Biometrika（2026年），
xx，x，*p*.1印于大不列颠

arXiv:2603.04608v1 [math.ST] 4 Mar 2026

KRAFTY：联合簇恢复的 Khatri-Rao 框架

作者：高思怡统计学系，弗吉尼亚大学，美国弗吉尼亚州夏洛茨维尔，邮编*22904*sg7nx@virginia.edu

Zachary Lubberts
统计学系，弗吉尼亚大学，美国弗吉尼亚州夏洛茨维尔，邮编 *22904*zlubberts@virginia.edu

Marianna Pensky
数学系，中佛罗里达大学，美国佛罗里达州奥兰多，邮编 *32816*Marianna.Pensky@ucf.edu

摘要

当多个数据集描述同一组实体的互补信息时，例如某个人在不同时期的脑部扫描、跨年度的全球贸易网络，或用户在不同社交媒体平台上的信息，将这些快照整合起来，可以让我们看到更全面的图景。识别数据结构的一种常见方式是聚类；虽然可以分别对每个数据集进行聚类，但在多视图设定下，通过识别联合簇，我们能够获得更多信息。我们研究一种聚类问题：每个视图都将若干联合簇混合在一起，只揭示部分信息，而我们的目标是恢复真实的聚类结构。我们提出了这一多视图聚类模型及其恢复方法：转置 Khatri–Rao 联合簇恢复框架（KRAFTY）。该模型具有灵活性，可适用于多种数据生成过程，包括随机点积图中的潜在位置和高斯混合。KRAFTY 的一项关键优势在于，它能在足够维度的空间中表示联合簇，使得每个联合簇在转置 Khatri-Rao 矩阵中占据一个正交子空间，因此碎石图会在真实联合簇数量处出现明显的骤降，从而便于进行模型选择。我们的模拟结果表明，当联合簇的数量超过各个视图中簇数量之和时，无论是在联合聚类准确率还是联合簇数量的估计方面，我们的方法都优于现有方法。

一些关键词：社区检测；Khatri-Rao积；多视图聚类

1. 引言

1.1. 背景与动机

社区检测和聚类是统计学与机器学习中的核心任务 Jaeger & Banks (2023); Milligan & Cooper (1987);Abbe (2023);Fortunato & Newman (2022)。聚类

c○ 2021 Biometrika Trust

分析是指将对象划分为若干组（簇），使组内相似性高、组间相似性低；相似性可由网络连通性、特征空间距离或其他亲和性来定义 Jaeger & Banks (2023); Milligan & Cooper (1987); von Luxburg (2007)。聚类分析已应用于多种现代领域的数据，例如基因图谱 Toh & Horimoto (2002) 和社交网络快照 Handcock等人 (2007)，以发掘有意义的潜在模式；此外，针对单数据集聚类，人们已开发出众多技术 Jaeger & Banks (2023); Milligan & Cooper (1987); von Luxburg (2007);Jin etal. (2023)。

多视图设定下的早期研究侧重于从多个来源中恢复单一且一致的聚类结果 Abbe（2023）；Fortunato & Newman（2022）；Jin等人（2023）。近年来，多视图聚类受到越来越多的关注 Yang & Wang（2018）；Vall`es-Catal`a等人（2016），因为许多应用会提供来自不同来源或视图的数据集，而联合聚类能够得到信息量更丰富的划分。人们也日益关注研究不同网络或聚类结果之间子空间的差异 MacDonald等人（2021）；Gallagher等人（2021）；Arroyo等人（2020）；Agterberg等人（2025），由此涌现出许多多视图谱聚类方法 Jing 等人（2021）；Lei & Lin（2023）。一种常见的多视图谱聚类方法是将各视图的谱嵌入拼接起来，再进行第二次谱分解，随后执行聚类；Arroyo 等人（2020）将其称为多重邻接谱嵌入（**MASE**）。这类方法有两个缺点。首先是秩亏：每个视图的嵌入维数通常设为该视图自身的簇数，因此拼接得到的空间维度至多为各视图簇数之和，而联合簇的真实数量可能高达各视图簇数的乘积。其次是难以确定联合簇数：这些方法通常缺乏明确的联合簇数估计器，而是依赖碎石图中的“肘部”，但肘部可能并不明确。为应对这些挑战，我们提出转置 Khatri-Rao 积框架。该框架为估计联合簇数提供了足够的秩，并且在估计联合簇数和联合聚类方面均表现出色。

1.2. 记号

对于正整数 n，记 $[n] = \{1, \dots, n\}$。我们用 $\mathrm{O}_{r,k}$ 表示具有标准正交列的 $r \times k$矩阵集合，其中 $\mathrm{O}_r = \mathrm{O}_{r,r}$。我们用 $\mathrm{C}_{n,K}$表示将 n 个对象划分为 K 个簇的聚类矩阵集合。这里，$\mathrm{C}_{n,K}$ 由 $n \times K$ 个二元矩阵组成，每个矩阵的每行恰有一个值为 1，其余均为 0。我们分别用 $A(i,:)$ 和 $A(:, j)$表示矩阵 A 的第 i行和第 j列。我们用 I_m表示 m 阶单位矩阵。我们用 $\mathrm{SVD}_K(X)$表示矩阵 X 的 K 个左奇异向量组成的矩阵。我们分别用 $\|\cdot\|_F$、$\|\cdot\|$ 和 $\|\cdot\|_{2,\infty}$ 表示 Frobenius 范数、谱范数和 $2 \to \infty$ 矩阵范数，其中$\|A\|_{2,\infty} = \max_i \|A(i,:)\|$。我们分别用 $A \circ B$ 和 $A \otimes B$表示矩阵 A和B 的 Hadamard 积和克罗内克积。对于具有相同行数的矩阵 $A \in \mathbb{R}^{n \times m}$ 和 $B \in \mathbb{R}^{n \times q}$，定义转置 Khatri-Rao 积为 $A \bigcirc\!\!\!\!\!\!\text{KR}\; B \in \mathbb{R}^{n \times mq}$，其中对于 $i \in [n]$,

$$(A \text{KR} B)(i,:) = A(i,:) \otimes B(i,:) = [A(i,1)B(i,:) \,|\, \cdots \,|\, A(i,m)B(i,:)].$$

该乘积的一些性质见附录 A.6。我们使用 C_τ 表示仅依赖于 τ 的通用常数，并且该常数在全文中可以取不同的值。

1.3. *The Model*

本文考虑一种常见情形：数据来自 V 个来源（视图）。每个 V 视图都对应一个聚类分配 $z_v : [n] \to K_v$，以及相应的聚类矩阵 $Z_v \in \{0,\ 1\}^{n \times K_v}$，其中 K_v表示视图 $v \in [V]$中的簇数。

这些聚类分配可以具有任意联合配置。我们的目标是恢复联合聚类分配函数 z：$[n] \to [K]$，或相应的聚类矩阵 $Z \in \{0, 1\}^{n\times K}$，以描述簇的联合配置。具体而言，设联合分配中的簇总数K为某个值

$$\max_{v\in[V]} K_v \le K \le \prod_v K_v, \tag{1}$$

并令 $Z \in \{0,1\}^{n\times K}$ 为真实联合聚类矩阵，每一行恰有一个非零元素。个体聚类矩阵 $Z_v = ZP_v$ 中，$P_v \in \{0,1\}^{K\times K_v}$ 的每一列至少有一个非零元素，且每一行恰有一个非零元素。这确保所有个体聚类结果都与联合聚类一致，但可能会将一些联合簇合并在一起。通常，各视图单独包含的簇数都少于真实联合簇的数量。

接下来，我们假设 V 较小，且数据由以下两者之一构成：特定视图的估计聚类矩阵 $Z_v \in \mathbb{C}_{n,\ K_v}$，$v \in [V]$，或可用于聚类的奇异向量估计矩阵 $U_v \in \mathbb{O}_{n,\ K_v}, v \in [V]$。我们不对数据来源作其他假设；不过，一种常见情形是对 V 个不同视图应用 k-均值算法所得的结果。

尽管这个问题乍看起来很简单，但实际上并非如此。考虑一个简单的情形，其中$V = 2$。设 $K_1 = K_2 = 3$，令 n 很大，并假设两个视图中的簇均以 1/3 的概率均匀随机生成。那么，聚类分配 z_1 和 z_2 相互独立，组合聚类由 $K_1 \times K_2 = 9$ 个大小几乎相等的组构成。若使用拼接矩阵 $\widetilde{Z} = [Z(1)|Z(2)] \in \{0,1\}^{n\times 6}$，其中 $Z(1), Z(2)$是各个聚类矩阵，作为联合聚类的基，则 $\mathrm{rank}(Z) = 5$。例如，若 $n = 1000$，则 Z 的奇异值为 (25.8852, 18.7956, 18.5849, 18.2198, 17.3010, 0)。因此，矩阵 Z 秩亏，要确定联合分配中的簇数极具挑战性。

作为补救措施，我们在此提议将 $Z^{(1,2)} = Z^{(1)} \circ Z^{(2)}$。很容易发现，矩阵 $Z^{(1,2)} \in \mathbb{C}_{n,K_1K_2}$，因此它是一个聚类矩阵，并且具有 9 个几乎相同的奇异值（11.1803、11.0454、10.9087、10.7238、10.6301、10.5830、10.3923、9.8489、9.4340）。我们稍后将证明，一般而言，矩阵 $Z^{(1,2)}$ 的秩可用作联合分配中簇数的代理指标，而其估计版本（以及基于奇异值估计矩阵的版本）为组合聚类分配恢复奠定了坚实基础。

因此，我们提出了用于联合簇恢复的转置 Khatri-Rao 框架（KRAFTY）。KRAFTY 的一项关键优势是，它能在维度足够的空间中表示联合簇，从而更可靠地估计联合簇的数量。图 1 展示了应用转置 Khatri-Rao 积的核心思想。这里，$Z(1)$和 $Z^{(2)}$是针对同一组实体的两个数据集的聚类矩阵。在这个例子中，$V = 2$，且两个视图都有 $K_1 = K_2 = 3$ 个簇。簇 1 在两个视图中是相同的，而视图1中簇 2 和簇 3 的元素可能分散在视图2的簇 2 和簇 3 中。这会形成一个包含 $K = 5$ 个不同簇的联合配置。矩阵 $Z^{(1,2)} = Z^{(1)} \circ Z^{(2)}$ 有 $3 \times 3 = 9$ 列，其中索引为 (k_1, k_2) 的列对应于视图1和视图2中簇 k_1 和 k_2 的组合。容易看出，$Z^{(1,2)}$ 中仅有非空簇组合对应的列为非零，因此它有 4 个零列。同样，如果 $n = 1000$，矩阵 $Z^{(1,2)}$ 有 5 个较大的奇异值（18.7617, 13.1149,12.8452, 12.4900, 12.4499），之后是 4 个零。

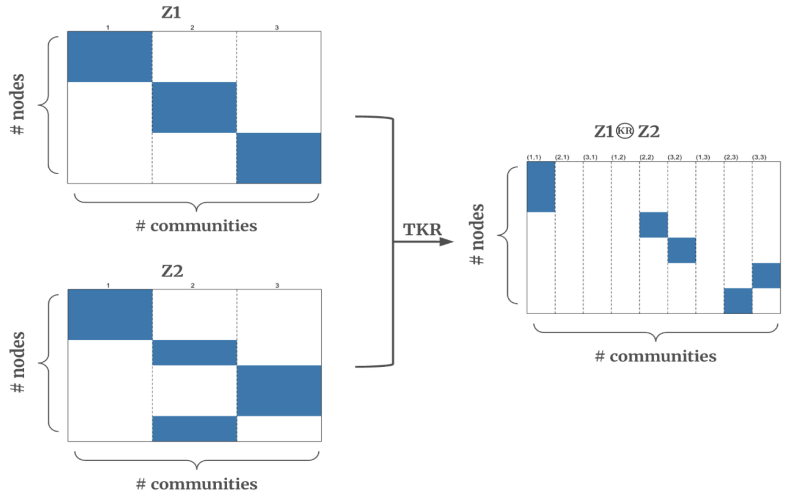


图 1：转置 Khatri-Rao 积背后直觉的图形示意。

1.4. 相关工作

近年来，基于社区的统计模型受到广泛关注，因为这类模型将簇视为数据背后的生成机制，而不仅仅是数据的特征。Fortunato & Newman（2022）。突出的例子包括高斯混合模型（高斯混合模型）和随机块模型（随机块模型）Holland 等（1983）。这些模型推动了单个数据集上的社区检测研究蓬勃发展 Abbe（2023）；Fortunato & Newman（2022），也催生了一系列谱聚类方法（Jin（2015）；von Luxburg（2007）；Qin & Rohe（2013）；Rohe 等（2011）；Lei & Lin（2023）；L¨offler 等（2020））。其核心思想是利用矩阵分解（例如特征分解或奇异值分解）提取低维嵌入，再使用 k-均值或层次聚类等方法对其进行聚类 von Luxburg（2007）。这些方法在实践中表现出色，并具有坚实的理论保证：在适当条件下，它们是一致的，甚至能够实现完美聚类（Lei & Rinaldo（2015）；Lyzinski 等（2015）；L¨offler 等（2020）；Pensky（2024）；Sussman 等（2012））。我们的方法利用每个数据集的谱嵌入，建立在单源谱聚类的优良性质之上。

转向多视图聚类，近期已有若干研究在多层随机块模型框架下开展（Vall`es-Catal`a 等人（2016））。研究者开发了多种聚类方法，包括谱聚类方法（Jing 等人（2021）；Lei & Lin（2023）；Agterberg等人（2025）；Domenico 等人（2014）；Han 等人（2015））和矩阵分解（Paul & Chen（2018））。许多方法会先从每个图中获取谱嵌入，再将其拼接起来进行后续谱分解，例如 MASE（Arroyo 等人（2020））和 UASE（Gallagher等人（2021））。然而，这种并排拼接存在局限：当联合簇的数量超过各层簇数之和时，由于每层的嵌入维数通常设为该层自身的簇数，拼接矩阵的最大秩至多为这些簇数之和，因此表示可能会秩亏，无法捕捉完整的联合结构。另一项重要研究方向是估计多视图数据中的簇数。现有方法通常依赖于对碎石图“肘部”的目视检查（Ledesma 等人（2015）），或使用 Zhu & Ghodsi（2006）提出的自动维数选择方法。一个常见问题是，特征值可能没有明显的下降，导致即使采用自动程序也难以确定维度。为应对这些挑战，我们

的工作提出了转置 Khatri-Rao 积框架，该框架能够提供足够的秩来估计联合簇的数量，并且在联合簇数量对应的位置形成清晰的“肘部”。

本文其余部分安排如下。第 2 节介绍我们的模型，并提出转置 Khatri-Rao 积方法。第 3 节研究所提方法的理论统计性能。第 4 节评估该方法在聚类和估计联合簇数量方面的实证性能，并将其与 MASE 进行比较。第 5 节展示该方法在 FAO 食品和农产品贸易数据上的应用，说明该方法能够揭示全球贸易网络中国家的联合簇。

2. 方法

2.1. *KRAFTY* 联合聚类方法

由于 V 是一个较小的整数，为简单起见，我们考虑 $V = 2$的情况。在介绍方法之前，我们先仔细考察个体聚类矩阵 $Z(1)$、 $Z(2)$、它们的转置 Khatri-Rao 积

$$Z^{(1,2)} = Z^{(1)} \circledR Z^{(2)} \in \{0,1\}^{n \times K_1 K_2} \tag{2}$$

以及真实联合聚类矩阵 $Z \in \mathsf{C}_{n,K}$之间的关系。

从第 1.3 节和图 1 可以很容易看出，矩阵 Z 可以通过删除矩阵 $Z^{(1,2)}$ 中的零列得到。为此，可以引入矩阵 $H \in \{0,1\}^{K_1 K_2 \times K}$，其中 H 由单位矩阵中与 $Z(1,2)$ 的非零列相对应的列构成。因此，稍微滥用一下符号，对于 $k_1 \in [K_1]$、 $k_2 \in [K_2]$和 $k \in [K]$，有 $H((k_1, k_2), :) = 0$当且仅当 $Z(1,2)(:, (k_1, k_2)) = 0$，且矩阵 H 的每一列都有一个元素等于 1。这些定义意味着以下关系

$$Z = Z^{(1,2)} H, \quad Z^{(1,2)} = Z H^{\top}, \quad H^{\top} H = I_K, \ H \in O_{K_1 K_2, K}. \tag{3}$$

此处， K 满足条件(1)，从而 $K \le K_1 K_2$。此外， $Z \in \mathsf{C}_{n,K}$，因此它是一个聚类矩阵。

算法1. 使用KRAFTY进行联合聚类

输入： $\widehat{Z}_{(v)} \in \mathsf{C}_{n,K_v}$或 $\widehat{U}_{(v)} \in \mathsf{O}_{n,K_v}$, $v = 1,2$；最终簇数 K；参数 r
$\epsilon > 0$**输出：** 估计的联合聚类函数$\hat{z}$： $[n] \to [K]$**步骤：** **1：** 获取矩阵 U，使用公式(6)或(8)**2：** 使用$(1+\epsilon)$-近似 k-均值聚类，将 U 的 n 行聚类为 K 个簇，从而得到估计的聚类函数$\hat{z}$。

如果已知真实矩阵 $Z(1)$和 $Z(2)$，那么 (3) 中的矩阵 Z 就能给出联合聚类分配。然而，我们得到的要么是它们的估计版本 $\widehat{Z}_{(v)} \in \mathsf{C}_{n,K_v}$，要么是奇异向量的估计矩阵 $\widehat{U}_{(v)} \in \mathsf{O}_{n,K_v}$, $v = 1,2$，这些矩阵可用于求出各自的聚类分配。为设计统一的聚类程序，考虑对角

矩阵

$$D^{(1,2)} = \left(Z^{(1,2)}\right)^{\top} Z^{(1,2)}, \quad D = Z^{\top}Z, \tag{4}$$

并注意到，由于 (3)，有 $D = H^{\top}D\,(1,\,2)\,H$。这里，$D(k,\ k) = n_k$，即联合簇 $k \in [K]$ 中元素的数量。由 (2) 和 (4) 可得

$$U = ZD^{-1/2} = Z^{(1,2)}HD^{-1/2} \in \mathcal{O}_{n,K} \quad \Longrightarrow \quad Z^{(1,2)} = \left(Z^{(1)} \circledKR Z^{(2)}\right) = UD^{1/2}H^{\top}, \tag{5}$$

并且第二个等式给出了矩阵 $Z\,(1,\,2)$ 的奇异值分解。这里，U 的行彼此不同，并对应于 K 个簇，因此可以通过对矩阵 U 的行进行聚类来获得聚类分配 $z\colon [n] \to [K]$。此外，U 可以作为 $Z\,(1,\,2)$ 的前 K 个左奇异向量获得，即 $U = \text{奇异值分解}_K\left(Z\,(1,\,2)\right)$。当有 $\widehat{Z}\,(v)$ 而不是 $Z\,(v)$ $v = 1,\ 2$ 时，矩阵 U 可以替换为

$$\widehat{U} = \mathrm{SVD}_K\left(\widehat{Z}^{(1,2)}\right), \quad \widehat{Z}^{(1,2)} = \left(\widehat{Z}^{(1)} \circledKR \widehat{Z}^{(2)}\right). \tag{6}$$

现在，考虑数据由矩阵组成的情况

$$U^{(v)} = Z^{(v)}\left(D^{(v)}\right)^{-1/2} \in \mathcal{O}_{n,K_v}, \quad D^{(v)} = (Z^{(v)})^{T}Z^{(v)},\ v = 1, 2.$$

然后，根据 Khatri-Rao积的性质**P1**（见附录 A.6）以及克罗内克积的性质，可得

$$U^{(1)} \circledKR U^{(2)} = \left(Z^{(1)}(D^{(1)})^{-1/2}\right) \circledKR \left(Z^{(2)}(D^{(2)})^{-1/2}\right) = Z^{(1,2)}\left(D^{(1)} \otimes D^{(2)}\right)^{-1/2}$$

因此，由式（5）可知，U和 $U^{(1)}{}_{\mathrm{KR}}\ U^{(2)}$之间的关系为

$$U^{(1,2)} := U^{(1)} \circledKR U^{(2)} = UD^{1/2}H^{\top}(D^{(1)} \otimes D^{(2)})^{-1/2}.$$

该等式右侧并非 $U(1,2)$的奇异值分解。为得到 $U(1,2)$的奇异值分解，考虑一个对角矩阵 $\widetilde{D}$：当第 k 个联合簇对应于单个视图中的簇 k_1 和 k_2 时，该矩阵在第 k 个对角元素处放置 $n_{k_1}^{(1)}n_{k_2}^{(2)}$：

$$\widetilde{D} = H^{\top}\left(D^{(1)} \otimes D^{(2)}\right)H \in \mathbb{R}^{K\times K}.$$

$\widetilde{D} \neq D$由于 $D(1,2) \neq D(1) \otimes D(2)$。我们有 $H^{\top}\left(D(1) \otimes D(2)\right)^{-1/2} = \widetilde{D}^{-1/2}H^{\top}$，因此

$$U^{(1,2)} = (U^{(1)} \circledKR U^{(2)}) = U\,D^{1/2}\widetilde{D}^{-1/2}H^{\top}. \tag{7}$$

由于矩阵 $D^{1/2}\widetilde{D}^{-1/2}$是对角的，该方程给出了 $U^{(1,2)}$的一个奇异值分解，其左奇异向量矩阵为 U，即 $U = \text{奇异值分解}_K\left(U(1,2)\right)$.同样，如果可用的是矩阵 $\widehat{U}(v)$而非 $U(v)$，$v = 1,2$，则矩阵 U可以替换为 $\widehat{U}$，其中

$$\widehat{U} = \mathrm{SVD}_K\left(\widehat{U}^{(1,2)}\right), \quad \widehat{U}^{(1,2)} = \left(\widehat{U}^{(1)} \circledKR \widehat{U}^{(2)}\right). \tag{8}$$

附录图 13 展示了将 $\widehat{U}^{(v)}$或 $\widehat{Z}\,(v)$ 作为 KRAFTY 的输入时的结果可视化比较。

2.2. 联合聚类算法

要从转置 Khatri-Rao 矩阵中获得联合聚类，只需对通过公式 (6) 或 (8) 得到的 $\hat{U}$ 的行应用聚类算法即可。具体过程见算法1。

如上所述，获得联合簇的另一种方法是拼接矩阵 $\widehat{Z}^{(v)}$ 或 $\widehat{U}_{(v)}$，$v = 1, 2$。我们将此方法称为 MASE，因为它可视为 Arroyo 等人（2020）中 MASE 的一个版本。步骤见算法 2。

算法 2. 基于 MASE 的联合聚类

输入： $\widehat{Z}_{(v)} \in \mathbb{C}^{n,\ K_v}$ 或 $\widehat{U}_{(v)} \in \mathbb{O}^{n,\ K_v}$，$v = 1, 2$；最终簇数 K；参数 r $\epsilon > 0$ **输出：** 估计的联合聚类函数 $\hat{z}: [n] \to [K]$ **步骤：**

p

1: Construct matrix $\widetilde{Z} = \left[\widehat{Z}^{(1)} \mid \widehat{Z}^{(2)}\right]$ or $\widetilde{U} = \left[\widehat{U}^{(1)} \mid \widehat{U}^{(2)}\right]$

2: Find $\widehat{U} = \mathrm{SVD}_K\left(\widetilde{Z}\right)$ or $\widehat{U} = \mathrm{SVD}_K\left(\widetilde{U}\right)$

3: 使用 $(1+\epsilon)$-近似 k-均值聚类，将 U 的 n 行聚类为 K 个簇，以得到估计的聚类函数 $\hat{z}$。

第 4 节将详细比较算法 1 和算法 2。这里，我们只作几点说明。矩阵 Z 是聚类矩阵，而 Z 不是。此外，虽然 $\mathrm{rank}(Z_{(1,2)}) = \mathrm{rank}(U_{(1,2)})$ 等于簇数，但 $\widetilde{Z} = [Z_{(1)} | Z_{(2)}]$ 或 $\widetilde{U} = [U_{(1)} | U_{(2)}]$ 并非如此。矩阵 $\widetilde{Z}$ 和 $\widetilde{U}$ 的秩至多为 $K_1 + K_2$，因此当 $K > K_1 + K_2$ 时可能秩亏，正如我们在第 1.3 节所展示的那样。

算法 1 和算法 2 中的聚类算法建议采用 $(1+\epsilon)$-近似 k-均值聚类，但这并非唯一可行的选择。例如，可以使用层次聚类或任何其他合适的聚类算法。在本文的所有模拟和真实数据分析中，我们都使用层次聚类，因为我们发现，在重复测试中，它比 k-均值或高斯混合建模产生的结果更稳定。

注记*1.* **情形** $V \geq 3$。虽然本节仅给出 $V = 2$的情形，但将KRAFTY 方法推广到 $V \geq 3$的情形相当直接。具体而言，对于算法1，公式(6)中的 $\widehat{Z}_{(1,2)}$ 以及(8)中的 $\widehat{U}^{(1,2)}$ 应分别替换为

$$\widehat{Z}^{(1,\ldots,V)} = \left(\widehat{Z}^{(1)} \circledKR \ldots \circledKR \widehat{Z}^{(V)}\right), \quad \widehat{U}^{(1,\ldots,V)} = \left(\widehat{U}^{(1)} \circledKR \ldots \circledKR \widehat{U}^{(V)}\right).$$

若使用算法2，则在步骤1中，

$$\widetilde{Z} = \left[\widehat{Z}^{(1)} \mid \ldots \mid \widehat{Z}^{(V)}\right], \quad \widetilde{U} = \left[\widehat{U}^{(1)} \mid \ldots \mid \widehat{U}^{(2)}\right].$$

由于秩$(Z) \leq K_1 + \cdots + K_V$ 可能远小于 K，后者将再次导致秩亏聚类。

2.3. 一个常见的数据驱动示例。

虽然本文对数据生成过程不作任何假设，但最常见的情形之一是 k-均值场景，其中观测数据 v 表示为矩阵

$$Y^{(v)} = G^{(v)} + \Xi^{(v)} \in \mathbb{R}^{n \times d_v}, \quad v \in [V], \tag{9}$$

其中矩阵 $G_{(v)}$有 K_v 个不同的行，并且 $\mathbb{E}(\Xi^{(v)}) = 0$。后者意味着存在聚类函数 z_v: $[n] \to [K_v]$，使得 $G^{(v)}$的第 i 行为 $G_{(v)}(i,:) = \mu_k^{(v)}, i \in [n], k \in [K_v]$。如果 $\mathrm{M}_{(v)} \in \mathbb{R}^{K_v \times d_v}$的行是 $\mu_k^{(v)}$，那么用矩阵形式表示为

$$G^{(v)} = Z^{(v)} M^{(v)}, \quad v = 1, 2. \tag{10}$$

在上述情况下，可以得到矩阵 $\widehat{U}_{(v)} =$ 奇异值分解$_{K_v}(Y_{(v)})$，还可以对$\widehat{U}_{(v)}$的行进行聚类，从而得到矩阵 $\widehat{Z}_{(v)}$、 $v \in [V]$。式（9）和（10）中的矩阵 $Y_{(v)}$ 和 $G_{(v)}$ 可以来自任意来源。例如， $Y_{(v)}$ 的行 $Y_{(v)}(i, :)$ 可以相互独立，其协方差矩阵由$k = z_v(i)$决定，因此 $\mathrm{Cov}\left(Y_{(v)}(i, :)\right) = \Sigma_k^{(v)}$; 或者，它们可以是随机图的邻接矩阵。

为了进一步说明转置 Khatri-Rao 积如何用于恢复联合簇，在图 2 中，我们以可视化方式展示了将我们的方法应用于由具有三个潜在簇的随机块模型生成的两个邻接矩阵的过程。我们假设联合聚类结构遵循图 1，因此共有五个联合簇。我们首先分别对每个矩阵执行奇异值分解，并提取矩阵 $\widehat{U}_{(v)}, v = 1, 2$。然后，我们使用公式（8）构造 $\widehat{U}_{(1,2)}$，并得到 $\widehat{U}$，它是 $\widehat{U}_{(1,2)}$ 的左奇异向量矩阵。 $\widehat{U}_{(1,2)}$ 的奇异值碎石图显示，第五个奇异值之后存在明显间隔，表明有 $K = 5$个潜在联合簇。最后，将 k-均值算法与 $K = 5$结合，应用于矩阵 $\widehat{U}_{(1,2)}$ 的前五个左奇异向量，以恢复联合簇。

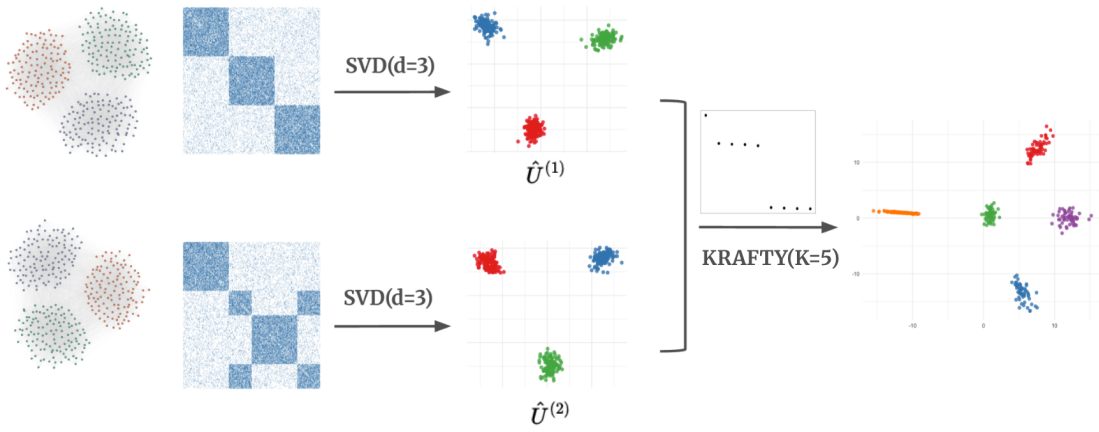


图 2：一个常见数据驱动示例的图形表示。

3. 理论结果

3.1. 假设

本节给出使用 KRAFTY（算法1）进行联合簇恢复的保证。与之前一样，为简单起见，我们考虑 $V = 2$ 的情形，并将真实聚类矩阵记为 $\widehat{Z}_{(v)}$、 $v = 1, 2$。我们假设数据以 $\widehat{Z}_{(v)}$ 或 $\widehat{U}_{(v)}$、 $v = 1, 2$ 的形式给出。对于前一种情形，令 $J_v \subseteq \{1, \ldots, n\}$ 表示被 $\widehat{Z}_{(v)}$ 错误分类的元素集合，并令 $\mathrm{E}_{v,n} = |J_v|$be$\widehat{\ }$ 表示其基数。类似地，我们将真实聚类矩阵记为 Z，其估计版本记为 Z。$\widehat{\ }$令 $J \subseteq \{1, \ldots, n\}$ 表示被 Z 错误分类的元素集合，并令 $\mathrm{E}_n = |J|$ 表示聚类错误的数量。

如果数据以 $\widehat{U}(v)$、 $v = 1$、2 的形式给出，我们将对 $U(v)$ 与 $\widehat{U}(v)$ 之差的范数给出上界。由于矩阵 $\widehat{U}(v)$ 仅在旋转意义下确定，我们考虑矩阵 $W(v)\,U$，使其在弗罗贝尼乌斯范数意义下尽可能接近 $U(v)$ 与 $\widehat{U}^{(v)}$。特别地，若 $(U(v))^{\top}\widehat{U}(v) = W_1 D_W W_2^T$ 是 $(U(v))^{\top}\widehat{U}(v)$ 的奇异值分解，则 W $(U^{v} = W_1 W_2^T \in)$ $\mathrm{O}K_v$。

我们称由 $\widehat{Z}$ 得到的聚类是一致 的，如果 $n^{-1}\mathrm{E}_n \to 0$ 随着 $n \to \infty$。如果$\mathrm{E}_n \to 0$ 随着 $n \to \infty$，我们称该聚类是完美的。接下来，我们证明在简明的正则性假设下，算法1的聚类精度与个体聚类一样好，并且还可以正确识别联合簇数。我们首先陈述一些正则性假设。

假设*1*（均衡的簇大小）。 假定单视图聚类与联合聚类都具有均衡的簇大小：若 $n_k^{(v)}$ 是视图 v中簇 k 的元素个数， $k \in [K_v]$，且 n_k 是联合簇 $k \in [K]$的元素个数，则对某些正常数，有

$$\left(c_o^{(v)}\right)^2 \frac{n}{K_v} \le \min_k n_k^{(v)} \le \max_k n_k^{(v)} \le \left(C_o^{(v)}\right)^2 \frac{n}{K_v}, \tag{11}$$

$$c_o^2 \frac{n}{K} \le \min_k n_k \le \max_k n_k \le C_o^2 \frac{n}{K}. \tag{12}$$

假设 *2*（有限个簇）。 每个视图中的簇数，因而联合簇数，都是有限的

$$K_1 = O(1), \quad K_2 = O(1), \quad K = O(1), \quad n \to \infty. \tag{13}$$

假设 *3*（各视图中的一致聚类）。 若数据形如 $\widehat{Z}_{(v)}$， $v = 1, 2,$，则聚类错误数$\mathrm{E}_{v,n}$在 $\widehat{Z}_{(v)}$ 的基础上满足

$$n^{-1}\,\mathcal{E}_{v,n} \;\to\; 0, \quad \text{as } n \to \infty. \tag{14}$$

假设 *4*（一致子空间估计）。如果数据的形式为 $\widehat{U}_{(v)}$， $v = 1$， 2，则对于某些非随机量 $\mathrm{E}_{(v)on}$和 $\mathrm{E}_{2\,\infty,}^{(v)}$ $,n$, $v = 1$， 2 以及任意常数 $\tau > 0$，有

$$\mathbb{P}\left\{\left\|\widehat{U}^{(v)} - U^{(v)} W_U^{(v)}\right\| \le C_\tau\,\mathcal{E}_{0,n}^{(v)}\right\} \ge 1 - n^{-\tau}, \tag{15}$$

$$\mathbb{P}\left\{\left\|\widehat{U}^{(v)} - U^{(v)} W_U^{(v)}\right\|_{2,\infty} \le C_\tau\,\mathcal{E}_{2,\infty,n}^{(v)}\right\} \ge 1 - n^{-\tau}, \tag{16}$$

其中，常数 C_τ 仅取决于 τ。

假设*5*（ $\widehat{U}_{(v)}$的非相干性）。对于 $v = 1$或 $v = 2$，以及仅取决于 τ的常数 C_τ，有

$$\mathbb{P}\left\{\left\|\widehat{U}^{(v)}\right\|_{2,\infty} \le C_\tau \frac{\sqrt{K_v}}{\sqrt{n}}\right\} \ge 1 - n^{-\tau}. \tag{17}$$

上述假设相当常见。假设1允许簇的大小在常数因子范围内变化。如果各簇极不均衡，那么就很难区分错误的簇和真正很小的簇。假设 2指出，我们关注的是每个视图中的簇数固定且不随 n增长的情形。假设 3针对数据以估计的聚类矩阵形式给出的情况，此时估计得到的各个聚类分配以高概率保持一致。假设 4处理数据表现为奇异向量的估计矩阵的情形。它假定，

以高概率，差值的谱范数和二到无穷范数均有非随机上界。唯一看起来有些不同寻常的条件是假设5。不过，这一假设可以通过多种方式满足。容易看出，若 $\mathrm{E}_{2\,\infty}^{(v)}$„ $_n \le C\,K_v^{1/2}\,n^{-1/2}$，则式 (17) 成立。例如，在 k-均值模型 (9) 中，误差矩阵 $\Xi^{(v)}$ 的条目相互独立且服从次高斯分布时，后者便成立。比如，以下结论可直接由 Xie & Zhang（2025）的定理3.1推出。

引理1（矩阵的非相干性 $\widehat{U}_{(v)}$）。考虑模型(9)，其中 $G_{(v)}$ 由(10)给出。设$d_v \ge K_v$，且$\mathcal{M}^{(v)}$ 的最小非零特征值与最大非零特征值之比以某个常数为下界，即

$$\sigma_{K_v}(G^{(v)}) \Big/ \sigma_1(G^{(v)}) \ge C_\sigma .$$

假设矩阵 $\Xi^{(v)}$ 在(9) 中的元素是独立的次高斯或次指数随机变量，从而对任意 $t>0$ 及某个绝对常数 $C>0$，有

$$\mathbb{P}\left\{\sigma^{-1}|\Xi(i,j)| > t\right\} \le C\,\exp(-t^a),$$

其中 $a=2$ 对应于次高斯情形，$a=1$ 对应于次指数情形。于是，只要满足以下条件，假设5中的条件(17) 即成立

$$\sigma\sqrt{n}(\log n)^{1+\xi} \Big/ \sigma_{K_v}(G^{(v)}) = O(1)$$

对某个 $\xi>0$。

确保假设5成立的另一种方法是通过增加一个正则化步骤来修改算法1。将正则化算子$\mathrm{P}_\delta(U,K)$定义为

$$\mathcal{P}_\delta(U) = \mathrm{SVD}_K(U_*) \quad \text{where} \quad U_*(i,:) := U(i,:)\cdot \min\{\delta, \|U(i,:)\|\} \Big/ \|U(i,:)\|, \quad i\in[n].$$

在应用算法1之前，将矩阵 $\widehat{U}^{(v)}$ 之一替换为 $\widehat{U}_\delta^{(v)}$，

$$\widehat{U}_\delta^{(v)} = \mathcal{P}_{\delta_v}\left(\widehat{U}^{(v)}, K_v\right), \quad \delta_v = C_v\,K_v^{1/2}\,n^{-1/2}, \quad v=1,2,$$

其中，C_v 是一个足够大的常数。随后，在算法1的步骤1中使用公式计算矩阵 $\widehat{U}$

$$\widehat{U} = \mathrm{SVD}_K\left(\widehat{U}^{(1,2)}\right), \quad \widehat{U}^{(1,2)} = \left(\widehat{U}_{\delta_1}^{(1)} \circledR \widehat{U}^{(2)}\right) \text{ or } \widehat{U}^{(1,2)} = \left(\widehat{U}^{(1)} \circledR \widehat{U}_{\delta_2}^{(2)}\right)$$

而非（8）。在这种情况下，$\widehat{U}_{\delta_v}^{(v)}$ 将满足假设5，结合假设1，对于该视图，方程（16）在 $\mathrm{E}_{2\,\infty}^{(v)}{}_{,n} = n^{-1/2}\,K_v^{1/2}$ 时成立。此外，根据正则化的性质（例如，参见 Kim 等人（2018）或 Kivel¨a 等人（2014）），有

$$\left\|\widehat{U}_{\delta_v}^{(v)} - U^{(v)}W_{U,\delta}^{(v)}\right\| \le 2\sqrt{2}\,\left\|\widehat{U}^{(v)} - U^{(v)}W_U^{(v)}\right\|,$$

因此，对于正则化版本 $\widehat{U}_{(v)\delta}$ of $\widehat{U}_{(v)}$，方程（15）在相同的 $\mathrm{E}_{(v)}0_{,n}$ 下成立，但常数 C_τ 不同。

3.2。联合簇恢复的准确性

在本节中，我们评估算法1的精度。假设每个视图中的簇数K_v $v=1$，2，以及联合簇数 K已知。我们先讨论观测到 $\widehat{Z}_{(v)}$、 $v=1$，2的情况。

定理1（基于 $\widehat{Z}_{(v)}$的联合簇恢复精度）。设假设*1*、*2*、*3*成立，并且通过对 $\widehat{Z}_{(v)}$、$v = 1$、2应用算法*1*得到联合聚类分配$\hat{z}$。若 n足够大，则聚类$\hat{z}$的聚类错误数$\mathcal{E}_n$至多为单个视图 $\widehat{Z}_{(1)}$和 $\widehat{Z}_{(2)}$中的聚类错误数之和，即

$$\mathcal{E}_n \leq \mathcal{E}_{1,n} + \mathcal{E}_{2,n}. \tag{18}$$

因此，如果对于 $v = 1$，$_{(v)}$，2 的聚类结果是一致的，那么基于 $\widehat{Z}$ 的聚类也是一致的。如果聚类结果 $\widehat{Z}_{(v)}$、 $v = 1$、2 是完美的，那么基于 $\widehat{Z}$ 的聚类也同样完美。

现在，考虑数据以奇异向量的估计矩阵 $\widehat{U}^{(v)}$、 $v = 1,2$ 的形式给出的情况。

定理 2（基于 $\widehat{U}_{(v)}$ 的联合聚类的一致恢复）。 假设 *1*、*2*、*4* 和 *5* 成立，并且对于 $v = 1$，2，$\mathcal{E}_0^{(v)},_n = o(1)$ 作为 $n \to \infty$。那么，使用算法*1*对 $\widehat{U}_{(v)}$、 $v = 1$、2 应用算法进行聚类是一致的。

定理 3（完美聚类）。设定理 *2* 的条件成立，此外，对于$v = 1$，2

$$\sqrt{n/K_v}\,\mathcal{E}_{2,\infty,n}^{(v)} = o(1), \quad as\ n \to \infty. \tag{19}$$

那么，对于足够大的 n，使用算法*1* 应用于 $\widehat{U}$ $_{(v)}$ 、 $v = 1$，2进行聚类，以高概率实现完美聚类。

如果使用基于 $\widehat{U}$ $_{(v)}$ 的算法1 版本，则条件 E $_{(v)}$ 0，$_n = o(1)$ 对于各视图中的一致聚类 v 是必要的。为了在假设 1、2、4 和 5 下实现完美聚类，需要使 $\sqrt{n/K_v}\,\mathrm{E}_{2\,\infty}^{(v)}$，，$_n$ 由一个较小的绝对常数界定；在这一较弱条件下，定理仍然成立（尽管该常数难以明确给出）。

在我们的模拟中，我们发现，与 k-均值或高斯混合建模相比，采用完全链接的层次聚类（见算法 3）在联合聚类中表现出更稳定的性能，因此我们针对算法1 的一个版本给出了类似结果，在该版本中，我们用此算法替换（$1 + \epsilon$）-近似 k-均值。

^定理 4（基于层次聚类的完美聚类）。 设定理 *3* 的条件成立，并令 U为将算法*1* 应用于 $\widehat{U}_{(v)}$、 $v = 1$、2 的结果。则当 n足够大时，使用应用于 U 各行的算法 *3* 对 ^进行聚类，会在步骤 $t = n - K + 1$ 以高概率得到完美聚类。

3.3. 估计联合簇的数量

如前所述，KRAFTY 算法的优势之一在于，它能够更准确地估计联合簇的数量，尤其是在 $K > K_1 + K_2$ 时。以下陈述证实，分别在式 (6) 和式 (8) 中定义的矩阵 $\widehat{Z}_{(1,2)}$和 $\widehat{U}_{(1,2)}$的奇异值在第 K 个奇异值处呈现肘部。

算法 3。完全链接层次聚类

输入： $\widehat{U}(i,:) \in \mathbb{R}^K, i \in [n]$.**输出：** 聚类划分 $\mathrm{G}_{(t)}$ 以及合并高度 $h(t)$, 对应于 $t = 1,\ldots,n$。**步骤：**

1:初始化 $\mathrm{G}_1^{(0)} = \{\{1\}, \{2\}, \ldots\ldots, \{n\}\}$;
2: for $t = 2$ 到 n **执行**
找出使以下量最小的簇对（$G_{(t-1)i}$， $G_j^{(t-1)}$）：

$$d(G_i^{(t-1)}, G_j^{(t-1)}) = \max\{\|\widehat{U}(a,:) - \widehat{U}(b,:)\| : a \in G_i^{(t-1)}, b \in G_j^{(t-1)}\}$$

Merge $G_i^{(t-1)}$ and $G_j^{(t-1)}$ to form $G_{new}^{(t)} = G_i^{(t-1)} \cup G_j^{(t-1)}$;
Set $h(t) = \max\{\|\hat{U}(a,:) - \hat{U}(b,:)\| : a, b \in G_{new}^{(t)}\}$.;
Update $\mathcal{G}^{(t)} = \left(\mathcal{G}^{(t-1)} \setminus \{G_i^{(t-1)}, G_j^{(t-1)}\}\right) \cup \{G_{new}^{(t)}\}$;

定理5（基于 $\widehat{Z}_{(v)}$ 对联合簇数量的一致估计）。假设 *1*、*2* 和 *3* 成立，并且 $\mathrm{E}_{v,\,n}$ 定义如(14)所示。则

$$\sigma_K(\widehat{Z}^{(1,2)}) - \sigma_{K+1}(\widehat{Z}^{(1,2)}) \geq \frac{c_o\sqrt{n}}{\sqrt{K}}\left(1 - \frac{2\sqrt{2}}{c_o}\sqrt{\frac{K(\mathcal{E}_{1,n} + \mathcal{E}_{2,n})}{n}}\right), \tag{20}$$

$$\sigma_{K+j}(\widehat{Z}^{(1,2)}) - \sigma_{K+j+1}(\widehat{Z}^{(1,2)}) \leq \frac{c_o\sqrt{n}}{\sqrt{K}}\frac{2\sqrt{2}}{c_o}\sqrt{\frac{K(\mathcal{E}_{1,n} + \mathcal{E}_{2,n})}{n}}, \quad j \geq 1. \tag{21}$$

由于 n^{-1}（$\mathrm{E}_{1,\,n} + \mathcal{E}_{2,\,n}$）$= o$（1）当 $n \to \infty$ 时，当 n 足够大时，奇异值会在 K 处呈现肘部。

定理6（基于 $\widehat{U}_{(v)}$ 的一致簇数估计）。令假设 *1*、*2*、*4* 和 *5* 成立，并令 $\mathrm{E}_{o}^{(v),\,n}$ 在（15）中满足 $\mathrm{E}_{(v)\,o,\,n} = o(1)$ 当 $n \to \infty$，$v = 1,\ 2$ 时。则

$$\sigma_K(\widehat{U}^{(1,2)}) - \sigma_{K+1}(\widehat{U}^{(1,2)}) \geq \frac{c_D\sqrt{K}}{\sqrt{K_1K_2\,n}}\left[1 - \frac{2\,C_\tau\,K_1K_2}{c_D\,\sqrt{K}}(\mathcal{E}_{o,n}^{(1)} + \mathcal{E}_{o,n}^{(2)})\right], \tag{22}$$

$$\sigma_{K+j}(\widehat{U}^{(1,2)}) - \sigma_{K+j+1}(\widehat{U}^{(1,2)}) \leq \frac{c_D\sqrt{K}}{\sqrt{K_1K_2\,n}}\left[\frac{2\,C_\tau\,K_1K_2}{c_D\,\sqrt{K}}\,(\mathcal{E}_{o,n}^{(1)} + \mathcal{E}_{o,n}^{(2)})\right], \quad j \geq 1. \tag{23}$$

由于 $\mathcal{E}_{o}^{(1),\,n} + \mathcal{E}_{(2)\,o,\,n} = o$（1）当 $n \to \infty$ 时，当 n 足够大时，奇异值会在 K 处呈现肘部。

如果基于合并高度 $h(t)$ 的肘部来估计簇数，则对于采用完全链接的层次聚类情形，也有类似的结果。

定理 7（使用层次聚类一致估计联合簇的数量）。设定理 *4* 的条件成立，并考虑对 $\widehat{U}^{(1,2)}$ 的行应用算法 *3* 进行聚类。当 n 足够大时，以高概率，合并高度在步骤 $n - K + 2$ 处出现一个肘部，具体而言：对于任意固定的 $\eta \in (0,1/6)$，当 n 足够大时，以

高概率有

$$h(n-K+2)-h(n-K+1)\geq(1-4\eta)\frac{\sqrt{2K_1K_2}}{c_o^{(1)}c_o^{(2)}n},$$
$$h(k+1)-h(k)\leq 2\eta\frac{\sqrt{2K_1K_2}}{c_o^{(1)}c_o^{(2)}n},\quad k\leq n-K.$$

S展示这一现象的模拟结果见附录图 12 .

4. 模拟结果

4.1. 模拟设计

本节通过模拟评估 KRAFTY 在联合聚类和簇数估计中的实证性能，并在两种输入设置下将其与 MASE 进行比较：(1) $\widehat{U}_{(v)}$对每个视图分别进行奇异值分解得到的矩阵，以及 (2) $\widehat{Z}_{(v)}$通过对 $\widehat{U}_{(v)}$应用 k-means 得到的矩阵。

模拟数据由高斯混合生成，其特征由三个参数决定：维度（p）、簇内方差（σ^2）和成分数（K）。为获得联合簇标签，我们初始化一个 $K_1\times K_2$ 概率矩阵，其非零元素的数量等于真实联合簇数 K。确保每个联合簇中至少包含一个点后，其余点按照指定的概率向量从多项分布中抽取。对于每个数据集 $v\in 1,\ 2$，我们从标准多元正态分布 $\mathbf{N}(\mathbf{0},I_p)$ 中抽取 K_v 个不同的簇中心。所得簇均值记为$\boldsymbol{\mu}_k^{(v)}$ ， $k\in[K_v]$。

随后，生成每个数据集中的观测值，方法如下：

$$\mathbf{Y}_i^{(v)}\sim\mathcal{N}(\boldsymbol{\mu}^{(v)}_{\tau_v(i)},\sigma^2I_p),\quad i=1,\ldots,n.$$

为评估不同条件下的性能，我们设计了三种情景，每种情景改变一个参数，同时保持其他参数不变。

生成数据集后，我们为 KRAFTY 和 MASE 准备输入。两种方法都需要两个输入：数据矩阵和联合簇数，获取方式如下。对于矩阵输入，使用 $\widehat{U}_{(v)}$ 个矩阵时，我们对每个数据集进行奇异值分解，并提取前 K_1 个和 K_2 个左奇异向量，构成 $\widehat{U}_{(v)}$。这两个矩阵用作 KRAFTY 或 MASE 的输入。使用 $\widehat{Z}_{(v)}$ 个矩阵时，我们对 $\widehat{U}_{(v)}$ 应用 k 均值聚类，使用 $k=K_1$ 或 $k=K_2$，并将得到的 $\widehat{Z}_{(v)}$ 个矩阵作为输入。为突出比较 KRAFTY 和 MASE，我们假设各个数据集中的簇数 K_1 和 K_2 已知。在实际应用中，这些数值也需要估计，但可以采用标准方法进行估计（Zhu & Ghodsi，2006年）。对于另一个输入，即联合簇数，当 K 未知时，我们采用 Zhu & Ghodsi（2006年）的自动碎石图选择方法，将第二个肘部选作 K，因为第一个肘部通常对应的是单一的主导维度，而非真实的聚类结构。采用自动碎石图选择方法主要有两个原因：（1）它为估计 K 提供客观且可复现的标准；（2）它便于在本研究所需的大量实验重复中实现全自动估计。在 K 已知的条件下，我们将 K 作为输入。

对单个视图应用 k-均值聚类以获得 $\widehat{Z}$ 能提供稳定结果，但在应用 KRAFTY 或 MASE 后，对 U 的行进行聚类时一致性较差，因此我们在所有实验中均采用完全链接层次聚类进行最终聚类。

对于评估指标，我们使用 Hubert 与 Arabie（1985）提出的调整兰德指数（ARI）评估联合聚类性能，该指数用于比较聚类结果与真实标签。我们使用绝对误差评估联合簇数量的估计准确性，$\widehat{\ }$其定义为估计的 K 与真实 K之间的差值。结果取 100 次重复的平均值。在所有图中，点表示均值，阴影区域表示经验 95% 置信区间（均值 ± 1.96SE），KRAFTY 在图例中简称为“KR”。

4.2. 变化的联合簇数量

在本模拟研究中，我们将参数固定为 $p = 20, n = 1000, \sigma^2 = 0.1, K_1 = K_2 = 4$，同时将真实联合簇数量从4变化到16，对应于从$\max(K_1, K_2)$到 $K_1 \times K_2$的范围。

图3比较了KRAFTY与MASE在已知和未知 K条件下，分别使用$\widehat{U}^{(v)}$或 $\widehat{Z}^{(v)}$矩阵时的联合聚类性能。在左图中，当以$\widehat{Z}^{(v)}$矩阵作为输入时，三条曲线相互重叠，并在 $K > K_1 + K_2$时位于未知 K 条件下的MASE曲线之上，表明KRAFTY在该情形下的优势来自对 K的正确估计。在右图中，当以 $\widehat{U}^{(v)}$ 矩阵作为输入时，在已知和未知 K 条件下，当 $K > K_1 + K_2$ 时KRAFTY均优于MASE。

图 4 进一步比较了未知 K 设置下，四种方法与输入类型组合（KRAFTY-$\widehat{Z}^{(v)}$、KRAFTY- $\widehat{U}^{(v)}$、MASE-$\widehat{Z}^{(v)}$ 和 MASE- $\widehat{U}^{(v)}$）的联合聚类性能。结果展示了 K 的三个代表性取值（4、9 和 15），分别对应 K小于、接近和大于 $K_1 + K_2$的情况；此外，还展示了六种噪声水平下的结果，其中 $\sigma^2 \in \{0.01, 0.05, 0.1, 0.25, 0.5, 1\}$。结果表明，在不同噪声水平和输入类型下，当 $K > K_1 + K_2$时，KRAFTY 的性能更好。此外，当 K较小时，使用 $\widehat{Z}^{(v)}$ 矩阵的 KRAFTY 与使用 $\widehat{Z}^{(v)}$ 的 MASE 性能相当，且优于另外两种组合；而随着 K增大，使用 $\widehat{U}^{(v)}$矩阵的 KRAFTY 性能更好。

在附录图 19 和图 20 中，我们比较了已知 K和未知 K条件下 KRAFTY 与 MASE 的表现，并比较了在各种增长率 $K = f(K_1, K_2)$ 下将 $\widehat{U}^{(v)}$与 $\widehat{Z}^{(v)}$矩阵作为这些方法输入时的效果。

接下来，我们在图 5 中考察了两种方法对联合簇数量的估计性能。热图颜色表示 KRAFTY 与 MASE 的平均绝对误差之差；暖色表示 KRAFTY 的表现更好，冷色表示 MASE 的表现更好。两个面板（分别以 $\widehat{U}^{(v)}$或 $\widehat{Z}^{(v)}$作为输入）均证实，当 K较大时，KRAFTY 的表现优于 MASE。还值得注意的是，使用 $\widehat{Z}^{(v)}$矩阵时（右侧面板），即使 K较小，只要噪声水平较低（此处为 $\sigma^2 < 0.5$），KRAFTY 的表现也与 MASE 相当。这一发现与我们先前的结论一致：当 K较小且噪声水平较低时，采用 $\widehat{Z}^{(v)}$矩阵的 KRAFTY 更可取。附录图 16 通过直接比较使用 $\widehat{Z}^{(v)}$矩阵和 $\widehat{U}^{(v)}$矩阵，进一步支持了这一结论。附录图 21 和图 22 将分析扩展到更多 K_1, K_2取值以及增长率 $K = f(K_1, K_2)$，为我们的发现提供了进一步证据。

4.3. 变化的信噪比

在本次模拟中，我们研究维度和噪声水平如何影响 KRAFTY 的性能。我们固定 $K_1 = K_2 = 4$，设置 $K \in \{6，12\}$，并改变噪声水平 $\sigma^2 \in \{0。01、0。25、1、5\}$ 和维度 $p \in \{2、4、6、8、10、20、30、40、50\}$。图 6 展示了在 K 未知时对 KRAFTY 和 MASE 进行联合簇聚类的结果。对于这两种方法，随着维度增加，性能提升到某一点后趋于稳定。这一模式符合直觉：随着维度增加，由于噪声数据的重叠减少，簇变得更加易于分离；

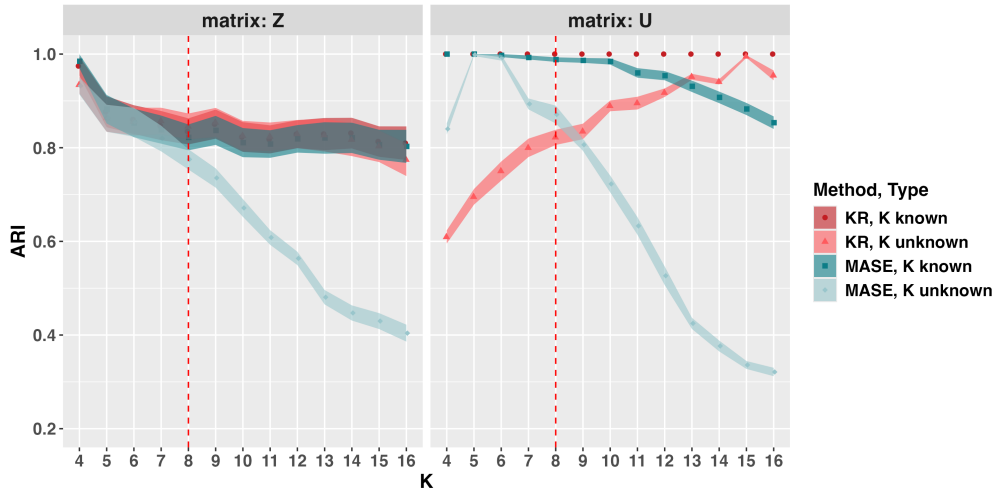


图3：在已知和未知 K条件下，使用 $\widehat{Z}_{(v)}$或 $\widehat{U}^{(v)}$矩阵评估KRAFTY与MASE的联合聚类性能。调整兰德指数衡量估计联合簇与真实联合簇之间的一致性。 $K_1 = K_2 = 4, \sigma^2 = 0.1, p = 20.$

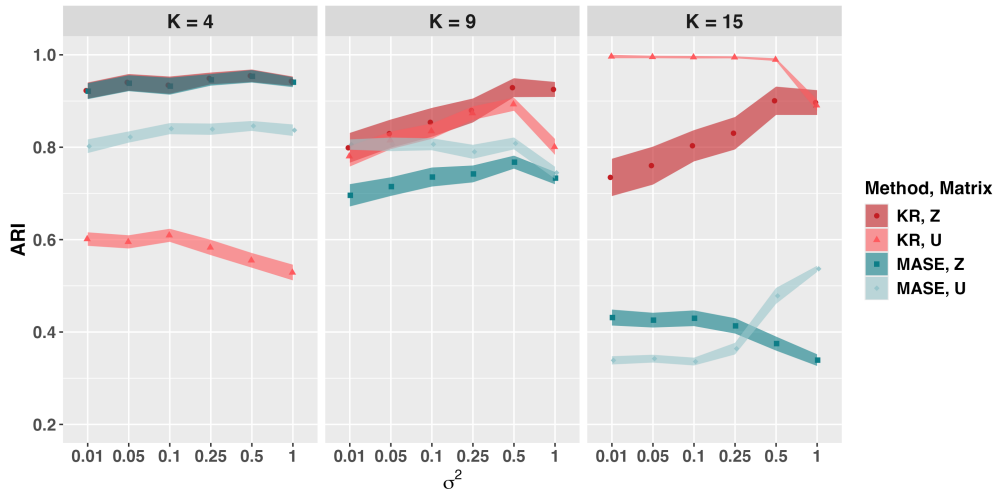


图4：在**未知** K 设定下，使用KR与MASE对 $\widehat{Z}_{(v)}$和 $\widehat{U}^{(v)}$矩阵进行的联合聚类性能。真实联合簇数量 K 在各实验中保持为4、9和15不变，同时 σ^2 发生变化。调整兰德指数将估计联合簇与真实值进行比较。此处， $K_1 = K_2 = 4, p = 20.$

但当它们已经足够清晰可分时，进一步增加维度只能带来很小的额外收益。

在大规模 K 设置下，我们看到 KRAFTY 在多个维度和噪声水平下均优于 MASE。在小规模 K 设置下，MASE 与 KRAFTY（ $\widehat{Z}_{(v)}$ ）的表现相当出色，只有在噪声极低时，才更倾向于使用以 $\widehat{U}^{(v)}$ 矩阵为输入的 MASE。对于低维数据，在不同的 K 取值和噪声水平下，我们看到 KRAFTY

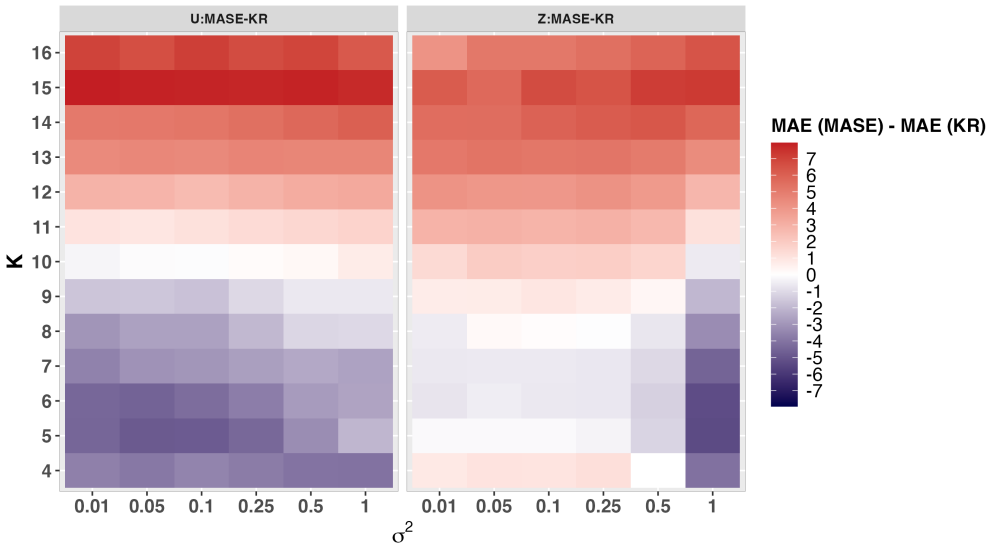


图 5：在改变真实簇数量 K和噪声水平 σ^2时，比较 KR 和 MASE 在 $\widehat{Z}(v)$ 与 $\widehat{U}^{(v)}$矩阵上的联合簇数估计性能。热图颜色表示使用 MASE 与 KR 估计 K时平均绝对误差的差值（对 100 次重复取平均）。对于 $\widehat{U}^{(v)}$矩阵，最大可信区间的宽度为 1.06（= 2(1.96) SE）；对于 $\widehat{Z}(v)$ 矩阵，该值为 1.15。冷色 $\rightarrow$表示 MASE 表现更好；暖色 $\rightarrow$ 表示 KR 表现更好。 $K_1 = K_2 = 4, p = 20$。

使用 $\widehat{Z}^{(v)}$ 矩阵是一个极佳的选择，这或许有些出人意料，因为我们原本预计 MASE 在 K 取值较小时会有出色的性能。关于这一比较的另一种视角见附录图 17 （第一行第四个面板），进一步印证了这一结果：在低维和高维 K 设置下，暖色的右下三角区域均表明 KRAFTY 在高噪声条件下表现更好。

比较 $U^{(v)}$矩阵作为输入时，我们发现，在噪声和维度设置相同的情况下，当 K较小时，使用 $\widehat{Z}$（v）矩阵的 KRAFTY 始终优于使用 $\widehat{U}$（v）矩阵的 KRAFTY；而当 K较大时，在信噪比足够高的情况下，使用 $\widehat{U}$（v）矩阵的 KRAFTY 表现更好，这一点在较高维度下尤为明显。图7进一步印证了这一规律，它直接比较了 $\widehat{U}^{(v)}$与 $\widehat{Z}$（v）在所有维度和噪声水平组合下的结果。当 $K = 6$时，所有热图均呈冷色，表明使用 $\widehat{Z}$（v）矩阵的 KRAFTY 表现更好。相反，当 $K = 12$时，热图左上区域（对应于相对于噪声而言维度较高的情况）呈暖色，而下方区域呈冷色，这表明在高维、低噪声条件下，使用 $\widehat{U}^{(v)}$的 KRAFTY 表现更好。关于 $\widehat{U}^{(v)}$ 与 $\widehat{Z}$（v）矩阵的更全面比较，请参见附录图 18。

图8展示了联合簇数估计的结果。图中左侧四个面板比较了每种方法的 $\widehat{U}^{(v)}$和 $\widehat{Z}_{(v)}$，右侧四个面板则比较了每种输入类型下的 KRAFTY 和 MASE。当 K较小时，MASE 的表现更好；但在相对高维的设置下使用 $\widehat{Z}^{(v)}$时例外，此时 KRAFTY 与 MASE 之间的差异较小。相反，当 K较大时，KRAFTY 明显优于 MASE。比较 KRAFTY 的 $\widehat{U}^{(v)}$和 $\widehat{Z}_{(v)}$，我们发现，当维度足够高时，$\widehat{Z}_{(v)}$对联合簇数的估计更准确；而当维度较低且噪声水平较高时，$\widehat{U}^{(v)}$表现更好。在这些设置下，$\widehat{Z}_{(v)}$往往会过度

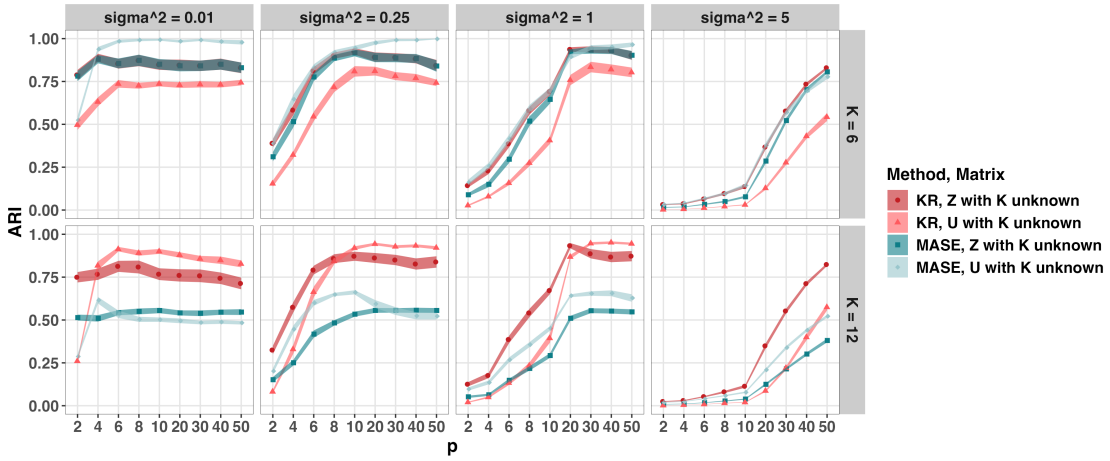


图 6：当真实联合簇数（K）未知且维度 p 变化时，KR 与 MASE 在 $\widehat{Z}^{(v)}$ 和 $\widehat{U}^{(v)}$ 矩阵上的联合聚类性能比较。通过将估计的联合簇与真实值进行比较来计算 ARI。我们设定 $K_1 = K_2 = 4$，即每个单图中的簇数。

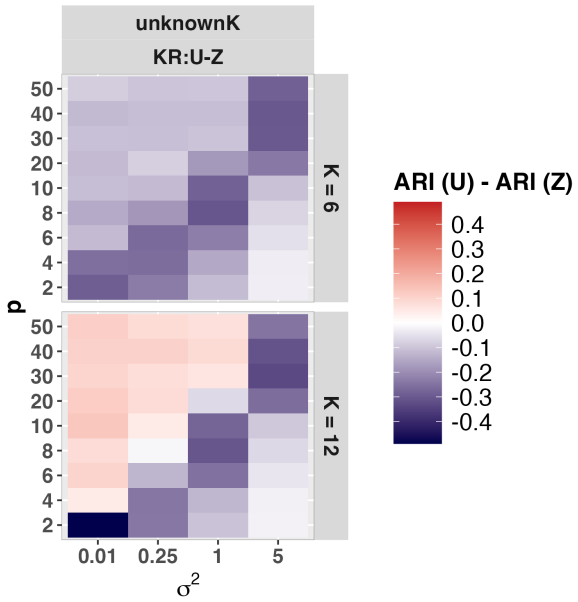


图 7：KRAFTY 中 $\widehat{Z}^{(v)}$ 与 $\widehat{U}^{(v)}$ 的联合聚类性能比较。热图颜色表示联合簇数 K 未知时平均 ARI（对 100 次重复取平均）的差异。**冷色 → 表示 Z 更优；暖色 → 表示 U 更优**。 $K_1 = K_2 = 4$。

估计真实簇数：这会导致联合簇数估计的误差较大，但不会使以调整兰德指数衡量的聚类性能明显下降。

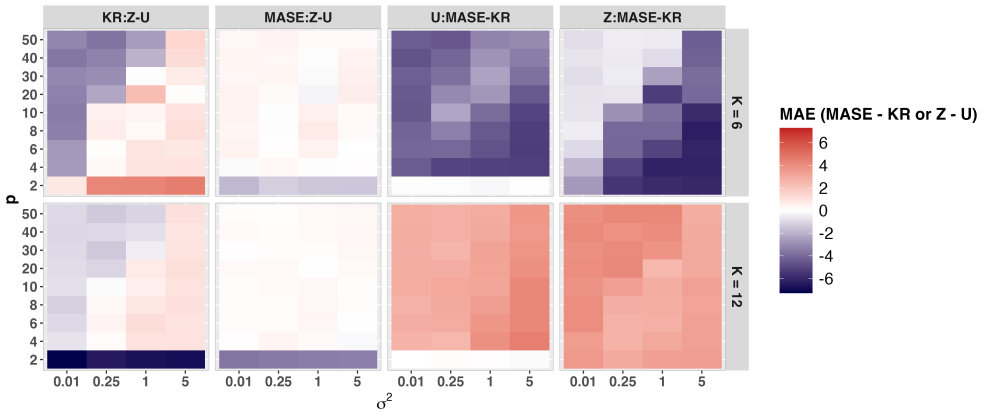


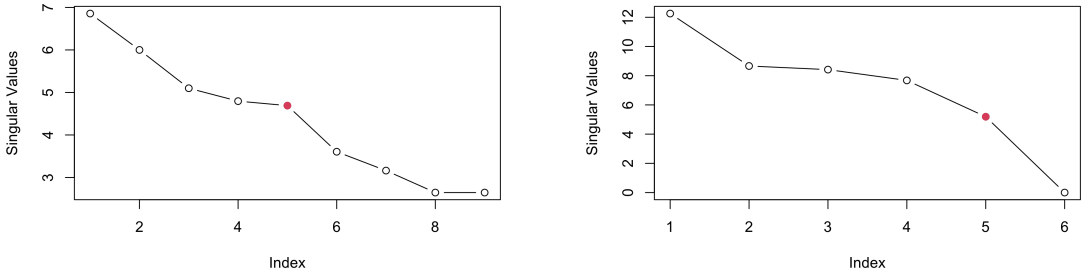
图 8：随着维度 p 和噪声水平 σ^2 变化，KR 与 MASE 在 Z 矩阵和 U矩阵上的联合簇数估计性能比较。热图左侧两列显示使用 Z 与 U 估计 K 时平均绝对误差（对 100 次重复取平均）的差异。**冷色 → Z 更优；暖色 → U 更优**。 右侧两列显示 KR 与 MASE 方法之间平均绝对误差的差异。**冷色 → MASE 更优；暖色** → KR 更优。</style}此处，$K_1 = K_2 = 4$。

5. 全球贸易数据分析

我们使用粮食及农业组织（FAO）提供的食品和农产品贸易数据集，该数据集也被 Jing 等人（2021）使用，并可在 http://www.fao.org 公开获取。该数据包含一系列有向加权贸易网络，每种产品对应一个网络，其中节点代表国家，边从出口国指向进口国，边的权重为该产品的美元贸易额。

沿用先前工作 Domenico 等人 (2014)；Jing 等人 (2021)；Agterberg 和 Zhang (2024)，我们选取 2010 的数据。我们还将 2023（可获得的最近年份）的数据纳入分析。本分析旨在：(i) 揭示能够刻画各国在 2010 出口与进口行为中联合隶属关系的联合聚类结构；(ii) 揭示可跨年份泛化的、出口商特异与进口商特异的联合聚类结果。我们聚焦一种产品（生鸡肉），并分析其贸易网络的两个快照（2010 和 2023），二者共享由 191 个国家组成的相同顶点集。对于每个年份 $v \in \{2010, 2023\}$，令 $A^{(v)}$ 表示观测到的贸易矩阵（条目为该产品从行国家到列国家的美元出口额）。我们计算其奇异值分解 $A^{(v)} = U(v)\,\Sigma^{(v)} V(v)^\top$，并保留 2010 的前三个维度以及 2023 的前两个维度，这些维度由 Zhu & Ghodsi (2006年) 选出。然后，我们通过将$U(v)\,(\Sigma^{(v)})^{1/2}$和 $V(v)\,(\Sigma^{(v)})^{1/2}$按行归一化至单位长度，估计各国作为出口商和进口商的潜在位置。归一化后的出口商与进口商嵌入分别记为 $\widehat{U}(v)$ 和 $\widehat{V}(v)$。随后，我们对 $\widehat{U}(v)$ 和 $\widehat{V}(v)$ 应用带有 $k = 3$和 $k = 2$的 k-means，从而得到 $\widehat{Z}(v)$: $v \in \{(2010, \text{imp}), (2010, \text{exp}), (2023, \text{imp}), (2023, \text{exp})\}$ 矩阵作为我们的输入。我们在本节详述单年度分析的发现，并将双年度分析留至附录，因为它们遵循相同的分析流程。

在本研究中，我们将 KRAFTY 和 MASE 方法应用于 $\widehat{Z}^{(v)}$: $v \in \{(2010,\text{imp}), (2010, \text{exp})\}$ 矩阵，这些矩阵分别是 2010 年进口商和出口商的簇分配，用于恢复联合群组。图9展示了碎石图，其中红点标出了 KRAFTY（图9(a)）和 MASE（图9(b)）估计的联合簇数量。碎石



(a) KRAFTY (b) MASE

图9：2010 年的特征值碎石图；红点表示选定维度 .

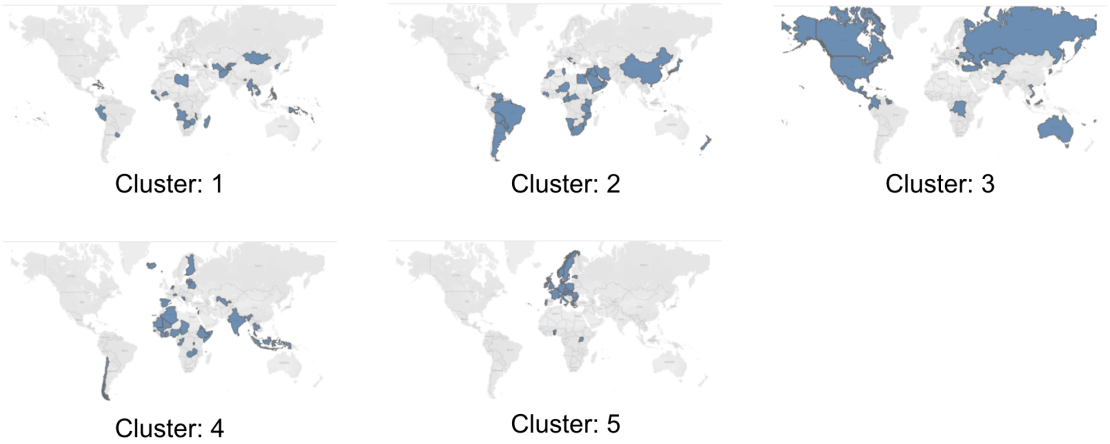
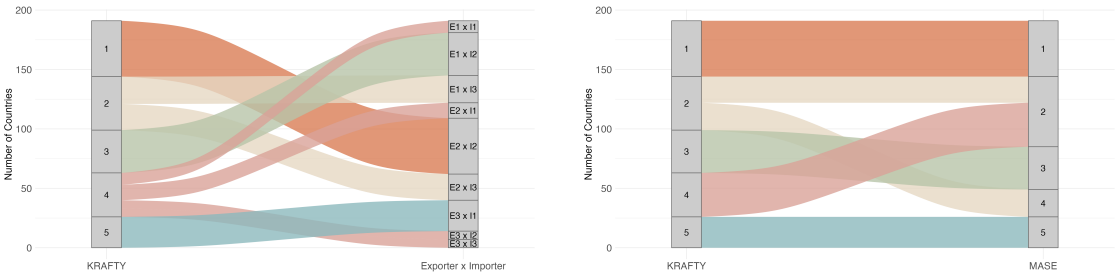


图10：按 2010 年 KRAFTY联合聚类着色的世界地图；蓝色表示聚类归属。



(a) 联合（KRAFTY）与单源 (b) KRAFTY 与 MASE

图11：比较2010年聚类分配的桑基图。(a) 对比 KRAFTY 的联合簇与单视图聚类的组合；相同颜色表示同一组实体，区块高度表示每个簇中的国家数量。(b) 对比 KRAFTY 的联合簇与 MASE 得出的簇。

KRAFTY 的碎石图呈现出明显的肘部，使选择更为容易，这与我们此前的发现一致。附录中针对两年期出口商和进口商分析的碎石图也显示出相同的优势。两种方法均采用层次聚类，并比较其聚类分配。调整兰德指数为 0.79，表明两种方法的结果高度一致。这与我们的模拟结果相符：当联合簇的数量小于单源聚类数量之和时，使用 $\widehat{Z}^{(v)}$ 输入的 KRAFTY 与 MASE 表现相近。

接下来，我们根据联合簇为世界地图着色，以观察它们揭示的模式；结果见图 10。从着色后的地图可以看出，簇主要呈现区域性分布：欧洲、北美洲，以及一个亚洲—非洲—南美洲群组。这可能反映了生鸡肉极易腐坏的特性，因此更适合进行距离较短的区域贸易。不过也有例外：簇 1 和簇 4 似乎分别混合了亚洲—非洲和欧洲—亚洲—非洲的部分子集。这些例外对应的群组，其贸易量明显低于主要区域簇，详见附录图 30 中的贸易网络热图。因此，联合簇既捕捉了区域模式，也反映了各国的贸易价值行为。此外，我们使用桑基图（图 11）展示联合簇如何由单源聚类形成，并比较 KRAFTY 的联合分配与 MASE 的结果略有不同之处。

6. 讨论

在这项工作中，我们提出了 KRAFTY，这是一种基于转置 Khatri-Rao 积的新型联合聚类框架。KRAFTY 不需要对单视图数据的来源作出任何假设，并直接处理每个视图的聚类分配矩阵（$\widehat{Z}_{(v)}$）或低维谱嵌入（$\widehat{U}_{(v)}$）。这种设计使 KRAFTY 成为一种灵活且适应性强的工具，可用于恢复多个视图间的联合聚类结构。KRAFTY 的另一项优势是，它提供了足够的维度，使联合簇能够占据正交子空间；而拼接的秩则不可能大于各单视图簇数之和。我们的主要结果表明，当数据点数量 n 较大时，联合聚类误差的上界为各单视图聚类误差之和。如果每个单视图聚类都具有一致性，则联合聚类也具有一致性；如果单视图聚类完美，则联合聚类也完美。这是一项理想的性质，因为研究人员可以独立并行地使用现有的高质量单视图聚类方法，从而获得准确的联合聚类结果。此外，我们证明，当 n 较大时，KRAFTY 生成的联合嵌入会在真实联合簇数 K处呈现清晰的肘部，从而能够利用碎石图或自动维数选择方法可靠地估计 K。在模拟中，我们将 KRAFTY 与广泛使用的 MASE 方法进行了系统比较，并评估了它在不同输入类型下的性能：$\widehat{Z}_{(v)}$和 $\widehat{U}_{(v)}$。KRAFTY 在聚类分配准确率和联合簇数估计方面均表现出色，尤其是在真实联合簇数 K 超过单视图簇数之和时。即使 K 小于该总和，KRAFTY 仍能取得相当的性能。在世界贸易数据分析中，我们应用 KRAFTY，根据各国作为出口国和进口国所扮演的角色来恢复国家的联合簇。

最后，我们讨论未来工作的几个方向。尽管 KRAFTY 假设单视图聚类结果保持一致，但真实世界情景通常涉及含噪、不精确或存在偏差的聚类结果。探索能否利用联合结构，通过 KRAFTY 对此类含噪输入进行去噪或优化，将会很有意义。此外，许多真实世界数据集中不断演化的聚类结构表明，KRAFTY 在动态社区发现方面具有有趣的应用前景。

7. 补充材料

补充材料包含定理 1–7 的证明、额外的模拟、KRAFTY 的可视化示例，以及真实数据应用的另一个示例。模拟和真实数据分析的代码可在 GitHub 获取：https://github.com/Ciel-Gao/ KRAFTY。

参考文献

Abbe, E.（2023）。社区检测与随机块模型。Abbe, E.、Fan, J. 和 Wang, K.（2022）。PCA与谱聚类的 l_p 理论。《统计年鉴》50，2359–2385。Agterberg, J.、Lubberts, Z. 和 Arroyo, J.（2025）。多层度校正随机块模型中的联合谱聚类。Agterberg, J. 和 Zhang, A.（2024）。通过 $l_{2,\infty}$ 张量扰动界估计高阶混合成员。Arroyo, J.、Athreya, A.、Cape, J.、Chen, G.、Priebe, C. E. 和 Vogelstein, J. T.（2020）。具有公共不变子空间的多个异质网络的推断。Domenico, M. D.、Nicosia, V.、Arenas, A. 和 Latora, V.（2014）。多层网络的结构可约性。《自然通讯》6。Fortunato, S. 和 Newman, M. E. J.（2022）。网络社区检测的20年。《物理评论》18，848–850。Gallagher, I.、Jones, A. 和 Rubin-Delanchy, P.（2021）。具有稳定性保证的动态网络谱嵌入。载于《神经信息处理系统进展》，M. Ranzato、A. Beygelzimer、Y. Dauphin、P. Liang 和 J. W. Vaughan 编，第34卷。Curran Associates, Inc。Han, Q.、Xu, K. S. 和 Airoldi, E. M.（2015）。动态和多层块模型的相合估计。Handcock, M. S.、Raftery, A. E. 和 Tantrum, J. M.（2007）。社交网络的基于模型的聚类。《英国皇家统计学会会刊A辑：社会统计学》170，301–354。Holland, P.、Laskey, K. B. 和 Leinhardt, S.（1983 5, 109 137 ）。随机块模型：初步研究。《社交网络》–。Hubert, L. 和 Arabie, P.（1985）。比较划分。《分类学杂志》2，193–218。Jaeger, A. 和 Banks, D.（2023）。聚类分析：现代统计综述。《WIREs计算统计学》15，e1597。Jin, D.、Yu, Z.、Jiao, P.、Pan, S.、He, D.、Wu, J.、Yu, P. S. 和 Zhang, W.（2023）。社区检测方法综述：从统计建模到深度学习。《IEEE知识与数据工程汇刊》35，1149–1170。Jin, J.（2015）。基于得分的快速社区检测。《统计年鉴》43。Jing, B.-Y.、Li, T.、Lyu, Z. 和 Xia, D.（2021）。通过正则化张量分解对混合多层网络进行社区检测。《统计年鉴》49，3181–3205。Kim, C.、Bandeira, A. S. 和 Goemans, M. X.（2018）。超图随机块模型：统计极限与半定规划方法。arXiv预印本 arXiv:1807.02884。Kivel¨a, M.、Arenas, A.、Barthelemy, M.、Gleeson, J. P.、Moreno, Y. 和 Porter, M. A.（2014）。多层网络。《复杂网络杂志》2，203–271。Ledesma, R. D.、Valero-Mora, P. 和 Macbeth, G.（2015）。碎石检验与因子数：一种动态图形方法。《西班牙心理学杂志》18，E11。Lei, J. 和 Lin, K. Z.（2023）。多层随机块模型中的偏差校正谱聚类。《美国统计学会杂志》118，2433–2445。Lei, J. 和 Rinaldo, A.（2015）。随机块模型中谱聚类的相合性。《统计年鉴》43。Lyzinski, V.、Sussman, D.、Tang, M.、Athreya, A. 和 Priebe, C.（2015）。通过邻接矩阵谱嵌入实现随机块模型图的完美聚类。L¨offler, M.、Zhang, A. Y. 和 Zhou, H. H.（2020）。高斯混合模型中谱聚类的最优性。MacDonald, P. W.、Levina, E. 和 Zhu, J.（2021）。具有共享结构的多路复用网络潜在空间模型。Milligan, G. W. 和 Cooper, M. C.（1987）。方法综述：聚类方法。《应用心理测量》11，329–354。Paul, S. 和 Chen, Y.（2018）。用于多层网络一致性社区检测的谱方法与矩阵分解方法。Pensky, M.（2024）。二到无穷范数下的Davis-Kahan定理及其在完美聚类中的应用。Qin, T. 和 Rohe, K.（2013）。度校正随机块模型下的正则化谱聚类。Rohe, K.、Chatterjee, S. 和 Yu, B.（2011）。谱聚类与高维随机块模型。《统计年鉴》39。

Sussman, D. L.、Tang, M.、Fishkind, D. E. 和 Priebe, C. E.（2012）。随机块模型图的一种相合邻接矩阵谱嵌入。Toh, H. 和 Horimoto, K.（2002）。通过聚类分析与图形高斯建模相结合的方法推断遗传网络。《生物信息学》**18**，287–297。Vall`es-Catal`a, T.、Massucci, F. A.、Guimer `a, R. 和 Sales-Pardo, M.（2016）。多层随机块模型揭示复杂网络的多层结构。《物理评论*X*》**6**，011036。von Luxburg, U.（2007）。谱聚类教程。Xie, F. 和 Zhang, Y.（2025）。低秩随机矩阵的高阶逐元素特征向量分析：偏差校正、埃奇沃思展开与自助法。《统计年鉴》**53**，1667–1696。Yang, Y. 和 Wang, H.（2018）。多视图聚类：综述。《大数据挖掘与分析》**1**，83–107。Zhu, M. 和 Ghodsi, A.（2006年）。利用轮廓似然从碎石图中自动选择维度。《计算统计与数据分析》**51**，918–930。

A. 论文中各结论的证明。

A.1. *Proof of Theorem 1.*

为简便起见，记

$$X = Z^{(1)} \circledR Z^{(2)} = Z H^\top, \quad \widehat{X} = \widehat{Z}^{(1)} \circledR \widehat{Z}^{(2)}, \quad \Xi = \widehat{X} - X, \tag{A1}$$

其中，矩阵 H 在式 (3) 中定义。令

$$X = UDH^\top, \quad \widehat{X} = \widehat{U}\,\widehat{D}\,\widehat{V}^\top + \widehat{U}^\perp\,\widehat{D}^\perp\,(\widehat{V}^\perp)^\top \tag{A2}$$

为 X 和 $\widehat{X}$ 的奇异值分解。这里，$D = \mathrm{diag}\,(\sqrt{n_1}, \cdots, \sqrt{n_K})$。

回想一下，J_v 是被 $\widehat{Z}^{(v)}$ 错误聚类的元素集合，$\mathrm{E}_{v,\,n}$ 是其数量，$v = 1,\ 2$。于是，J_v^c 是正确聚类元素的集合。记作 $J = J_1 \cup J_2$、$J^c = J_1^c \cap J_2^c$，并注意到，若 $i \in J^c$，则 $\Xi(i,\ :) = 0$。这里，由于 $|J_v^c| = n - \mathrm{E}_{v,\,n}$，有 $|J^c| \geq n - (\mathrm{E1}_{,\,n} + \mathrm{E2}_{,\,n})$。为了评估算法 1 的聚类精度，注意到，根据假设 1

$$\sigma_K(X) = \min_{k\in[K]} \sqrt{n_k} \geq \frac{c_o\sqrt{n}}{\sqrt{K}}. \tag{A3}$$

由于 $\left\|\widehat{X}(i,\ :) - X(i,\ :)\right\|^2 = 0$ 若 $i \in J^c$，且 $\left\|\widehat{X}(i,\ :) - X(i,\ :)\right\|^2 = 2$ 若 $i \in J$，我们有

$$\|\widehat{X} - X\|_F^2 = 2\,|J| \leq 2(\mathcal{E}_{1,n} + \mathcal{E}_{2,n}). \tag{A4}$$

因此，$\|\Xi\| = \|\widehat{X} - X\| \leq \sqrt{2(\mathrm{E1}_{,\,n} + \mathrm{E2}_{,\,n})}$。个体聚类的相合性（假设 3）意味着 $n^{-1}\mathrm{E}_{v,\,n} \to 0$。然后，对于足够大的 n，有

$$\sigma_K(\widehat{X}) \geq \sigma_K(X) - \|\widehat{X} - X\| \geq \frac{c_o\sqrt{n}}{\sqrt{K}}\left(1 - \frac{\sqrt{2K}}{c_o}\frac{\sqrt{\mathcal{E}_{1n} + \mathcal{E}_{2n}}}{\sqrt{n}}\right) \geq \frac{c_o\sqrt{n}}{2\sqrt{K}}. \tag{A5}$$

注意，由假设 3，当 $n \to \infty$ 时，

$$\epsilon_{o,n} := \frac{\sqrt{\mathcal{E}_{1n} + \mathcal{E}_{2n}}}{\sqrt{n}} = o(1). \tag{A6}$$

然后，由 Wedin定理以及假设 2 和假设 3，当 $n \to \infty$ 时，

$$\max\left(\|\widehat{U} - U\,W_U\|_F^2, \|\widehat{V} - H\,W_V\|_F^2\right) \leq \frac{2\,\|\Xi\|_F^2}{\sigma_K^2(X)} \leq \frac{4K\,(\mathcal{E}_{1,n} + \mathcal{E}_{2,n})}{c_o^2\,n} \asymp \epsilon_{o,n}^2 = o(1). \tag{A7}$$

这里，W_U 和 W_V 分别是使 U 与 U、V 与 HW_V 之间匹配程度最高的旋转。具体而言，若 $U^\top\widehat{U}$ 的奇异值分解为 $U^\top\widehat{U} = W_1 D_W W_2^\top$，则 $W_U = W_1 W_2^\top$。

为了根据 X 评估聚类错误的数量，我们应用 Abbe等人（2022）的引理D1），现为方便读者起见将其列于此处。

引理 A1. **(Abbe等人（2022）的引理D1)**. 设矩阵 $B \in \mathbb{R}^{r\times m}$ 的行是 $B(k,:)$，$k \in [K]$，该矩阵为真实均值矩阵，且 $z:\ [n] \to [K]$ 为真实聚类函数。对于数据矩阵 $X \in \mathbb{R}^{n\times m}$、任意矩阵 $\widetilde{B} \in \mathbb{R}^{K\times m}$ 和任意聚类函数 $\tilde{z}:\ [n] \to [K]$，定义

$$L\left(\widetilde{B}, \tilde{z}\right) = \sum_{i=1}^{n}\left\|\widehat{X}(i,:) - \widetilde{B}(\tilde{z}(i),:)\right\|^2. \tag{A8}$$

设 $\widehat{B} \in \mathbb{R}^{K\times m}$ 和 $\hat{z}[n] \to [K]$ 是$(1+\epsilon)$–近似 k 均值问题的解，即

$$L\left(\widehat{B}, \hat{z}\right) \leq (1+\epsilon)\min_{\widetilde{B},\tilde{z}} L\left(\widetilde{B}, \tilde{z}\right).$$

设 $s = \min_{i \neq j} \|B(i,:) - B(j,:)\|$，$n_{\min}$ 为最小簇大小。若存在某个 $\delta \in (0, s/2)$ 使得

$$L(B,z) = \sum_{i=1}^{n} \left\| \widehat{X}(i,:) - B(z(i),:) \right\|^2 \leq \frac{\delta^2 n_{\min}}{K(1+\sqrt{1+\epsilon})^2}, \tag{A9}$$

则存在一个置换 $\phi:\ [K] \to [K]$ 使得

$$\{i:\ \hat{z}(i) \neq \phi(z(i))\} \subseteq \left\{ i:\ \|\widehat{X}(i,:) - B(z(i),:)\| \geq s/2 - \delta \right\}, \tag{A10}$$

$$\#\{i:\ \hat{z}(i) \neq \phi(z(i))\} \leq (s/2-\delta)^{-2} L(B,z). \tag{A11}$$

我们应用引理 A1，其中 $s = \min_{i \neq j} \|U(i,:) - U(j,:)\| \geq (2/\max(n_k))^{1/2}$, $X = U W_U$, $B(z(i),:) = X(i,:)$ 以及 $L(B,z) \leq \left\| \widehat{U} - U W_U \right\|_F^2$。因此，

$$\delta^2 \leq \frac{\|\widehat{U} - U W_U\|_F^2 K(1+\sqrt{1+\epsilon})^2}{\min(n_k)},$$

从而，根据 (A7)，我们得到

$$\frac{s}{2} \geq \frac{\sqrt{K}}{C_o \sqrt{2n}}, \quad \delta \leq \frac{2K(1+\sqrt{1+\epsilon})}{c_o^{3/2}} \frac{\sqrt{\mathcal{E}_{1,n} + \mathcal{E}_{2,n}}\sqrt{K}}{n}. \tag{A12}$$

根据引理 A1，当 $\|e_i^\top (\widehat{U} - U W_U)\| < s/2 - \delta$ 时，行 i 正确聚类，其中

$$\frac{s}{2} - \delta \geq \frac{\sqrt{K}}{C_o\sqrt{2n}} \left(1 - \frac{2\sqrt{2} C_o K(1+\sqrt{1+\epsilon})}{c_o^{3/2}} \frac{\sqrt{\mathcal{E}_{1,n} + \mathcal{E}_{2,n}}}{\sqrt{n}} \right),$$

且括号中的第二项由于假设 3 为 $o(1)$ $n \to \infty$。因此，行 i 的正确聚类由下式保证

$$\|e_i^\top (\widehat{U} - U W_U)\| = o\left(\sqrt{K}/\sqrt{n} \right). \tag{A13}$$

现在，我们考虑满足 $i \in J^c$ 的行，并证明这些行都被正确聚类。对于 $i \in J^c$, $e_i^\top \Xi = \Xi(i,:) = 0$。

由 (A1) 可得

$$R_i := \|e_i^\top (\widehat{U} - U W_U)\| \leq R_{i1} + R_{i2} + R_{i3} + R_{i4} + R_{i5} \tag{A14}$$

$\widehat{\ }$满足 $\Xi = X - \widehat{X}$ 且

$$\begin{aligned} &R_{i1} = \|e_i \Xi H W_V \widehat{D}^{-1}\| = 0, \quad R_{i2} = \|e_i^\top \Xi (\widehat{V} - H W_V)\widehat{D}^{-1}\| = 0, \\ &R_{i3} = \|e_i^\top U U^\top \Xi H W_V \widehat{D}^{-1}\|, \quad R_{i4} = \|e_i^\top U U^\top \Xi (\widehat{V} - H W_V)\widehat{D}^{-1}\|, \\ &R_{i5} = \|e_i^\top U (U^\top \widehat{U} - W_U)\|. \end{aligned} \tag{A15}$$

注意，当 n 足够大时，由 (A5) 可得

$$\|\widehat{D}^{-1}\| = \sigma_{\min}^{-1}(\widehat{X}) \leq C\sqrt{\frac{K}{n}}.$$

因此，利用 (A4)、(A6) 和 (A7)，当 $n \to \infty$ 时，可得

$$R_{i3} \leq \|U\|_{2,\infty} \|\Xi\| \|\widehat{D}^{-1}\| \leq C(K/\sqrt{n})\epsilon_{o,n} = o(\sqrt{K}/\sqrt{n}), \tag{A16}$$

$$R_{i4} \leq \|U\|_{2,\infty} \|\Xi\| \|\widehat{D}^{-1}\| \|\widehat{V} - H W_V\| \leq C(K/\sqrt{n})\epsilon_{o,n}^2 = o(\sqrt{K}/\sqrt{n}), \tag{A17}$$

$$R_{i5} \leq \|e_i^\top U (U^\top \widehat{U} - W_U)\| \leq C(\sqrt{K}/\sqrt{n})\epsilon_{o,n}^2 = o(\sqrt{K}/\sqrt{n}). \tag{A18}$$

结合 (A14)–(A18) 可知 (A13) 成立，因此保证当 n 足够大时， J^c 中的所有元素都能正确聚类。于是 (18) 成立，定理1得证。

A.2. *Proof of Theorems 2 and 3* .

不失一般性，假设假设5在 $v = 2$ 下成立。为了基于 U 评估聚类错误的 $\widehat{}$ 数量，我们对 $s = \min_{z(i)\neq z(j)}\|U(i,:) - U(j,:)\|$ 应用引理 A1，
$X = U W_U$, $B(z(i),:) = X(i,:)$ 和 $L(B,z) \le \left\|\widehat{U} - U W_U\right\|_F^2$。因此，

$$s \ge \frac{\sqrt{2}}{\sqrt{\max(n_k)}} \ge \frac{\sqrt{2K}}{C_o\sqrt{n}}, \quad \delta \le \frac{\|\widehat{U} - U W_U\|_F \times \sqrt{K}\,(1+\sqrt{1+\epsilon})}{\sqrt{\min(n_k)}}. \tag{A19}$$

回顾一下， U =奇异值分解$\widehat{U}^{1} \bigcirc \widehat{U}^{2}$ ^{7} $\bigcirc$ ^{K} $\bigcirc$ $=$ $(U$ $_{(1,2)}$) 以及 $\widehat{U}$ =奇异值分解 ($\widehat{U}_{(1)}$ ⊛ $\widehat{U}_{(2)}$)，其中 $U_{(1,2)} = U_{(1)}$ ⊛ $U_{(2)}$，而 $\widehat{U}_{(1,2)}$ ⊛ ()。还需回顾，根据式 () ，

$$U^{(1,2)} = U D V^\top, \quad V \equiv H, \quad D = \left(D_Z \widetilde{D}_Z^{-1}\right)^{1/2}. \tag{A20}$$

这里， $D_Z = Z^\top Z$，因此，对于 $k_1 \in [K_1]$、 $k_2 \in [K_2]$，使用双指标可以写成

$$D_Z(k,k) = n_k, \quad \widetilde{D}_Z(k,k) = n_{k_1}^{(1)} n_{k_2}^{(2)} I(H((k_1,k_2),k) = 1), \quad k \in [K].$$

因此，

$$D(k,k) = \frac{\sqrt{n_k}}{\sqrt{n_{k_1}^{(1)} n_{k_2}^{(2)}}} I\left(H((k_1,k_2),k) = 1\right). \tag{A21}$$

根据假设1，

$$\frac{c_D\sqrt{K}}{\sqrt{K_1K_2}\sqrt{n}} \le D(k,k) \le \frac{C_D\sqrt{K}}{\sqrt{K_1K_2}\sqrt{n}}, \tag{A22}$$

其中 $c_D = c_o/(C_{o}^{(1}\ C_o^{(2))})$ 和 $C_D = C_o/(c_o^{(1)} c_o^{(2)})$。后者与（A20）一起给出

$$\sigma_{\min}(U^{(1,2)}) \ge \frac{c_D\sqrt{K}}{\sqrt{K_1K_2}\sqrt{n}}. \tag{A23}$$

将 $\widehat{U}^{(1,2)}$ 的奇异值分解写为

$$\widehat{U}^{(1,2)} = \widehat{U}\,\widehat{D}\,\widehat{V}^\top + \widehat{U}^\perp\,\widehat{D}^\perp\,(\widehat{V}^\perp)^\top.$$

记 $\Xi^{(v)} = \widehat{U}^{(v)} - U_{(v)}W_{(v)U}$、 $v = 1,\ 2$，以及 $\Xi = \widehat{U}_{(1,\ 2)} - U_{(1,\ 2)}W_{(1,\ 2)U}$，其中 $W_{(v)U}$在弗罗贝尼乌斯范数下提供最接近的匹配，而 $W_{(1,2}$ $_U = W_U^{(1)} \otimes W^{)(2)U}$。令 W_V 为一个旋转矩阵，它在 V 与 VW_V 之间的弗罗贝尼乌斯范数下提供$\widehat{}$最接近的匹配。

回顾一下，根据假设2和假设4，当 $n \to \infty$ 时，有

$$\left\|\Xi^{(v)}\right\|_F = \left\|\widehat{U}^{(v)} - U^{(v)}W_U^{(v)}\right\|_F \le C_\tau\sqrt{K_v}\,\mathcal{E}_{0,n}^{(v)} = o(1), \quad v = 1,2. \tag{A24}$$

因此，

$$\Xi = \Xi^{(1)} \circledR \widehat{U}^{(2)} + (U^{(1)}W_U^{(1)}) \circledR \Xi^{(2)} \tag{A25}$$

并且，根据 Khatri-Rao积的性质**P2** ，可得

$$\|\Xi\|_F \le \|\Xi^{(1)}\|_F \|\widehat{U}^{(2)}\|_{2,\infty} + \|\Xi^{(2)}\|_F \|U^{(1)}\|_{2,\infty}$$

那么，由假设1和假设5以及（A24）可知，其概率至少为 $1-3n^{-\tau}$,

$$\|\Xi\|_F \le \frac{\widetilde{C}_\tau \sqrt{K_1}\sqrt{K_2}}{\sqrt{n}} \left(\mathcal{E}_{0,n}^{(1)} + \mathcal{E}_{0,n}^{(2)}\right), \tag{A26}$$

~其中， C_τ 是一个通用的绝对常数，取决于 τ 以及假设1中的常数。容易看出，(A19) 和 (A26) 中的第二个不等式立即推出

$$\delta \le \frac{\widetilde{C}_\tau (1+\sqrt{1+\epsilon})\, K \sqrt{K_1\, K_2}\, (\mathcal{E}_{0,n}^{(1)} + \mathcal{E}_{0,n}^{(2)})}{c_o n} \tag{A27}$$

应用 (A19) 中的第一个不等式，可得

$$\frac{s}{2} - \delta \ge \frac{\sqrt{K}}{C_o\sqrt{2n}} \left(1 - \frac{\widetilde{C}_\tau\, C_o\, (1+\sqrt{1+\epsilon})\, \sqrt{2KK_1K_2}\, (\mathcal{E}_{0,n}^{(1)} + \mathcal{E}_{0,n}^{(2)})}{c_o\sqrt{n}}\right). \tag{A28}$$

若 n 足够大，根据定理的条件，可得

$$\frac{s}{2} - \delta \ge \frac{\sqrt{K}}{2\sqrt{2}\, C_o\, \sqrt{n}}.$$

因此，根据引理 A1，当满足以下条件时，可保证第 i 行被正确聚类

$$\|e_i^\top (\widehat{U} - UW_U)\| < \frac{\sqrt{K}}{2\sqrt{2}\, C_o\, \sqrt{n}}. \tag{A29}$$

为了利用 (A29)，我们推导 $\|e_i^\top(\widehat{U} - UW_U)\|$ 的一个上界。将量 $\Xi V W_V \widehat{D}^{-1}$ 展开为

$$\begin{aligned}
\Xi V W_V \widehat{D}^{-1} &= UU^\top \Xi V W_V \widehat{D}^{-1} - (I - UU^\top)\Xi(\widehat{V} - VW_V)\widehat{D}^{-1} + (I - UU^\top)\Xi\widehat{V}\widehat{D}^{-1} \\
&= UU^\top \Xi V W_V \widehat{D}^{-1} - (I - UU^\top)\Xi(\widehat{V} - VW_V)\widehat{D}^{-1} + (I - UU^\top)(\widehat{U} - UW_U),
\end{aligned}$$

利用 $(I - UU^\top)U^{(1,2)} = 0$ 这一事实，且 $\widehat{U}$ （1，2） $\hat{V}\hat{}D^{-1} = \widehat{U}$。于是，对于任意 $i \in [n]$，有

$$R_i = \|e_i^\top (\widehat{U} - UW_U)\| \le R_{i1} + R_{i2} + R_{i3} + R_{i4} + R_{i5}, \quad i \in [n]. \tag{A30}$$

其中

$$\begin{aligned}
R_{i1} &= \left\|e_i^\top \Xi V W_V \widehat{D}^{-1}\right\|, & R_{i2} &= \left\|e_i^\top \Xi(\widehat{V} - VW_V)\widehat{D}^{-1}\right\|, \\
R_{i3} &= \left\|e_i^\top UU^\top \Xi V W_V \widehat{D}^{-1}\right\|, & R_{i4} &= \left\|e_i^\top UU^\top \Xi(\widehat{V} - VW_V)\widehat{D}^{-1}\right\|, \\
R_{i5} &= \left\|e_i^\top U(U^\top\widehat{U} - W_U)\right\|.
\end{aligned}$$

注意到，根据 (A22),

$$\min_k(\widehat{D}(k,k)) \ge \min_k(D(k,k)) - \|\Xi\| \ge \frac{c_D\sqrt{K}}{\sqrt{K_1K_2}\sqrt{n}} \left(1 - \frac{\widetilde{C}_\tau K_1K_2(\mathcal{E}_{0,n}^{(1)} + \mathcal{E}_{0,n}^{(2)})}{c_D\sqrt{K}}\right). \tag{A31}$$

因此，如果 n 足够大，则以高概率有:

$$\min_k(\widehat{D}(k,k)) \ge \frac{c_D\sqrt{K}}{2\sqrt{K_1K_2}\sqrt{n}}, \tag{A32}$$

并且我们以高概率得到类似的上界

$$\max_k(\widehat{D}(k,k)) \le \frac{2\, C_D\sqrt{K}}{\sqrt{K_1K_2}\sqrt{n}}. \tag{A33}$$

由于 $e_i^\top \Xi = \Xi(i,:)$，(A32) 和 (A33) 得出

$$R_{i1} = \|e_i^\top \Xi V W_V \widehat{D}^{-1}\| \le \|\Xi(i,:)\| \frac{2\sqrt{K_1 K_2}\sqrt{n}}{c_D \sqrt{K}}. \tag{A34}$$

根据Davis-Kahan定理和（A23），我们得到

$$\|\widehat{V} - V W_V\|_F \le \frac{C\|\Xi\|_F}{\sigma_{\min}(U^{(1,2)})} \le \widetilde{C}_\tau \frac{K_1 K_2}{\sqrt{K}} (\mathcal{E}_{0,n}^{(1)} + \mathcal{E}_{0,n}^{(2)}). \tag{A35}$$

因此，根据（A34），当 $n \to \infty$ 时，以高概率有，

$$R_{i2} \le \frac{2\widetilde{C}_\tau}{c_D} \|\Xi(i,:)\| \frac{(K_1 K_2)^{3/2}\sqrt{n}}{K} (\mathcal{E}_{0,n}^{(1)} + \mathcal{E}_{0,n}^{(2)}) = o\left(\|\Xi(i,:)\| \frac{\sqrt{K_1 K_2}\sqrt{n}}{\sqrt{K}} \right). \tag{A36}$$

现在，假设1和（A32）表明，若 n 足够大，则以高概率有，

$$R_{i3} \le \|U\|_{2,\infty} \|\Xi\| \, \|\widehat{D}^{-1}\| \le 2 (c_D \, c_o)^{-1} \sqrt{K_1 K_2} \, \|\Xi\|.$$

将（A26）中的 $\|\Xi\|_F$ 代入，我们得到

$$R_{i3} \le \frac{2\widetilde{C}_\tau \, K_1 \, K_2}{c_D c_o \sqrt{n}} \left(\mathcal{E}_{0,n}^{(1)} + \mathcal{E}_{0,n}^{(2)} \right). \tag{A37}$$

类似地，当 $n \to \infty$ 时，以高概率有，

$$R_{i4} \le \|U\|_{2,\infty} \|\Xi\| \|\widehat{D}^{-1}\| \|\widehat{V} - V W_V\| = o(\|U\|_{2,\infty} \|\Xi\| \|\widehat{D}^{-1}\|), \tag{A38}$$

其上界已在（A37）中给出。最后，

$$R_{i5} \le C \|U\|_{2,\infty} \|\widehat{U} - U W_U\|^2. \tag{A39}$$

公式（A35）中的界同样适用于 $\widehat{} \, U - U W_U$，因此我们有

$$R_{i5} \le C \frac{\sqrt{K}}{c_o \sqrt{n}} \widetilde{C}_\tau^2 \frac{(K_1 K_2)^2}{K} (\mathcal{E}_{0,n}^{(1)} + \mathcal{E}_{1,n}^{(2)})^2 \tag{A40}$$

根据定理的假设，$\mathrm{E}_0^{(v)}{}_{,\,n} = o(1)$，因此该上界远小于公式（A37）中的上界。因此，当 n 足够大且 $i \in [n]$ 时，结合 (A30)–(A40) 可得

$$\|e_i^\top (\widehat{U} - U W_U)\| \le C_\tau \left(\|\Xi(i,:)\| \frac{\sqrt{K_1 K_2}\sqrt{n}}{\sqrt{K}} + \frac{\sqrt{K_1 K_2}}{\sqrt{n}} (\mathcal{E}_{0,n}^{(1)} + \mathcal{E}_{1,n}^{(2)}) \right), \tag{A41}$$

其中，(A41) 中的第二项渐近地小于 $\sqrt{K}/\sqrt{n}$。

因此，不等式 (A29) 是否成立完全取决于 $\|\Xi(i,:)\|$ 的行为。注意，由 (A25) 可得

$$\Xi(i,:) = \left(U^{(1)}(i,:) W_U^{(1)} \right) \circledKR \, \Xi^{(2)}(i,:) + \Xi^{(1)}(i,:) \circledKR \, \widehat{U}^{(2)}(i,:).$$

因此，由假设1和假设5，

$$\|\Xi(i,:)\| \le \frac{C_\tau \sqrt{K_2}}{\sqrt{n}} \|\Xi^{(1)}(i,:)\| + \frac{\sqrt{K_1}}{c_o^{(1)} \sqrt{n}} \|\Xi^{(2)}(i,:)\|. \tag{A42}$$

那么，当 n 足够大时，该结论以高概率成立，

$$\|\Xi(i,:)\| \le C_\tau \left(\frac{C_\tau \sqrt{K_2}}{\sqrt{n}} \mathcal{E}_{2,\infty,n}^{(1)} + \frac{\sqrt{K_1}}{c_o^{(1)} \sqrt{n}} \mathcal{E}_{2,\infty,n}^{(2)} \right).$$

因此，定理 3 的假设意味着，对于足够大的 n，以及任意绝对常数$C_{K,\text{ small}}$

$$\|\Xi(i,:)\| \leq \frac{\sqrt{K_1K_2}}{n} C_{K,\text{small}}, \tag{A43}$$

其中$C_{K,\text{ small}}$是一个稍后选定的小正数。由式 (A41) 和 (A43) 可得

$$\begin{aligned}\|e_i^\top(\widehat{U}-UW_U)\| &\leq C_\tau \left(\|\Xi(i,:)\| \frac{\sqrt{K_1K_2}\sqrt{n}}{\sqrt{K}} + \frac{\sqrt{K_1K_2}}{\sqrt{n}} (\mathcal{E}_{0,n}^{(1)}+\mathcal{E}_{1,n}^{(2)})\right) \\ &\leq C_\tau \frac{\sqrt{K}}{\sqrt{n}} \left[(]\frac{K_1K_2}{K} C_{K,\text{small}} + \frac{\sqrt{K_1K_2}}{\sqrt{K}} \left(\mathcal{E}_{0,n}^{(1)}+\mathcal{E}_{0,n}^{(2)}\right)\right] \\ &\leq \frac{\sqrt{K}}{2\sqrt{2}C_o\sqrt{n}},\end{aligned} \tag{A44}$$

当 n 足够大，且$C_{K,\text{ small}}$取值足够小时。这意味着条件 (A29) 对任意 $i \in [n]$均成立，因此定理 3 得证。

现在，我们完成定理 2 的证明。令

$$\mathcal{J}_n^{(v)} \equiv \mathcal{J}_n^{(v)}(C_{\text{small}}) = \left\{i \in [n] : \|\Xi^{(v)}(i,:)\| > \frac{C_{\text{small}}\sqrt{K_v}}{\sqrt{n}}\right\}, \quad v = 1,2, \tag{A45}$$

and

$$\mathcal{E}_n^{(v)} = |\mathcal{J}_n^{(v)}|, \quad v = 1,2. \tag{A46}$$

若 $i \in (\mathcal{J}_n^{(1)})^c \cap (\mathcal{J}_{(2)n})^c$，则由 (A42) 可知界 (A43) 成立，并且根据条件 (A29) 和不等式 (A44)，行 i 被正确聚类。记

$$\mathcal{J}_n = \mathcal{J}_n^{(1)} \cup \mathcal{J}_n^{(2)}, \quad \mathcal{J}_n^c = (\mathcal{J}_n^{(1)})^c \cap (\mathcal{J}_n^{(2)})^c$$

于是,

$$\mathcal{E}_n = |\mathcal{J}_n| \leq \mathcal{E}_n^{(1)} + \mathcal{E}_n^{(2)}. \tag{A47}$$

根据假设 4,

$$\mathbb{P}\left\{\|\Xi^{(v)}\|_F^2 \leq C_\tau K_v (\mathcal{E}_{0,n}^{(v)})^2\right\} \geq 1 - n^{-\tau}. \tag{A48}$$

根据定理 2 和假设 2 的条件，可得

$$K_v(\mathcal{E}_{on}^{(v)})^2 = o(1), \quad \text{as } n \to \infty. \tag{A49}$$

于是，由 (A45) 可得

$$\|\Xi^{(v)}\|_F^2 = \sum_{i=1}^n \|\Xi^{(v)}(i,:)\|^2 \geq \sum_{i \in \mathcal{J}_n^{(v)}} \|\Xi^{(v)}(i,:)\|^2 \geq \frac{K_v}{n} |\mathcal{J}_n^{(v)}| C_{\text{small}}^2. \tag{A50}$$

因此，由 (A48)、(A49) 和 (A50) 可知

$$\mathbb{P}\left\{\frac{\mathcal{E}_n^{(v)}}{n} \leq \frac{C_\tau}{C_{\text{small}}^2} (\mathcal{E}_{0,n}^{(v)})^2\right\} \geq 1 - n^{-\tau}.$$

因此，由于 (A47),

$$\mathbb{P}\left\{\frac{\mathcal{E}_n}{n} \leq \frac{C_\tau}{C_{\text{small}}^2} \left[(\mathcal{E}_{on}^{(1)})^2 + (\mathcal{E}_{on}^{(2)})^2\right]\right\} \geq 1 - 2n^{-\tau}. \tag{A51}$$

因此，当 $n \to \infty$ 时，$\mathrm{E}n/n \to 0$ 以高概率成立，并且聚类是一致的。

A.3. *Proof of Theorem 5*

如定理1的证明中所述，令 $\widehat{X} = \widehat{Z}(1,\ 2) = \widehat{Z}(1) \text{KR} \widehat{Z}(2)$和 $X = ZH^{\top}$。于是，由 (A3) 和 (A4) 可得

$$\sigma_K(X) \geq \frac{c_o\sqrt{n}}{\sqrt{K}}, \quad \|\widehat{X} - X\| \leq \sqrt{2(\mathcal{E}_{1,n} + \mathcal{E}_{2,n})}.$$

注意到，由 (A5) 可得

$$\sigma_K(\widehat{X}) \geq \frac{c_o\sqrt{n}}{\sqrt{K}}\left(1 - \frac{\sqrt{2K}}{c_o}\frac{\sqrt{\mathcal{E}_{1n} + \mathcal{E}_{2n}}}{\sqrt{n}}\right). \tag{A52}$$

此外，对于 $1 \leq j \leq K_1K_2 - K$,

$$\sigma_{K+j}(\widehat{X}) \leq \sigma_{K+j}(X) + \|\widehat{X} - X\| \leq \frac{c_o\sqrt{n}}{\sqrt{K}}\frac{\sqrt{2K}}{c_o}\frac{\sqrt{\mathcal{E}_{1n} + \mathcal{E}_{2n}}}{\sqrt{n}}. \tag{A53}$$

于是，(20) 和 (21) 直接由 (A52) 和 (A53) 得出。如果聚类是一致的，且假设 1、2、3 成立，则 $K = O(1)$， $n^{-1}K(\mathrm{E}_{1n} + \mathrm{E}_{2n}) = o(1)$，因此该定理成立。

A.4. *Proof of Theorem 6*

如定理 2 和定理 3 的证明中所示，设

$$X = U^{(1,2)}W_U^{(1,2)} = (U^{(1)} \text{KR} U^{(2)})W_U^{(1,2)} = UDH^{\top}W_U^{(1,2)}, \quad \widehat{X} = \widehat{U}^{(1,2)} = \widehat{U}^{(1)} \text{KR} \widehat{U}^{(2)}$$

和 $\Xi = X - X$。回顾一下，由 (A23) 可得,

$$\sigma_K(X) \geq \frac{c_D\sqrt{K}}{\sqrt{K_1K_2\,n}}, \quad \sigma_{K+1}(X) = 0.$$

此外，由 (A31) 可知，在假设 1、2、4 和 5 下，以高概率

$$\sigma_K(\widehat{X}) \geq \frac{c_D\sqrt{K}}{\sqrt{K_1K_2}\sqrt{n}}\left(1 - \frac{\widetilde{C}_\tau K_1K_2(\mathcal{E}_{0,n}^{(1)} + \mathcal{E}_{0,n}^{(2)})}{c_D\sqrt{K}}\right).$$

此外，由 (A26) 可知

$$\|\Xi\|_F \leq \frac{\widetilde{C}_\tau\sqrt{K_1}\sqrt{K_2}}{\sqrt{n}}\left(\mathcal{E}_{0,n}^{(1)} + \mathcal{E}_{0,n}^{(2)}\right).$$

于是，根据 $\sigma_{K+j}(X) \leq \|\Xi\|$ for $j \geq 1$这一事实，定理得证。

A.5. 层次聚类结果的证明

本节证明，在定理 3 的条件下，对 U =奇异值分解$_K(\widehat{U}(1) \text{KR} \widehat{U}(2))$ 的各行应用完全链接层次聚类，可以得到完美聚类，并且合并高度在第 $n - K + 2$ 步呈现肘部。算法 3 展示了完全链接层次聚类过程。

定理 4 的证明： 我们先介绍证明思路。在采用完全链接的层次聚类中，初始时有 n 个簇，每个簇都只包含一个点；每次合并后，簇数都会减少 1。假设属于每个真实簇的点集半径小于真实簇之间的间隔。那么，当点属于同一个真实簇时，我们预期合并高度较小；只有在所有可能的簇内合并都完成后，才会有一次合并将属于不同真实簇的点合在一起。这将在第 $n - K + 1$ 步发生，因为此时会有 $n -$ （$n - K + 1$）个 $= K - 1$ 簇，因此必有一个簇包含来自多个真实簇的点；我们也会看到这一步的合并高度出现跃升。现在来证明这些结论。

令 $\mathsf{G}^{(t)}$ 表示经过 t 次合并后得到的划分，并令 $G_{(t)i}$ 表示该划分中的第 i 个簇。合并两个簇时，有两种情况。首先，如果两个簇中的所有点 $G_{(t-1)i}, G_j^{(t-1)}$ 都属于同一个真实簇，那么对于任意 $a \in G_i^{(t-1)}$ 和 $b \in G_{(t-1)j}$，都有 $z(a) = z(b)$。因此，$e_a^\top U W_U = e_b^\top U W_U$，所以

$$\|(e_a^\top - e_b^\top)\widehat{U}\| \le \|e_a^\top(\widehat{U} - UW_U)\| + \|e_b^\top(\widehat{U} - UW_U)\| \le 2\|\widehat{U} - UW_U\|_{2,\infty} \tag{A54}$$

$$d(G_i^{(t-1)}, G_j^{(t-1)}) = \max\{\|\widehat{U}(a,:) - \widehat{U}(b,:)\| : a \in G_i^{(t-1)},\, b \in G_j^{(t-1)}\} \le 2\|\widehat{U} - UW_U\|_{2,\infty} \tag{A55}$$

其次，如果 $G_i^{(t-1)} \cup G_{(t-1)j}$ 中的点至少属于两个不同的真实簇，那么在这个并集中存在某些索引 a 和 b，使得 $z(a) \neq z(b)$。我们有

$$\begin{aligned}\|(e_a^\top - e_b^\top)\widehat{U}\| &\ge \|(e_a^\top - e_b^\top)UW_U\| - \|(e_a^\top - e_b^\top)(\widehat{U} - UW_U)\| \\ &\ge \|(e_a^\top - e_b^\top)UW_U\| - 2\|\widehat{U} - UW_U\|_{2,\infty}\end{aligned} \tag{A56}$$

令 $s = \min_{z(i)\neq z(j)}\|U(i,:) - U(j,:)\|$，我们有

$$\|(e_a^\top - e_b^\top)\widehat{U}\| \ge s - 2\|\widehat{U} - UW_U\|_{2,\infty} \tag{A57}$$

公式 (A19) 表明 $s \ge \frac{\sqrt{2K}}{C_o\sqrt{n}}$。另一方面，如果固定 $\eta \in (0, 1/6)$，那么在必要时于公式 (A43) 中选取更小的常数 $C_{K,\text{small}}$，便可由公式 (A44) 得到以下上界；当 n 足够大时，该上界以高概率成立：

$$\|\widehat{U} - UW_U\|_{2,\infty} \le \eta\frac{\sqrt{2K}}{C_o\sqrt{n}} \le \eta s. \tag{A58}$$

这意味着所有初始合并均发生在簇内，其合并高度 $h(t) \le 2\|U - UW_U\|_{2,\infty}$。在算法的第 t 阶段，有 $n - t + 1$ 个簇。因此，在第 $n - K + 1$ 步，恰有 K 个不同的簇，并且

$$h(n - K + 1) \le 2\|\widehat{U} - UW_U\|_{2,\infty}. \tag{A59}$$

因此，$G^{(n-K+1)}$ 给出了精确的簇标签。定理 4 的证明至此完成。

当我们可以从算法1获得 U 时，可以看出，对于后续步骤 $t \ge n - K + 2$，任何一次合并都必然涉及来自不同真实簇的点，因此得到

$$h(t) \ge s - 2\|\widehat{U} - UW_U\|_{2,\infty}. \tag{A60}$$

因此，在步骤 $n - K + 2$ 的合并高度上会出现一个间隙，将簇内合并与簇间合并区分开来，从而确保精确恢复。结合公式 (A58)、(A59) 和 (A60)，可得

$$h(n - K + 2) - h(n - K + 1) \ge (1 - 2\eta)\frac{\sqrt{2K}}{C_o\sqrt{n}}$$

$$h(k + 1) - h(k) \le 2\eta\frac{\sqrt{2K}}{C_o\sqrt{n}}$$

对于 $k \le n - K$。因此，对 U 应用算法 3 时，合并高度中会出现一个肘部，但这需要预先了解 K，才能选择 U 的维度。为了解决这一问题，我们考虑直接对 $\widehat{U}^{(1,2)}$ 应用算法 3。由此得到定理 7，现证明如下。

定理 7 的证明： 令 $s = \min_{z(i)\neq z(j)} \|U^{(1,2)}(i,:) - U(1,2)(j,:)\|$。根据假设1，我们有

$$s = \min_{z(i)\neq z(j)} \left(\frac{1}{n^{(1)}_{z_1(i)} n^{(2)}_{z_2(i)}} + \frac{1}{n^{(1)}_{z_1(j)} n^{(2)}_{z_2(j)}} \right)^{1/2} \geq \frac{\sqrt{2K_1K_2}}{c_o^{(1)} c_o^{(2)} n}. \tag{A61}$$

回顾 $\Xi = \widehat{U}^{(1,2)} - U(1,2)W(1,2)U$，并且对于较大的 n，公式 (A43) 给出了高概率界

$$\|\Xi\|_{2,\infty} \leq \frac{\sqrt{K_1K_2}}{n} C_{K,\text{small}}.$$

令 $\eta \in (0,1/6)$。然后在必要时选取更小的 $C_{K,\text{small}}$，我们可以确保对于较大的 n，以高概率，

$$\|\Xi\|_{2,\infty} \leq \eta \frac{\sqrt{2K_1K_2}}{c_o^{(1)} c_o^{(2)} n}.$$

综合以上，前 $n-K+1$ 次合并是簇内的，其合并高度

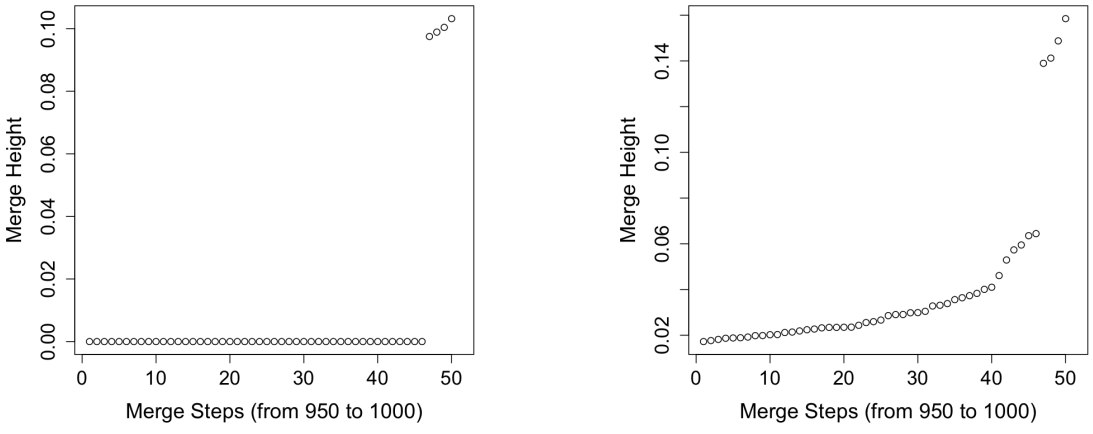
$$h(t) \leq 2\|\Xi\|_{2,\infty} \leq 2\eta \frac{\sqrt{2K_1K_2}}{c_o^{(1)} c_o^{(2)} n},$$

并且所有后续合并都涉及来自不同真实簇的点，其高度必须

$$h(t) \geq s - 2\|\Xi\|_{2,\infty} \geq (1-2\eta) \frac{\sqrt{2K_1K_2}}{c_o^{(1)} c_o^{(2)} n}.$$

这些不等式共同证明了定理 7。

图 12 展示了这样一种想法：在步骤 $n-K+2$ 处，合并高度出现了一个间隙。在两个子图（$n = 1000$ 和 $K = 5$）中，都可以观察到步骤 996 和 997 之间存在明显的跳跃。



(a) 输入：$SVD_K(\widehat{Z}(1,2))$。

（b）输入：SVD_K（$\widehat{U}$（1，2））。

图 12：$n = 1000$、$K = 5$、$K_1 = K_2 = 4$、$m = 30$ 和 $\sigma^2 = 0$。25 的完全链接层次聚类合并高度图。

A.6. *Khatri-Rao*积的性质及其证明

P1. $(AC) \odot (BD) = (A \odot B)(C \otimes D)$.

P2. $\|A \odot B\|_F \leq$最小$(\|A\|_F \|B\|_{2,\infty}, \|A\|_{2,\infty} \|B\|_F)$。

证明：

$$\begin{aligned}
\|A \odot B\|_F^2 &= \sum_{i=1}^{n} \|A(i,:) \otimes B(i,:)\|_2^2 \\
&= \sum_{i=1}^{n} [A(i,:) \otimes B(i,:)][A(i,:) \otimes B(i,:)]^\top \\
&= \sum_{i=1}^{n} \left[A(i,:)A(i,:)^\top\right] \otimes \left[B(i,:)B(i,:)^\top\right] \\
&\leq \max_j \|A(j,:)\|^2 \sum_{i=1}^{n} \|B(i,:)\|^2; \text{ or} \\
&\leq \max_j \|B(j,:)\|^2 \sum_{i=1}^{n} \|A(i,:)\|^2.
\end{aligned}$$

P3. $\|A \odot B\|_{2,\infty} \leq \|A\|_{2,\infty} \|B\|_{2,\infty}$.

证明：

$$\begin{aligned}
\|A \odot B\|_{2,\infty}^2 &= \max_i \|A(i,:) \otimes B(i,:)\|_2^2 \\
&= \max_i \|A(i,:)\|^2 \cdot \|B(i,:)\|^2 \\
&\leq \max_i \|A(i,:)\|^2 \cdot \max_v \|B(v,:)\|^2.
\end{aligned}$$

P4. $(A \odot B)^\top = A^\top * B^\top$，其中 $A^\top * B^\top$ 为Khatri-Rao（按列）积。

P5. $(A \odot B)(A^\top * B^\top) = (AA^\top) \circ (BB^\top)$，其中（$X_1 \circ X_2$）是 X_1与 X_2的 Hadamard积。

P6. $\|A \odot B\| \leq \|A \otimes B\| = \|A\|\|B\|$.

P7. 若 $Z_{(v)} \in \{0,1\}^{n \times K_v}$, $v = 1,2$, 是聚类矩阵，则 $Z_{(1)} \odot Z_{(2)} \in \{0,1\}^{n \times K_1 K_2}$和$\left(Z_{(1)} \odot Z_{(2)}\right)^\top \left(Z_{(1)} \odot Z_{(2)}\right) = \mathrm{D}_Z^{(1,2)}$是一个对角矩阵，其元素为

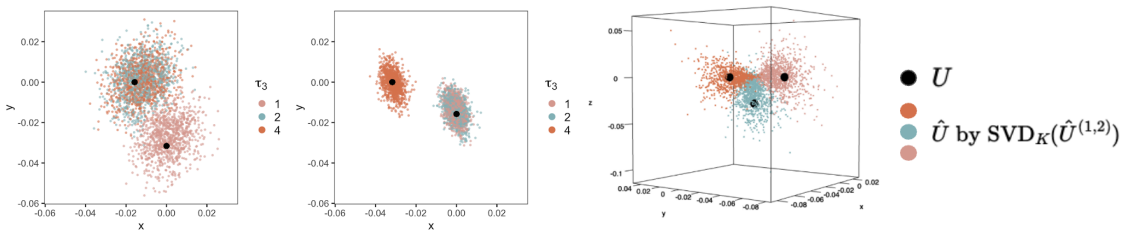
$$\mathcal{D}_{Z,(k_1,k_2)}^{(1,2)} = \sum_{i=1}^{n} I\left(i \in \mathcal{G}_{k_1}^{(1)}\right) \cdot I\left(i \in \mathcal{G}_{k_2}^{(2)}\right) = n_{(k_1,k_2)}^{(1,2)},$$

其中 $\mathrm{G}_{k_v}^{(v)}, v = 1,2$ 是按照聚类 $v \in \{1,2\}$ 对 k_v进行分组所得的组，$k_v \in [K_v]$。

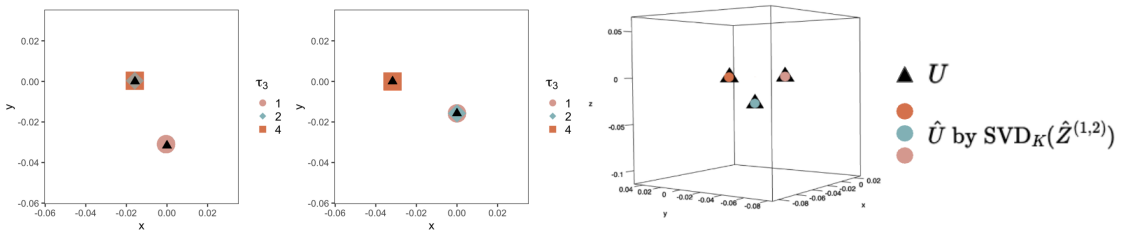
证明： 考虑 $\mathrm{L} = \left(Z_{(1)} \odot Z_{(2)}\right)^\top \left(Z_{(1)} \odot Z_{(2)}\right) \in \mathbb{R}^{K_1 K_2 \times K_1 K_2}$。然后，利用双指标，可以写作

$$\begin{aligned}
\mathcal{L}((k_1,k_2),(\tilde{k}_1,\tilde{k}_2)) &= \sum_{i=1}^{n} \left(Z^{(1)} \odot Z^{(2)}\right)(i,(k_1,k_2)) \left(Z^{(1)} \odot Z^{(2)}\right)(i,(\tilde{k}_1,\tilde{k}_2)) \\
&= \sum_{i=1}^{n} I\left(i \in \mathcal{G}_{k_1}^{(1)}\right) I\left(i \in \mathcal{G}_{k_2}^{(2)}\right) I\left(i \in \mathcal{G}_{\tilde{k}_1}^{(1)}\right) I\left(i \in \mathcal{G}_{\tilde{k}_2}^{(2)}\right) \\
&= \sum_{i=1}^{n} I\left(i \in \mathcal{G}_{k_1}^{(1)}\right) I\left(l \in \mathcal{G}_{k_2}^{(2)}\right) I(k_1 = \tilde{k}_1, k_2 = \tilde{k}_2) \\
&= n_{(k_1,k_2)}^{(1,2)}
\end{aligned}$$

B. $\widehat{U}^{(v)}$与 $\widehat{Z}$（v）输入的可视化比较



(a) $\widehat{U}^{(1)}$ (b) $\widehat{U}^{(2)}$ (c) 通过 SVD_K（$\widehat{U}$(1,2)）获得的 $\widehat{U}$



(d) $\widehat{Z}^{(1)}$ (e) $\widehat{Z}^{(2)}$ (f) $\widehat{U}$通过 $SVD_K(\widehat{Z}(1,2))$获得

图13：使用维度为 $p = 3$ 且包含$K = 3$个联合簇的模拟高斯混合展示 KRAFTY。(a) 视图1中的 $\widehat{U}(1)$；(b) 视图2中的 $\widehat{U}(2)$；(c) 对 $\widehat{U}(1)$ KR $\widehat{U}(2)$进行奇异值分解得到的左奇异向量。底行：(d) 视图1中的 $\widehat{Z}(1)$，按奇异值平方根缩放，并通过对 $\widehat{U}(1)$进行 k-均值聚类得到聚类结果；(e) 使用 $\widehat{U}(2)$得到视图2中 $\widehat{Z}(2)$的类似结果；(f) 对 $\widehat{Z}(1)$ KR $\widehat{Z}(2)$进行奇异值分解得到的左奇异向量。点的颜色表示真实类别归属；黑色标记表示真实簇中心。可以看出，尽管存在三个联合簇，但由于重叠，每个单视图中只有两个簇能够区分。图中展示了这样一种思想：$\widehat{U}^{(v)}$和 $\widehat{Z}(v)$都以相同的真实中心为目标；然而，$\widehat{Z}(v)$会将点压缩到质心附近，而 $\widehat{U}^{(v)}$保留了方差“云团”。无论使用$\widehat{U}(v)$输入还是 $\widehat{Z}(v)$输入，得到的 $\widehat{U}$ 都会估计联合视图的同一个 U。联合空间将所有簇分离到正交维度中。

C. 其他模拟结果

本节展示了三种设置下的扩展模拟结果：改变联合簇的数量、信噪比以及单视图簇的数量。

C.1. 联合簇数量变化

图 14 比较了不同噪声水平 σ^2（x 轴）下，三种联合簇大小 $K \in 4, 9, 15$ 时 MASE 和 KR 的表现。两个视图中的簇数均为 4（$K_1 = K_2 = 4$），观测值的维度为 $p = 20$，且在整个小节中保持不变。顶行使用 $\widehat{Z}(v)$作为输入；底行使用 $\widehat{U}(v)$。首先，考察以 $\widehat{Z}(v)$为输入的结果：当 $K = 4 < K_1 + K_2$ 时，无论 K 是否已知，MASE 和 KR 的表现都相近。当 $K > K_1 + K_2$（即 $K = 9, 15$）时，已知 K 的 MASE 与已知或未知 K 的 KR 表现相近，且都优于未知 K 的 MASE。这支持了 KR 的优势源自正确估计 K 这一发现。

对于 U 输入（底行），当 $K < K_1 + K_2$时，若未知 K，MASE 的表现优于 KR；若已知 K，两种方法的表现相近。当 $K > K_1 + K_2$时，MASE 和 KR 的表现

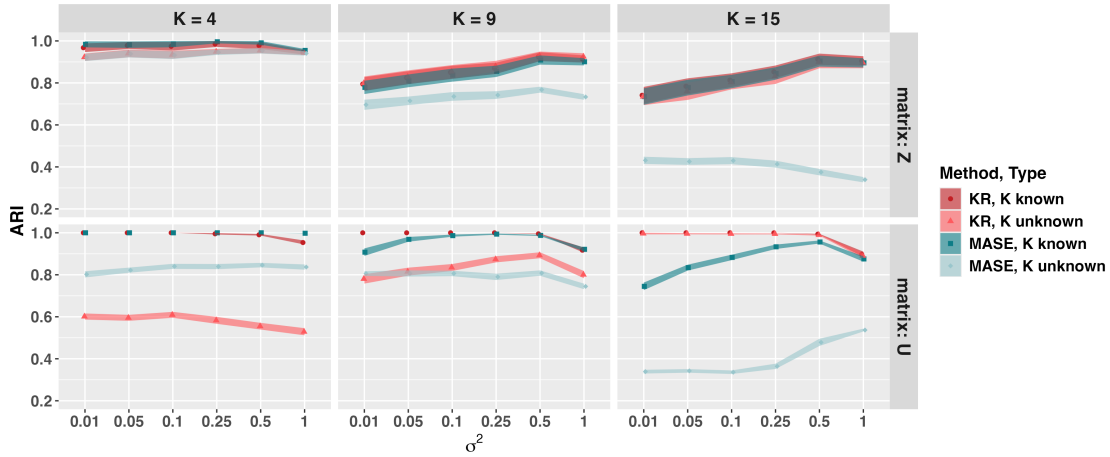


图 14：在已知和未知 K 的情况下，使用 $\widehat{Z}_{(v)}$ 或 $\widehat{U}^{(v)}$ 矩阵评估 KR 和 MASE 的联合聚类性能。联合簇的真实数量 K 固定为 4、9 和 15，而噪声水平 σ^2 不断变化。调整兰德指数将估计的联合簇与真实值进行比较。 $K_1 = K_2 = 4$， $p = 20$。

相近（当已知 K 时），但若未知 K，KR 和 MASE 之间的性能差距很大。

图 15 和图 16 研究了不同输入（$\widehat{Z}_{(v)}$ 与 $U_{(v)}$）在联合簇数量 K 未知时对 MASE 和 KR 性能的影响。在图 15 中，联合聚类性能（以调整兰德指数衡量）在噪声水平固定为 0.01 或 1 的情况下进行评估，而 K 在 4 到 16 之间变化。我们观察到，当 $K > K_1 + K_2$时，MASE 通常表现不佳。然而，在低噪声条件下，随着 K增加，$\widehat{Z}_{(v)}$ 的性能优于 $\widehat{U}_{(v)}$；而在高噪声条件下，情况相反（$\widehat{U}_{(v)}$ 的性能优于 $\widehat{Z}_{(v)}$）。KR 使用不同输入时的性能也会因噪声和联合簇数量的不同组合而变化。在低噪声条件下，$\widehat{U}_{(v)}$ 的性能优于 $\widehat{Z}_{(v)}$；而在高噪声条件下，当 $K > K_1 + K_2$时，$\widehat{Z}_{(v)}$ 的性能优于 $\widehat{U}_{(v)}$。图 16 展示了热图，说明在同时改变 K 和 σ^2 时估计 K 的性能。颜色表示两个输入在估计 K 方面的平均绝对误差（MAE）差异（计算方式为 $\text{MAE}_Z - \text{MAE}_U$）。因此，冷色表示 $\widehat{Z}^{(v)}$ 的性能更好（误差更低），而暖色表示 $\widehat{U}_{(v)}$ 的性能更好。结果显示，对于 MASE，$\widehat{U}_{(v)}$ 与 $\widehat{Z}_{(v)}$ 之间的差异可以忽略不计。对于 KR，在 K和 σ^2水平较低时，$\widehat{Z}_{(v)}$ 的结果更好；然而，随着这两个参数增加，这一趋势会反转，最终 $\widehat{U}_{(v)}$ 的性能优于 $\widehat{Z}_{(v)}$。

C.2. 变化的信噪比

图 17 展示了热图，说明随着维度和噪声水平变化，KR 和 MASE 的性能如何变化。颜色刻度表示调整兰德指数之差（$\text{ARI}_{\text{KR}} - \text{ARI}_{\text{MASE}}$）；因此，冷色表示 MASE 的性能更好，而暖色表示 KR 的性能更好。在已知 K设置中（前两列），无论维度和噪声水平如何变化，输入 $\widehat{Z}_{(v)}$ 对两种方法都产生相似的结果。对于输入 $\widehat{U}_{(v)}$，当 $K < K_1 + K_2$时，优先选择 MASE；而在高信噪比下，当 $K > K_1 + K_2$时，KR 的表现更好。在未知 K设置中，对于 $\widehat{U}_{(v)}$，当 K较小时，优先选择 MASE。当 $K = 12$且信噪比较高时，KR 更好。值得注意的是，使用 $\widehat{Z}_{(v)}$ 时，在所有信噪比下，无论 $\sim$ $\sim K$ 较小还是较大，KR 的表现始终优于 MASE。我们在图 18 中对 KR 和 MASE 的 U 和 Z 方法进行了类似的比较。

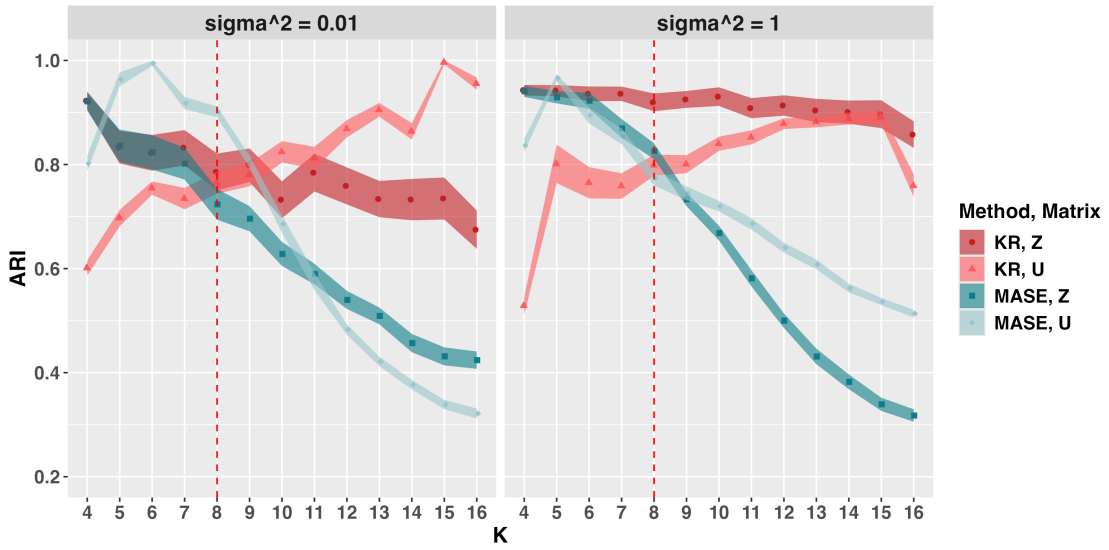


图 15：在未知的 K设置下，四种组合的联合聚类性能：KRAFTY 配合 $\widehat{Z}(\nu)$ 、KRAFTY 配合 $\widehat{U}(\nu)$ 、MASE 配合 $\widehat{Z}(\nu)$，以及 MASE 配合 $\widehat{U}(\nu)$。噪声水平 σ^2 设为 0.01 或 1。调整兰德指数用于比较估计的联合簇与真实值。$K_1 = K_2 = 4,\ \ p = 20$。

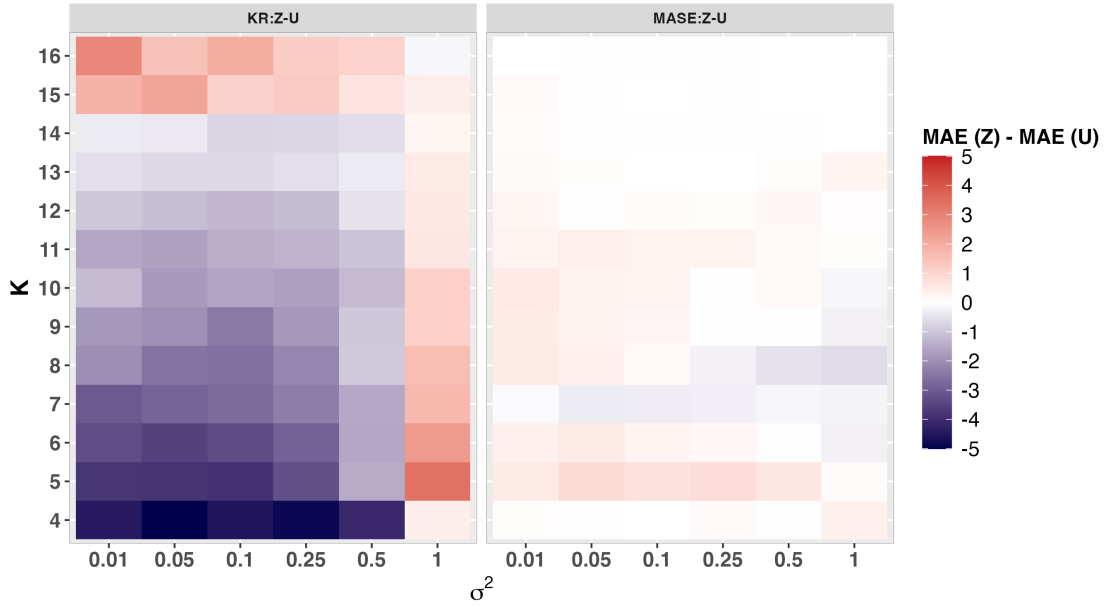


图 16：比较 KR 和 MASE 方法中 Z 与 U 的联合簇数估计性能。热图颜色表示使用 Z 相较于 U 估计 K时平均绝对误差的差异（对 100 次重复取平均）。**冷色 → 表示 Z 更优；暖色→表示 U 更优**。 $K_1 = K_2 = 4,\ \ p = 20$。

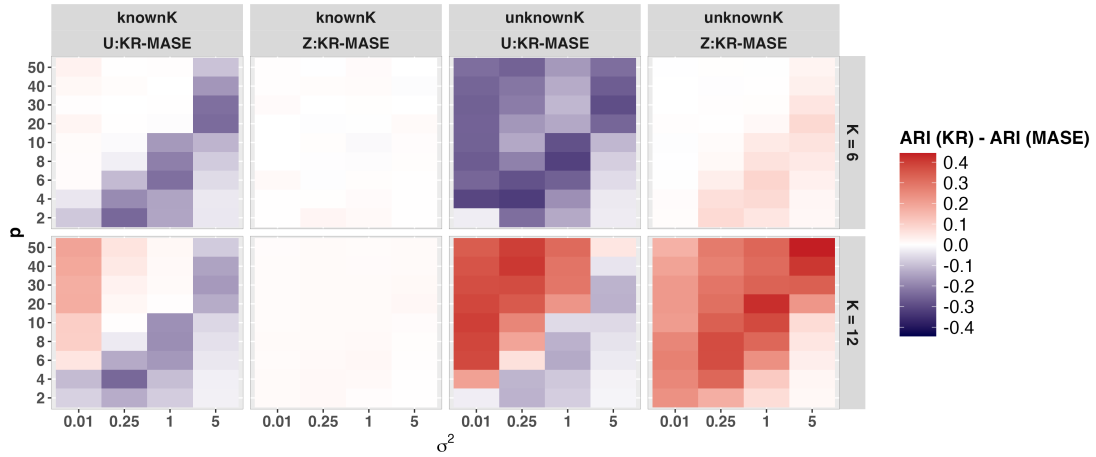


图 17：比较 KR 和 MASE 在 Z 矩阵与 U矩阵上的联合聚类性能。热图左侧 2 列显示当联合簇数量 K 已**知**时，使用 Z 矩阵或 U矩阵，KR 与 MASE 的平均 ARI（对 100 次重复取平均）之差。右侧 2 列显示当 K 为**未知**时，KR 与 MASE 的平均 ARI 之差。**冷色 →表示 MASE 更优；暖色 → 表示 KR 更优**。 $K_1 = K_2 = 4$。

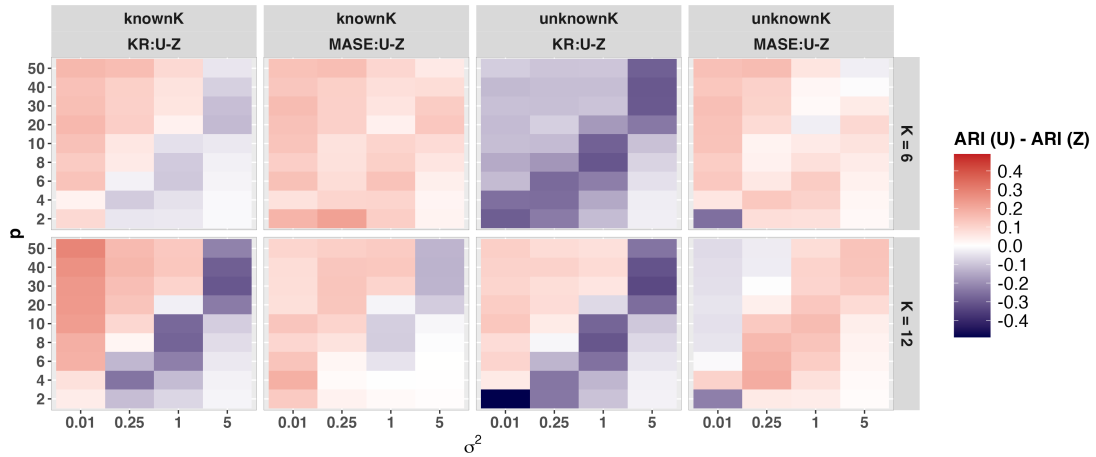


图 18：比较 KR 和 MASE 方法下 Z 与 U 的联合聚类性能。热图左侧 2 列显示当联合簇数量 K 已**知**时，使用 KR 或 MASE 方法，Z 与 U 的平均 ARI（对 100 次重复取平均）之差。右侧 2 列显示当 K为**未知**时，Z 与 U 的平均 ARI 之差。**冷色 → 表示 Z 更优；暖色 → 表示 U 更优**。 $K_1 = K_2 = 4$。

C.3. 单视图中的簇数变化

为得出更稳健的结论，我们进行了额外的模拟，其中 K_1 和 K_2 从 3 变化到 10，真实的联合簇数量 K 取四个具有代表性的值：两个位于$K \le K_1 + K_2$的范围内，具体为 $\max(K_1, K_2)$ 和 $K_1 + K_2$；另两个超出该范围，分别为 $\lfloor K_1 K_2/2 \rfloor$ 和 $K_1 K_2$。噪声水平固定为 $\sigma^2 = 0$。1，维度固定为 $p = 101$。结果以热图形式展示在图 19 和图 20 中。

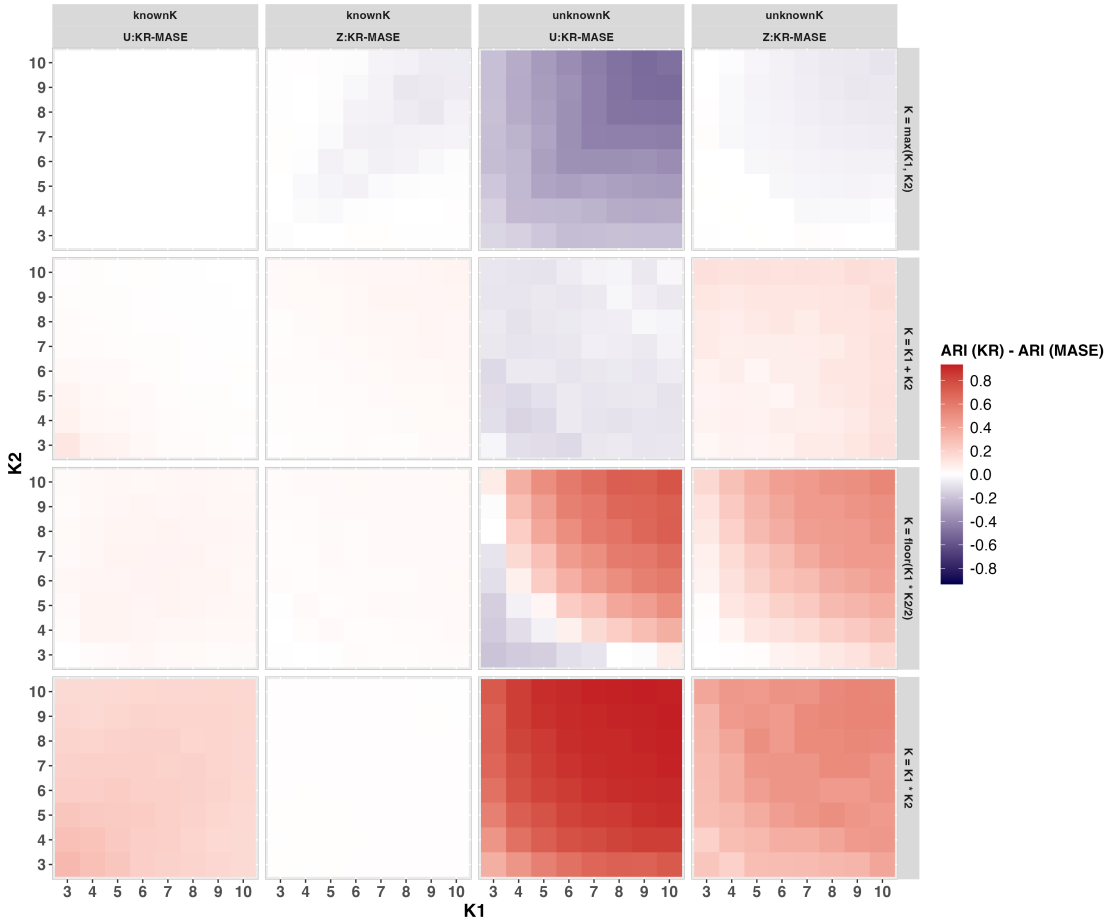


图 19：在 4 种不同的真实 K 设置下（每行对应一种），比较 KR 和 MASE 使用 Z 矩阵与 U矩阵时的联合聚类性能。热图左侧 2 列显示当联合簇数量 K **已知**时，KR 与 MASE 的平均 ARI（在 100 次重复中取平均）之差，所用矩阵为 Z 矩阵或 U矩阵。右侧 2 列显示当 K **未知**时，KR 与 MASE 的平均 ARI 之差。**冷色 → 表示 MASE 更优；暖色 →表示 KR 更优**。 $\sigma^2 = 0.1, p = 101$。

图 19 进一步证实，KRAFTY 的优势在于能够正确估计K。左侧两列（其中已知 K）中的热图几乎呈白色或浅色，表明在决定 K大小的不同情形下，KRAFTY 和 MASE 的联合聚类性能相近。相比之下，右侧两列（其中未知 K）呈现出更明显的暖色，且随着 K 增加而加深，反映出 KRAFTY 性能的提升。

在选择使用 $\widehat{Z}_{(v)}$矩阵还是 $\widehat{U}_{(v)}$矩阵时，图 20 的右侧第二列证实：当 K较小时，使用 $\widehat{Z}_{(v)}$ 矩阵的 KRAFTY 联合聚类性能更好；而当 K较大时，使用 $\widehat{U}^{(v)}$矩阵的 KRAFTY 表现更佳，这一点可从上方各面板中的冷色和下方各面板中的暖色看出。

图 21 展示了在单视图簇数量不同时对 K 进行估计的性能。与前两幅图类似，K_1 和 K_2 的范围为 3 到 10，维度固定为 101，噪声水平为 $\sigma^2 = 0.1$，且真实的 K 定义如下：$\max(K_1, K_2)$、$K_1 + K_2$、$\lfloor K_1 K_2/2 \rfloor$ 和 $K_1 K_2$。热图颜色表示平均绝对误差（$\mathrm{MAE}_{\mathrm{MASE}} - \mathrm{MAE}_{\mathrm{KR}}$）之间的差异；因此，冷色

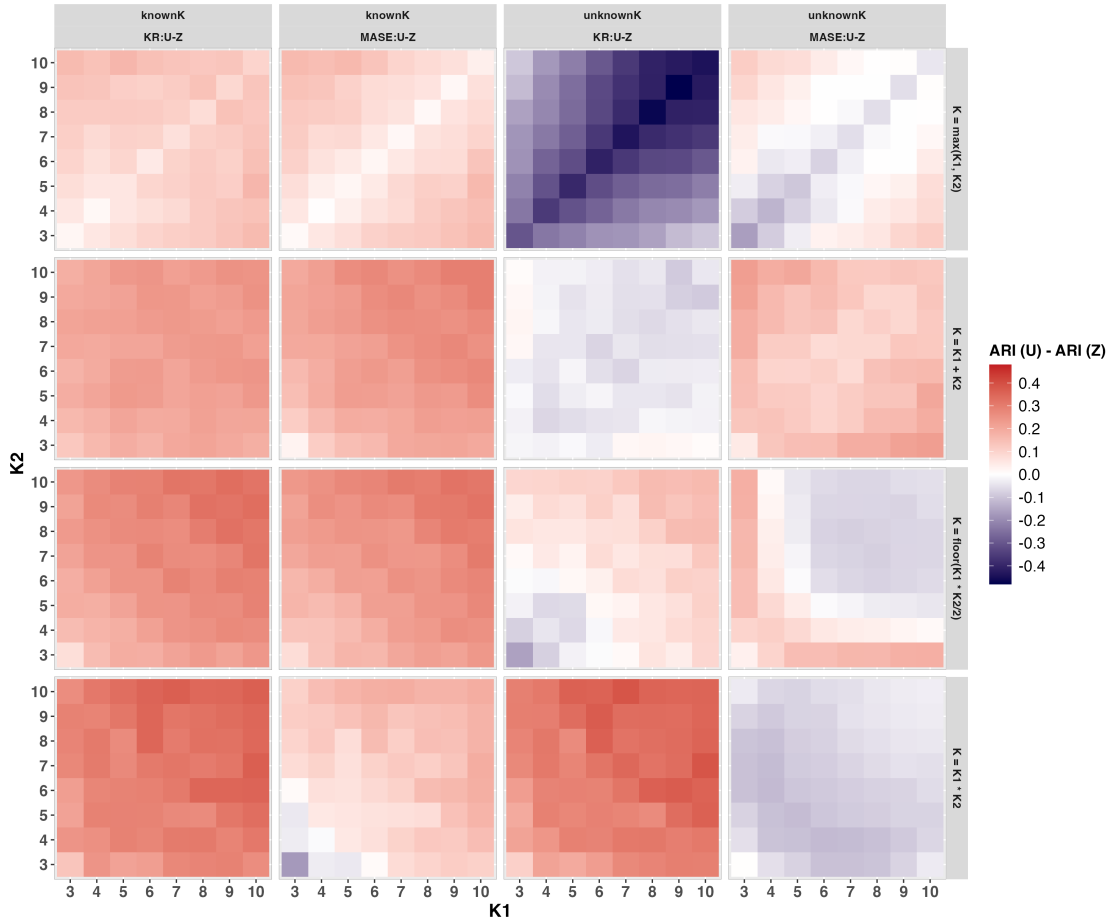


图 20：比较在 4 种不同的真实 K 设置下（每种设置显示于每一行）使用 KR 或 MASE 时 Z 与 U 的联合聚类性能。热图左侧 2 列显示当联合簇数量 K **已知**时，使用 KR 或 MASE 得到的 U 与 Z 之间的平均 ARI 差值（对 100 次重复取平均）。右侧 2 列显示当 K **未知**时 U 与 Z 之间的平均 ARI 差值。**冷色 → Z 更优；暖色 → U 更优**。$\sigma = 0.1$，$p = 101$。

表示 MASE 表现更好，暖色则表示 KR 表现更好。在情景 $K = \max(K_1, K_2)$ 和 $K = K_1 + K_2$ 下，MASE 通常优于 KR，因为 $K \le K_1 + K_2$。然而，在某些使用 $\widehat{Z}^{(v)}$ 作为输入的情况下，KR 可以达到相当的性能。当 $K > K_1 + K_2$ 时，随着 K 不断超过 $K_1 + K_2$，KR 的优势会更加明显。

图 22 展示了在四种真实 K 设置下，使用 $\hat{}Z$ 或 $\hat{}U$ 作为输入来估计 K 时的性能差异（平均绝对误差$_Z$ −平均绝对误差$_U$）。顶行（KR）和底行（MASE）呈现出不同的行为。对于 MASE，输入的选择几乎没有影响。然而，对于 KR，模式因设置而异：对于 $K \le K_1 + K_2$，$\widehat{Z}^{(v)}$ 在 K 较小时更优，但随着 K 增加，$\widehat{U}^{(v)}$ 的表现会更好；对于 $K = K_1 K_2$，U 是更好的输入。相反，在 $K = \lfloor K_1 K_2/2 \rfloor$ 设置下，当 $K < K_1 + K_2$ 时，$\widehat{Z}^{(v)}$ 产生的误差更低。

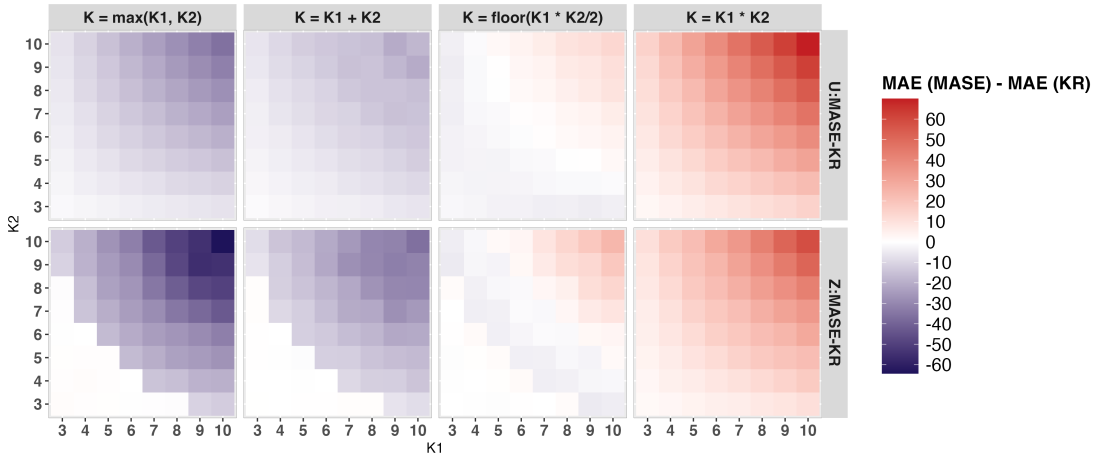


图 21：在 4 种真实的 K 设置下，比较使用 Z 矩阵或 U矩阵时 KR 与 MASE 的联合簇数估计性能。热图顶行显示使用 U矩阵时 MASE 与 KR 的平均绝对误差之差（对 100 次重复取平均）。底行显示使用 Z矩阵时 KR 与 MASE 的平均绝对误差之差。**冷色 → 表示 MASE 表现更好；暖色 →表示 KR 表现更好**。 $\sigma^2 = 0。1, p = 101。$

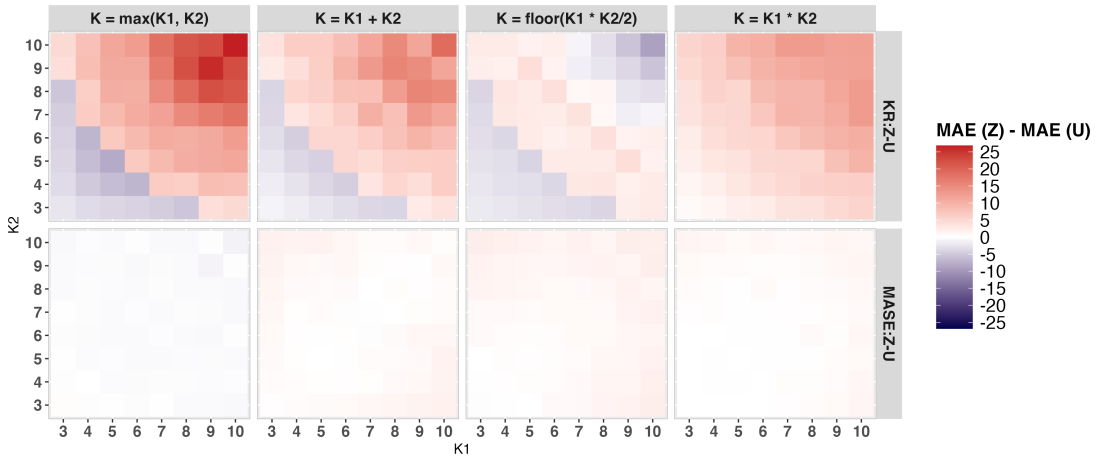


图 22：在 4 种不同的真实 K 设置下，比较使用 KR 或 MASE 时 Z 与 U 的联合簇数估计性能。热图顶行显示使用 KR 时，采用 Z 与采用 U 估计 K 所得的平均绝对误差之差（对 100 次重复取平均）。**冷色 → 表示 Z 表现更好；暖色 → 表示 U 表现更好**。 底行显示使用 MASE 时 Z 与 U 的平均绝对误差之差。 $\sigma = 0。1，\ p = 101。$

C.4。比较 $XandU$

本节展示了 $\hat{U}^{(v)}$与其缩放版本 $\hat{X}_{(v)} = \hat{U}_{(v)} (D_{(v)})^{1/2}$之间的性能相似性。与我们之前的分析一致，我们考察了相同的三种实验设置。

改变联合簇数

图 23 展示了 K在两个噪声水平下变化时的性能。结果表明，对于 KR 和 MASE，选择 $\hat{X}_{(v)}$还是 $\hat{U}^{(v)}$对联合聚类的影响都可以忽略不计。



图 23：在**未知** K 设置下，使用 KR 和 MASE 对 X 矩阵与 U矩阵进行联合聚类性能比较。噪声水平 σ 设置为 0.01 或 1。调整兰德指数比较估计的联合簇与真实值。此处，$K_1 = K_2 = 4$，$p = 20$。

图 24 展示了 σ^2 在三种不同 K 情景下变化时的性能。我们观察到，代表 $\widehat{X}^{(v)}$ 和 $\widehat{U}^{(v)}$ 的曲线相互重叠。这表明，对于 KR 和 MASE，选择这些输入对联合聚类性能的影响都可以忽略不计。

图 25 展示了比较输入性能的热图。$\widehat{U}^{(v)}$ 在估计 K 时的表现，观察到的误差差异可以忽略不计。$\widehat{X}^{(v)}$ 和 $\widehat{}$

变化的信噪比 我们比较输入的性能 (v) 在联合聚类（图 26）和估计 K（图 27）中，随着维度和噪声水平的变化。与之前的发现一致，两个输入之间的性能差异仍然很小。

单视图簇数变化 我们比较输入的性能 (v) 在联合聚类（图 28）和估计 K（图 29）中，针对不同的单视图簇数以及四种不同的 K 设置。与之前的发现一致，两个输入之间的性能差异仍然可以忽略不计。



图 24：在**未知** K 设置下，使用 KR 和 MASE 对 X 矩阵和 U矩阵进行联合聚类的性能。实验中，联合簇的真实数量 K 固定为 4、9 和 15，而 σ 则有所变化。调整兰德指数（ARI）用于比较估计的联合簇与真实值。此处，$K_1 = K_2 = 4$，$p = 20$。

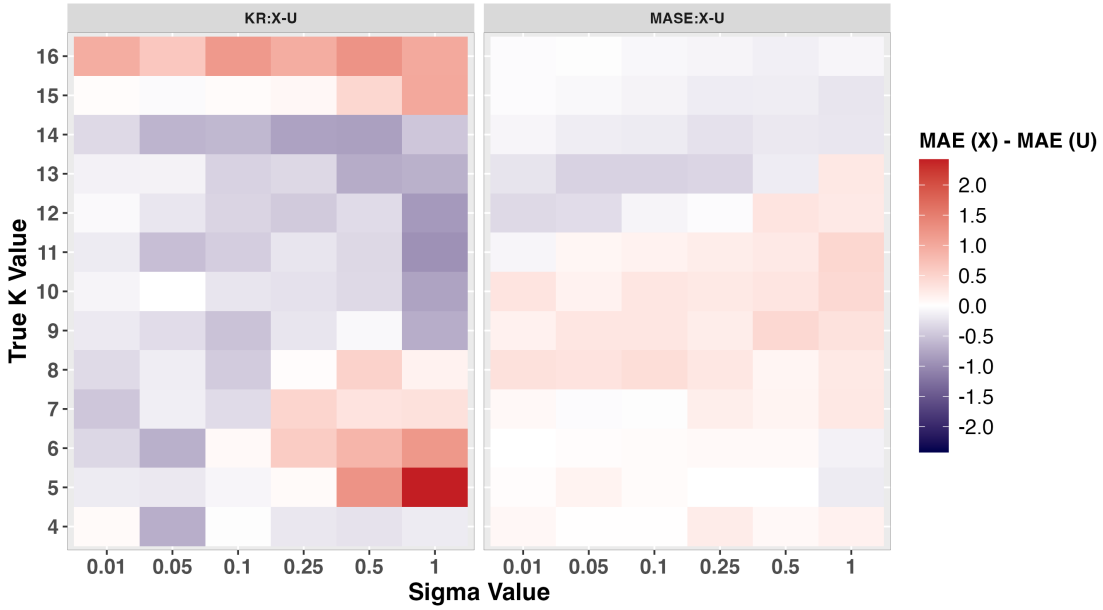


图 25：比较 KR 和 MASE 方法下 X 与 U 的联合簇数估计性能。热图颜色表示使用 X 与使用 U 估计 K 时平均绝对误差（按 100 次重复的结果取平均值）之间的差异。**冷色 → 表示 X 更优；暖色 → 表示 U 更优**。 $K_1 = K_2 = 4$，$p = 20$。

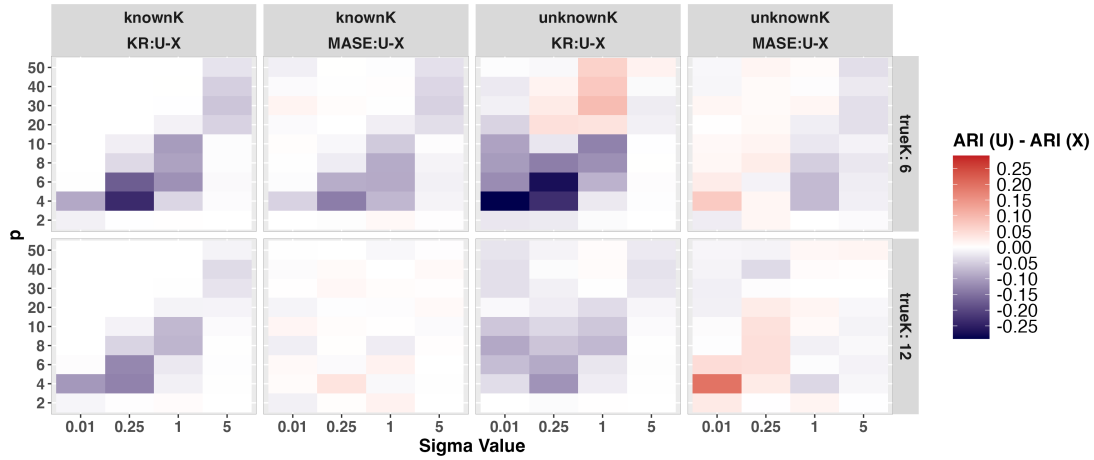


图 26：KR 和 MASE 方法下，X 与 U 的联合聚类性能比较。热图左侧 2 列显示在联合簇数 K **已知**时，X 与 U 的平均 ARI（基于 100 次重复计算的平均值）之差，分别采用 KR 或 MASE 方法。右侧 2 列显示在 K**未知**时，X 与 U 的平均 ARI 之差。**冷色 → 表示 X 更优；暖色 →表示 U 更优**。 $K_1 = K_2 = 4$。

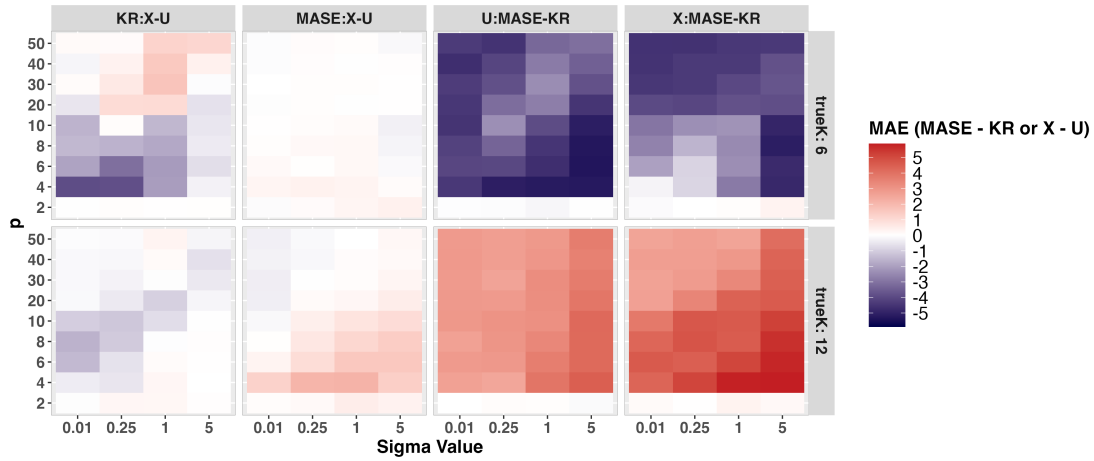


图 27：X 和 U矩阵下 KR 与 MASE 的联合簇数估计性能比较。热图左侧两列显示使用 X 与 U 估计 K 时平均绝对误差（基于 100 次重复计算的平均值）之差。**冷色→ 表示 X 更优；暖色 → 表示 U 更优**。右侧 2 列显示 KR 与 MASE 方法的平均绝对误差之差。**冷色 → 表示 MASE 更优；暖色 →表示 KR 更优**。此处， $K_1 = K_2 = 4$。

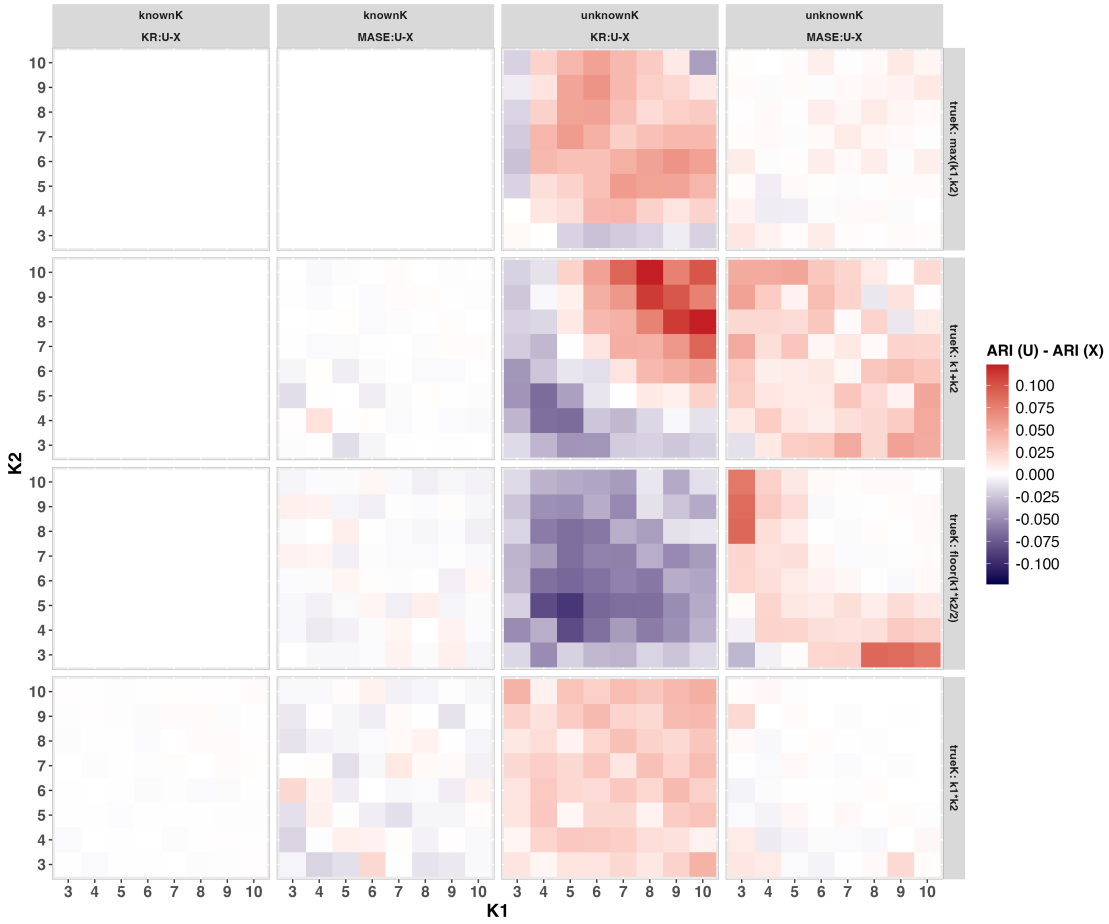


图 28：在 4 种不同的真实 K 设定下，使用 KR 或 MASE 比较 X 与 U 的联合聚类性能。热图左侧的 2 列显示在联合簇数量 K **已知**时，U 与 X 的平均 ARI（基于 100 次重复计算的平均值）之差，分别对应 KR 和 MASE。右侧的 2 列显示在 K **未知**时，U 与 X 的平均 ARI 之差。**冷色 →表示 X 更优；暖色 →表示 U 更优**。 $\sigma = 0。1$, $p = 101。$

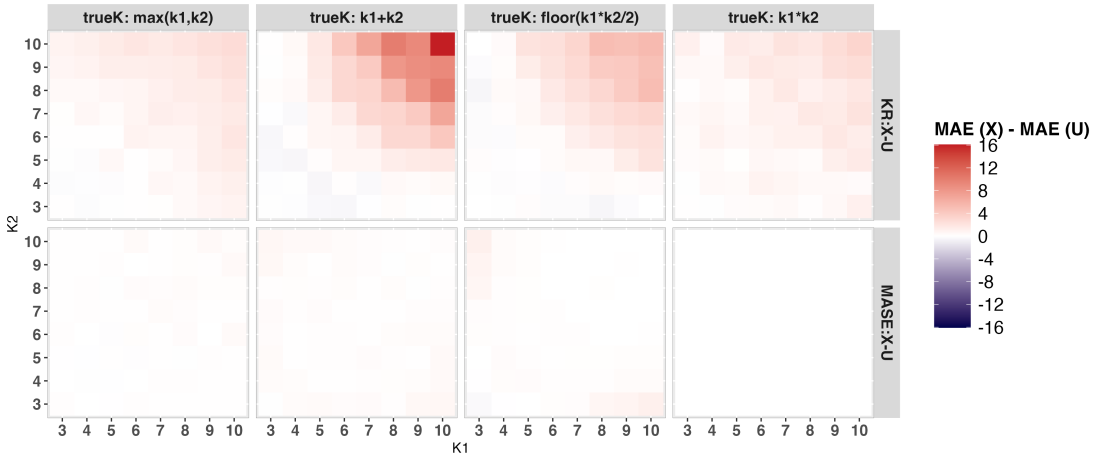


图 29：在 4 种不同的真实 K 设置下，比较使用 KR 或 MASE 时，基于 X 和 U 进行联合簇数量估计的性能。热图顶行显示使用 KR 时，基于 X 与基于 U 估计 K 的平均绝对误差之差（对 100 次重复取平均）。**冷色 → 表示 X 更优；暖色 → 表示 U 更优**。 底行显示使用 MASE 时，基于 X 与基于 U 的平均绝对误差之差。 $\sigma = 0。1$， $p = 101$。

D. 真实数据分析补充

本节提供真实数据分析的补充图，以及两年分析的相关图示。

D.1. *Single-year Results on Global Trade Data Analysis*

图 30 展示了生鸡肉贸易网络的邻接矩阵。对于每一对国家，将出口量和进口量相加，使矩阵对称化。行和列按照 KRAFTY 估计的簇排序，从簇 1 到簇 5（与图 10 中的顺序一致）。值得注意的是，簇 1 和簇 4 内部的贸易量较小。

D.2. 全球贸易数据分析的两年结果

我们的目标是识别作为出口商或进口商的国家之间时间不变的潜在结构，即“保持不变”的簇。我们将 KRAFTY 和 MASE 应用于 $\hat{Z}^{(v)}$ with $v \in \{(2010，\text{imp})，(2023，\text{exp})\}$ and $v \in \{(2010，\text{exp})，(2023，\text{exp})\}$。结果如图 31–38 所示。碎石图显示，KRAFTY 下存在明显的肘部，这与我们的理论一致；此外，KRAFTY 和 MASE 得出的联合簇高度一致（图 35、36），与我们的模拟结果相符。将 KRAFTY 得出的联合簇与单视图簇的组合进行比较，可以看出联合簇是如何从单视图划分中得出的（图 33、34）。一种解释是，各国的进出口行为随年份而变化，从而产生视图特定差异；与此同时，稳定的联合结构依然存在。

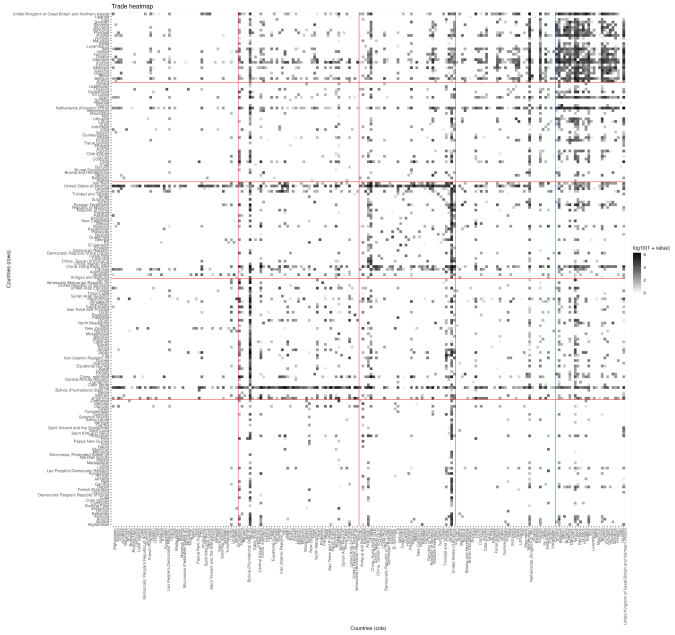
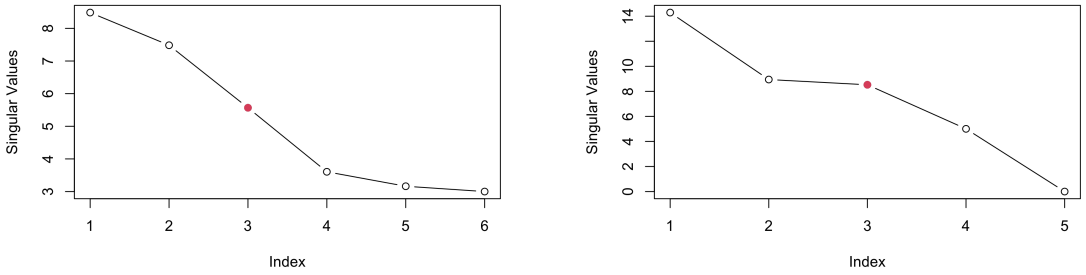


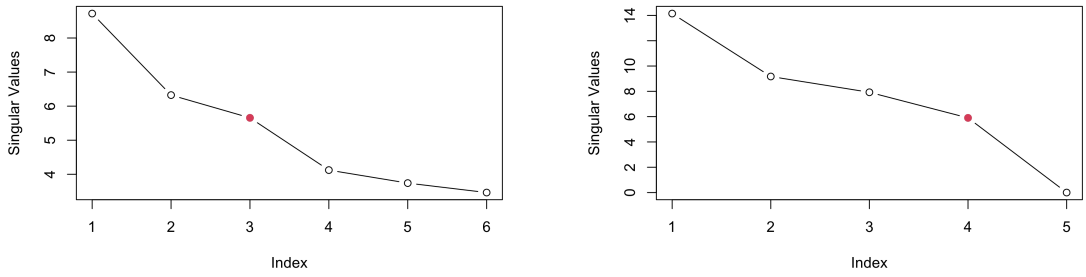
图30：全球贸易邻接热图；国家按KRAFTY联合聚类排序，边界以红线标示。



（a）KRAFTY

（b）MASE

图31：仅针对两年期出口商的特征值碎石图；红点表示选定维度。



(a) KRAFTY (b) MASE

图32：仅针对两年期进口国的特征值碎石图；红点表示选定维度。

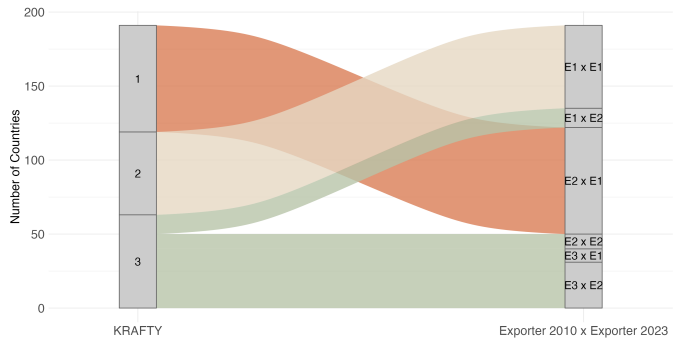


图33：KRAFTY联合簇与单一来源聚类（出口国）的桑基图；相同颜色表示同一组实体，色块高度反映每个簇中的国家数量。

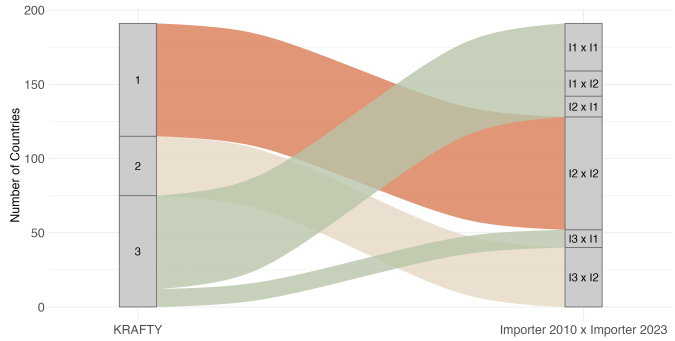


图34：KRAFTY联合簇与单一来源聚类（进口国）的桑基图；相同颜色表示同一组实体，色块高度反映每个簇中的国家数量。



图 35：按 KRAFTY 与 MASE 划分的联合簇桑基图（出口国）；相同颜色表示相同的实体集合，区块高度反映每个簇中的国家数量。



图 36：按 KRAFTY 与 MASE 划分的联合簇桑基图（进口国）；相同颜色表示相同的实体集合，区块高度反映每个簇中的国家数量。



图 37：按 KRAFTY联合聚类为两年期出口商着色的世界地图；蓝色表示聚类归属。



图 38：按 KRAFTY联合聚类为两年期进口国着色的世界地图；蓝色表示聚类归属。