text{tot}} - \hat{p} \le 0.5n'$ 的 $\hat{p}$，以下陈述成立

$$
-s_n(\hat{p}') \le -s_n(\hat{p}) \implies s_n(\hat{p}') \ge s_n(\hat{p}).
$$

因此，回到(30)，当我们处于 $\hat{C}_{\text{tot}} > \binom{M}{2}/2$ 的情形时，如果 $\frac{\hat{p}}{n} \le 0.5$，

$$
\hat{C}_{\text{tot}} > \frac{\binom{M}{2}}{2} = 0.5n + 0.5n' \implies \hat{q} = \hat{C}_{\text{tot}} - \hat{p} > \hat{C}_{\text{tot}} - 0.5n > 0.5n'.
$$

那么，$s_n\left(\max_{|N_p|=n}\hat{p}\right) \le s_n(\hat{p}')$，对所有 $\hat{p}' \le 0.5n$，即使 $\hat{p}' \le nC_{\text{tot}}\binom{M}{2}$。

另一方面，在 $\hat{C}_{\text{tot}} \le \binom{M}{2}/2$ 的情况下，可能同时存在 $\frac{\hat{p}}{n} \le 0.5$ 和 $\frac{\hat{q}}{n'} \le 0.5$。因此，为了

最小化 $s_n(\hat{p})$，我们应找到满足 $C_{\text{tot}-\hat{n}'}{}^{p} = \hat{n}'^{q} \leq 0.5$ 的最小 $\hat{p}$。将 $\hat{p}$ 替换为 $n\hat{p}_n$，并将 N_p（分别为 N_q）替换为 N'_p（分别为 N'_q），即可完成证明。 □

附录E

主动聚类问题的秩

在本节中，我们将证明，给定 M 个项目，在固定置信度设定下识别这些项目底层聚类的问题，其秩等于 $M-1$。在给出正式论证之前，我们先重述纯探索问题的秩的定义，该定义引自 [45]。

定义 1. [45, 定义 20] 固定常数 $d, S, M, R \in \mathbb{N}$. 若一个序贯识别问题由划分 $\mathcal{O} = \bigcup_{s=1}^{S} \mathcal{O}_s$ 指定，其中 $O_s \subseteq \mathbb{R}^d$ 对所有 $s \in [S]$ 均成立，则称其秩为 R，如果对每个 $s \in [S]$,

$$\mathcal{O} \setminus \mathcal{O}_s = \bigcup_{v=1}^{\bar{M}} \left\{ \boldsymbol{\lambda} \in \mathbb{R}^d : (\lambda_{k_1^{s,v}}, \ldots, \lambda_{k_R^{s,v}}) \in \mathcal{L}_{s,v} \right\} \tag{33}$$

对于一族指标 $k^{s}{}_{r}{}^{v} \in [d]$ 以及开集 $\mathcal{L}_{s,v}$，它们由 $r \in [R]$、$s \in [S]$ 和 $v \in [M]$ 索引。换言之，若对于每个 s，集合 $\mathcal{O} \setminus \mathcal{O}_s$ 都是有限个集合的并，且每个集合的定义仅涉及 R 元组，则该问题的秩为 R。

在我们的情境中，我们有

$$\mathcal{O} = \mathcal{C} = \bigcup_{K=1}^{M} \mathcal{C}^K = \bigcup_{K=1}^{M} \bigcup_{C \in \mathcal{C}^K} [C].$$

注意，形如 $[C]$的等价类划分了所有问题实例构成的空间 $\mathcal{C}$，并表示将 M 个项目划分为 $K \in [M]$ 个簇的所有可能的簇分配。因此，常数 M 等于 M 个项目所有可能划分的总数，这通常称为贝尔数 [41]。

若 $C \in \mathcal{C}^K$，则我们情境中与式 (33) 对应的结果为

$$\mathcal{C} \setminus [C] = \underbrace{\left(\bigcup_{\substack{\ell \in [M]: \\ \ell \neq K}} \bigcup_{C' \in \mathcal{C}^\ell} [C'] \right)}_{:= \mathcal{C}_1} \cup \underbrace{\left(\bigcup_{\substack{C' \in \mathcal{C}^K: \\ [C'] \neq [C]}} [C'] \right)}_{:= \mathcal{C}_2}.$$

接下来注意到，对于某个 $\ell \neq K$，每个 $C' \in \mathcal{C}_1$ 都由其对应的 $C'_=$ 矩阵中的 $M-\ell$ 个 1 指定，而每个 $C' \in \mathcal{C}_2$ 都由其对应的 $C'_=$ 矩阵中的 $M-K$ 个 1 指定。因此，问题的秩为

$$R = \max\{M - \ell : \ell \in [M]\} = M - 1.$$

相比之下，最佳臂识别问题的秩为 2，即使是在多目标最佳臂识别的情形下也是如此，正如 [51]中所指出的。因此，当 M取较大值时，众包聚类问题的秩远高于最佳臂识别问题的秩。

参考文献

[1] H. Driver 和 A. Kroeber，《文化关系的定量表达……》，《加利福尼亚大学美国考古学与民族学出版物》丛书。大学出版社，1932年。[在线]。网址：https://books.google.com.sg/books?id=4u9szQEACAAJ [2] E. H. Ruspini，“一种新的聚类方法”，《信息与控制》，第15卷，第1期，第22–32页，1969年。 [在线]。网址：https://www.sciencedirect.com/science/article/pii/S0019995869905919

[3] A. K. Jain、M. N. Murty 和 P. J. Flynn，“数据聚类：综述”，《ACM计算调查》，第31卷，第3期，第264–323页，1999年。[4] R. Xu 和 D. Wunsch，《聚类算法综述》。IEEE，2005年，第16卷。[5] J. B. MacQueen，“多变量观测的分类与分析方法”，载于《第五届伯克利数学统计与概率研讨会论文集》，第1卷。加利福尼亚大学出版社，1967年，第281–297页。[6] S. C. Johnson，“层次聚类方案”，《心理测量学》，第32卷，第3期，第241–254页，1967年。[7] A. Y. Ng、M. I. Jordan 和 Y. Weiss，“关于谱聚类：分析与算法”，载于《神经信息处理系统进展》，2002年，第849–856页。[8] A. Mazumdar 和 B. Saha，“带噪查询聚类”，载于《神经信息处理系统进展》，2017年，第5788–5799页。[9] I. P. Fellegi 和 A. B. Sunter，“记录链接理论”，《美国统计学会杂志》，第64卷，第328期，第1183–1210页，1969年。[10] A. K. Elmagarmid、P. G. Ipeirotis 和 V. S. Verykios，“重复记录检测：综述”，《IEEE知识与数据工程汇刊》，第19卷，第1期，第1–16页，2007年。[11] L. Getoor 和 A. Machanavajjhala，“实体解析：理论、实践与开放挑战”，载于《VLDB基金会会刊》。VLDB基金会，2012年，第5卷，第2018–2019页。[12] J. Leskovec、D. Huttenlocher 和 J. Kleinberg，“预测在线社交网络中的正向与负向链接”，载于《第19届国际万维网大会论文集》。ACM，2010年，第641–650页。

[13] M. Burke 和 R. E. Kraut，“收尾工作：维基百科晋升决策建模”，载于 ACM计算机支持的协同工作会议（CSCW）论文集，2008年，第27–36页。[14] M. J. Brzozowski、T. Hogg 和 G. Szabo，“朋友与敌人：意识形态社交网络”，载于 SIGCHI计算系统中的人为因素会议（CHI）论文集，2008年，第817–820页。[15] Y. Chen、M. Mitzenmacher 和 C. E. Tsourakakis，“复杂网络模型及其应用的推断”，载于第25届万维网国际会议（WWW）论文集，2016年，第689–699页。[16] N. Bansal、A. Blum 和 S. Chawla，“相关聚类”，载于《机器学习》，第56卷。施普林格，2004年，第89–113页。[17] N. Ailon、A. Bhattacharya、R. Jaiswal、R. Kumar、K. Nissim、B. Saha 和 S. Upadhyay，“近似相关聚类、基于同簇查询的聚类以及利用部分信息的聚类”，ACM经济学与计算汇刊（TEAC），第6卷，24，2018 1，1第。页 – 。[18] B. Saha 和 N. Subramanian，“基于同簇查询的相关聚类”，载于第36届国际机器学习会议（ICML）论文集，2019年，第5579–5588页。[19] P. W. Holland、K. B. Laskey 和 S. Leinhardt，“随机区块模型：初步研究”，《社会网络》，第5卷，第2期，第109–137页，1983年。[20] A. Condon 和 R. M. Karp，“植入划分模型上的图划分算法”，《随机结构与算法》，第18卷，第2期，第116–140页，2001年。[21] E. Abbe 和 C. Sandon，“一般随机区块模型中的社区检测：基本极限与高效算法”，载于 IEEE第56届计算机科学基础年会（FOCS）论文集，2015年，第670–688页。[22] E. Abbe，“社区检测与随机区块模型：近期进展”，《机器学习研究期刊》，第18卷，86，2018 1页 – 。[23] A. Garivier 和 E. Kaufmann，“固定置信度下的最优最佳臂识别”，载于学习理论会议。PMLR，2016年，第998–1027页。[24] Y. Chen、R. K. Vinayak 和 B. Hassibi，“通过主动查询实现众包聚类：具有理论保证的实用算法”，《AAAI人类计算与众包会议论文集》，第11卷，第1期，第27–37页，2023年11月。 [在线]。网址：https://ojs.aaai.org/index.php/HCOMP/article/view/27543[25] C. S. Mukherjee、P. Peng 和 J. Zhang，“随机区块模型中不平衡社区的恢复及其在错误预言机聚类中的应用”，《神经信息处理系统进展》，第36卷，第20 540–20 561页，2023年。[26] K. Green Larsen、M. Mitzenmacher 和 C. Tsourakakis，“使用错误预言机进行聚类”，载于《2020年万维网大会论文集》，丛书 WWW ’20。美国纽约州纽约：计算机协会，2020年，第2831–2834页。 [在线]。网址：https://doi.org/10 1145 3366423 3380045 。 / 。P. Peng 和 J. Zhang，“迈向查询最优且时间高效的错误预言机聚类算法”，arXiv预印本 arXiv:2106 10374，2021 . 。[27] [28] J. Xia 和 Z. Huang，“通过多臂老虎机实现含噪查询下的最优聚类”，载于国际机器学习会议。PMLR，2022年，第24 315–24 331页。[29] R. K. Vinayak、S. Oymak 和 B. Hassibi，“缺失数据下的图聚类：凸算法与分析”，载于《神经信息处理系统进展》，Z. Ghahramani、M. Welling、C. Cortes、N. Lawrence 和 K. Weinberger 编，第27卷。Curran Associates, Inc.，2014年。 [在线]。网址：https://proceedings.neurips.cc/paper files/paper/2014/file/9a2cb485e69e5ed18869dfcab39ab4b6-Paper.pdf [30] R. K. Vinayak 和 B. Hassibi，“众包聚类：查询边还是三角形”，载于《神经信息处理系统进展》，2016年，第1316–1324页。[31] R. G. Gomes、P. Welinder、A. Krause 和 P. Perona，“众包聚类”，载于《神经信息处理系统进展》，第24卷，2011年，第558–566页。[32] S. Ibrahim 和 X. Fu，“通过系统化边查询实现混合成员图聚类”，《IEEE信号处理汇刊》，第69卷，第5189–5205页，2021年。[33] Y. Jedra 和 A. Proutiere，“线性老虎机中的最优最佳臂识别”，《神经信息处理系统进展》，第33卷，第10 007–10 017页，2020年。[34] J. Yang、Z. Zhong 和 V. Y. F. Tan，“具有赌博机反馈的在线聚类”，《机器学习研究期刊》，第25卷，54，2024 186，1第。页 – 。[35] Y. Yue、B. Josef、R. Kleinberg 和 T. Joachims，“k臂对决老虎机问题”，《机器学习研究期刊》，第78卷，第5期，第1538–1556页，2012年。[36] T. T. Nguyen 和 H. W. Lauw，“上下文多臂老虎机的动态聚类”，载于第23届ACM信息与知识管理国际会议论文集，2014年，第1959–1962页。[37] C. Gentile、S. Li 和 G. Zappella，“老虎机的在线聚类”，载于国际机器学习会议。PMLR，2014年，第757–765页。[38] S. Li、A. Karatzoglou 和 C. Gentile，“协同过滤老虎机”，载于第39届ACM SIGIR信息检索研究与发展国际会议论文集，2016年，第539–548页。[39] D. Bouneffouf、S. Parthasarathy、H. Samulowitz 和 M. Wistuba，“在多臂老虎机中最优利用聚类与历史信息”，载于第28届人工智能国际联合会议论文集，2019年，第2016–2022页。[40] R. Singh、F. Liu、Y. Sun 和 N. Shroff，“具有相关臂的多臂老虎机”，arXiv预印本 arXiv:2010.09478，2020年。[41] L. Moser 和 M. Wyman，“贝尔数的渐近公式”，载于《贝尔数的渐近公式》。加拿大皇家学会，1955年。[42] A. Russo 和 A. Proutiere，“关于具有全局与局部结构的多任务老虎机中表征学习的样本复杂度”，《AAAI人工智能会议论文集》，第37卷，第8期，第9658–9667页，2023年6月。 [在线]。网址：https://ojs.aaai.org/index.php/AAAI/article/view/26155[43] E. O. Rivera 和 A. Tewari，“逻辑回归老虎机中的近似最优纯探索”，arXiv预印本 arXiv:2410.20640，2024年。[44] C. R´eda、A. Tirinzoni 和 R. Degenne，“处理固定置信度线性前m名识别中的模型设定错误”，《神经信息处理系统进展》，第34卷，第25 489–25 501页，2021年。[45] E. Kaufmann 和 W. M. Koolen，“混合鞅再探：在序贯检验与置信区间中的应用”，《机器学习研究期刊》，第22卷，第246期，第1–44页，2021年。[46] L. M. Ausubel 和 R. J. Deneckere，“广义最大值定理”，《经济理论》，第3卷，第1期，第99–107页，1993年。[47] E. Kaufmann、O. Capp´e 和 A. Garivier，“多臂老虎机模型中最佳臂识别的复杂度”，《机器学习研究期刊》，第17卷，第1期，第1–42页，2016年。[48] C. Berge，《拓扑空间：包括多值函数、向量空间与凸性的论述》，未删节重印版。纽约州米尼奥拉：多佛出版社，1997年。[49] W. Hoeffding，“有界随机变量之和的概率不等式”，《美国统计学会杂志》，第58卷，第301期，第13 30, 1963 [ ] 2282952 – 3[50] 。在线。网址：http://www.jstor.org/stable/ G. Grimmett 和 D. Stirzaker，《概率与随机过程》，rd 版。牛津大学出版社，2001 年。

[51] Z. Chen、P. N. Karthik、Y. M. Chee 和 V. Y. F. Tan，“固定置信度下的最优多目标最佳臂识别”，载于第*28*届人工智能与统计国际会议论文集，系列：机器学习研究论文集，Y. Li、S. Mandt、S. Agrawal 和 E. Khan 编，第258卷。PMLR，2025年5月3日至5日，第1162–1170页。 [在线]。获取地址：https://proceedings.mlr.press/v258/chen25b.html