Supplement to "Unfolding-Model-Based Visualization: Theory, Method and Applications"

A Bi-Cluster Analysis

The applications of multidimensional scaling, including multidimensional unfolding as a special case, are often followed by cluster analysis (e.g., Kruskal and Wish, 1978; Borg and Groenen, 2005) for better understanding and interpretation of the data visualization. In our context, it is often of interest to cluster the respondents and the items, respectively. This task is known as bi-clustering or co-clustering (Hartigan, 1972; Dhillon, 2001), which is often studied statistically under the stochastic co-blockmodel (Choi and Wolfe, 2014; Rohe et al., 2016), an extension of the widely used stochastic blockmodel (Holland et al., 1983).

Following multidimensional unfolding, it is natural to bi-cluster the respondents and the items based on the estimated ideal points, using the Euclidian distance as a natural measure of dissimilarity. In particular, we use the K-means algorithm (MacQueen, 1967) to cluster the respondents and the items into k_1 and k_2 clusters, respectively, for some pre-specified numbers of clusters k_1 and k_2. This two-step procedure for bi-cluster analysis is described in Algorithm A.1.

Algorithm A.1 (Two-step procedure for bi-cluster analysis)

***Step 1**: Apply Algorithm 1 and obtain estimates* $\{\hat{\boldsymbol{\theta}}_1, ..., \hat{\boldsymbol{\theta}}_N, \hat{\mathbf{a}}_1, ..., \hat{\mathbf{a}}_J\}$.

***Step 2**: Perform the K-means algorithm to* $\{\hat{\boldsymbol{\theta}}_1, ..., \hat{\boldsymbol{\theta}}_N\}$ *and* $\{\hat{\mathbf{a}}_1, ..., \hat{\mathbf{a}}_J\}$ *given* k_1 *and* k_2 *clusters, respectively.*

Output: *The cluster membership of respondents $\hat{\vartheta}_i \in \{1, ..., k_1\}$ and cluster membership of items $\hat{\upsilon}_j \in \{1, ..., k_2\}$ $(i = 1, ..., N;\ j = 1, ..., J)$.*

We provide a connection between the multidimensional unfolding model studied in this paper and the stochastic co-blockmodel. Consider a special case under the multidimensional unfolding model, where there are finite possible locations for the respondent ideal points and also for the item ideal points, independent of N and J. We denote the possible locations for the respondent ideal points as $\{\mathbf{b}_1^*, ..., \mathbf{b}_{k_1}^*\}$ and denote those for the item ideal points as $\{\mathbf{c}_1^*, ..., \mathbf{c}_{k_2}^*\}$. Under this setting, there exist k_1 respondent latent classes and k_2 item latent classes, regarding two respondents/items as from the same latent class when they have the same location. We denote $\vartheta_i^* \in \{1, ..., k_1\}$ and $\upsilon_j^* \in \{1, ..., k_2\}$ the true latent class memberships of respondent i and item j, respectively. In this sense, the model becomes a stochastic co-blockmodel, for which the distribution of Y_{ij} is only determined by the latent class memberships of respondent i and item j and Y_{ij}s are conditionally independent given all the latent memberships of the respondents and items. In what follows, we show that the proportions of misclassified respondents and items converge to 0 in probability, when both N and J grow to infinity, if the K-means algorithm in Algorithm A.1 has converged to the global optima.

Theorem A.1 *Suppose A0, A3 and A4 are satisfied, and $K_+ \geq K$. Further suppose the multidimensional unfolding model degenerates to a stochastic co-blockmodel, satisfying $\boldsymbol{\theta}_i^* \in \{\mathbf{b}_1^*, ..., \mathbf{b}_{k_1}^*\}$ and $\mathbf{a}_j^* \in \{\mathbf{c}_1^*, ..., \mathbf{c}_{k_2}^*\}$. If both K-means algorithms in Algorithm A.1 converge to the global optima, then the clustering result satisfies*

$$\min\left\{\max_{\zeta\in\mathcal{B}_{k_1}} \frac{\sum_{i=1}^N 1_{\{\hat{\vartheta}_i=\zeta(\vartheta_i^*)\}}}{N}, \max_{\zeta\in\mathcal{B}_{k_2}} \frac{\sum_{j=1}^J 1_{\{\hat{\upsilon}_j=\zeta(\upsilon_j^*)\}}}{J}\right\} \tag{A.1}$$

goes to 1 in probability as both N and J grow to infinity, where $\mathcal{B}_k$ denotes the set of all permutations on $\{1, ..., k\}$, for $k = k_1, k_2$.

Remark A.1 *To handle "label switching indeterminacy" in clustering, in the loss function* (A.1) *we find permutations that best match the true latent class memberships and their estimates for both the respondents and the items.*

B Proof of Theoretical Results

B.1 Definitions and Notations

In this appendix, we use c, C, C_1, C_2 to represent constants which do not depend on N, J, the values of which may vary according to the context. With a little abuse of notation, we use $\mathcal{A}_{N,J}$ to denote the specified events, which may differ in different proofs. For $\mathbf{x} \in \mathbb{R}^K$, we use $B_{\mathbf{x}}^K(C)$ to denote the closed ball in $\mathbb{R}^K$ centered at $\mathbf{x}$ with radius C. Unless otherwise specified, all balls in the appendix is assumed to be closed. For a set $G \subset \mathbb{R}^K$, let $\mathrm{int}(G)$ denote the set of all its interior points. For a positive integer n, we denote $[n] := \{1, ..., n\}$. We start with some notions which will be used in the proof of theorems, propositions and lemmas.

Definition B.1 *For points* $\mathbf{x}_i, \mathbf{x}'_i \in \mathbb{R}^K, i = 1, ..., n$, *we write* $(\mathbf{x}_1, ..., \mathbf{x}_n) \sim (\mathbf{x}'_1, ..., \mathbf{x}'_n)$, *if there exists an isometry* $F \in \mathcal{A}_K$, *such that* $\mathbf{x}'_i = F(\mathbf{x}_i), i = 1, ..., n$.

Remark B.1 *It is easy to show that "$\sim$" is an equivalence relation.*

Definition B.2 (Configuration) *We define an n-point configuration as an equivalence class. That is, we define a configuration*

$$[\mathbf{x}_1, ..., \mathbf{x}_n] := \{(\mathbf{x}'_1, ..., \mathbf{x}'_n) : (\mathbf{x}'_1, ..., \mathbf{x}'_n) \sim (\mathbf{x}_1, ..., \mathbf{x}_n)\}$$

as the equivalence class of $(\mathbf{x}_1, ..., \mathbf{x}_n)$.

Remark B.2 *By the property of isometry mapping, it is easy to see that all the elements in the same configuration have the same distance matrix.*

We now consider the space of all n-point configurations in $\mathbb{R}^K$, denoted by

$$\mathcal{H}_{n,K} := \left\{[\mathbf{x}_1, ..., \mathbf{x}_n] : \mathbf{x}_i \in \mathbb{R}^K, i = 1, ..., n\right\}.$$

For two configurations $\tau_1 = [\mathbf{x}_1, ..., \mathbf{x}_n], \tau_2 = [\mathbf{y}_1, ..., \mathbf{y}_n] \in \mathcal{H}_{n,K}$, we define

$$d(\tau_1, \tau_2) := \inf_{F \in \mathcal{A}_K} \sqrt{\sum_{1 \le i \le n} \|F(\mathbf{x}_i) - \mathbf{y}_i\|^2}.$$

First, we note that $d(\cdot, \cdot)$ is a well-defined mapping from $\mathcal{H}_{n,K} \times \mathcal{H}_{n,K}$ to $\mathbb{R}$. That is, for any $(\mathbf{x}'_1, ..., \mathbf{x}'_n) \in [\mathbf{x}_1, ..., \mathbf{x}_n]$ and $(\mathbf{y}'_1, ..., \mathbf{y}'_n) \in [\mathbf{y}_1, ..., \mathbf{y}_n]$,

$$\inf_{F \in \mathcal{A}_K} \sqrt{\sum_{1 \le i \le n} \|F(\mathbf{x}_i) - \mathbf{y}_i\|^2} = \inf_{F \in \mathcal{A}_K} \sqrt{\sum_{1 \le i \le n} \|F(\mathbf{x}'_i) - \mathbf{y}'_i\|^2}.$$

Second, we notice that $d(\cdot, \cdot)$ is a metric on $\mathcal{H}_{n,K}$, as summarized in Lemma B.1 below.

Lemma B.1 *$d(\cdot, \cdot)$ is a metric on $\mathcal{H}_{n,K}$.*

Remark B.3 *For $[\mathbf{x}_1, ..., \mathbf{x}_n] \in \mathcal{H}_{n,K}$, we have $[(\mathbf{x}_1^\top, 0)^\top, ..., (\mathbf{x}_n^\top, 0)^\top] \in \mathcal{H}_{n,K+1}$ in which sense we can say $\mathcal{H}_{n,K} \subset \mathcal{H}_{n,K+1}$. Thus $\mathcal{H}_{n,K_1} \subset \mathcal{H}_{n,K_2}$ if $K_1 \le K_2$. For $\tau_1 = [\mathbf{x}_1, ..., \mathbf{x}_n] \in \mathcal{H}_{n,K_1}, \tau_2 = [\mathbf{y}_1, ..., \mathbf{y}_n] \in \mathcal{H}_{n,K_2}$, the $d(\tau_1, \tau_2)$ is defined in the same way by seeing both τ_1 and τ_2 as elements in $\mathcal{H}_{n,\max\{K_1,K_2\}}$.*

We further denote $\mathcal{P}_{a,b,K}$ as the set of $a \times b$ partial distance matrices for configurations in $\mathbb{R}^K$:

$$\mathcal{P}_{a,b,K} := \left\{(\|\mathbf{x}_i - \mathbf{y}_j\|^2)_{a \times b} : [\mathbf{x}_1, ..., \mathbf{x}_a, \mathbf{y}_1, ..., \mathbf{y}_b] \in \mathcal{H}_{a+b,K}\right\}.$$

It is easy to check that $\mathcal{P}_{a,b,K} \subset \mathcal{P}_{a,b,K+1}$.

For $A_1, ..., A_n \subset \mathbb{R}^K$, denote $[A_1, ..., A_n]$ as a subset of $\mathcal{H}_{n,K}$:

$$[A_1, ..., A_n] := \{[\mathbf{x}_1, ..., \mathbf{x}_n] : \mathbf{x}_i \in A_i, i = 1, ..., n\}.$$

For $A, B \subset \mathcal{H}_{n,K}$, the distance between A and B is defined as

$$d(A, B) := \inf_{\tau_1 \in A, \tau_2 \in B} d(\tau_1, \tau_2). \tag{B.1}$$

We further denote

$$\mathcal{H}_{n,K,C} := \{[\mathbf{x}_1, ..., \mathbf{x}_n] \in \mathcal{H}_{n,K} : \|\mathbf{x}_i\| \le C\}$$

as a compact subset of $\mathcal{H}_{n,K}$, and

$$\mathcal{P}_{a,b,K,C} := \left\{(\|\mathbf{x}_i - \mathbf{y}_j\|^2)_{a \times b} : [\mathbf{x}_1, ..., \mathbf{x}_a, \mathbf{y}_1, ..., \mathbf{y}_b] \in \mathcal{H}_{a+b,K,C}\right\} \tag{B.2}$$

as a compact subset of $\mathcal{P}_{a,b,K}$. We consider a mapping defined as following:

$$\Phi_{a,b,K} : \mathbb{R}^{(a+b)\times K} \to \mathcal{P}_{a,b,K},$$
$$(\mathbf{x}_1, ..., \mathbf{x}_{a+b})^\top \mapsto D,$$

where D is the $a \times b$ partial distance matrix of $\{(\mathbf{x}_1, ..., \mathbf{x}_a), (\mathbf{x}_{a+1}, ..., \mathbf{x}_{a+b})\}$. It is not difficult to check that $\Phi_{a,b,K}$ is invariant with respect to isometry. Then, for $\tau = [\mathbf{x}_1, ..., \mathbf{x}_{a+b}]$, we denote

$$\Phi_{a,b,K}(\tau) := \Phi_{a,b,K}(X),$$

where $X^\top = (\mathbf{x}_1, ..., \mathbf{x}_{a+b})$.

Having introduced the notions above, we give the following lemma, which is crucial to the proof of Theorem 1. It essentially shows that for any partial distance matrix $D' \in \mathcal{P}_{k_1,k_2,K_+,M}$ that approximates to another partial distance matrix $D \in \mathcal{P}_{k_1,k_2,K,M}$, whose configuration τ contains a collection of anchor points, then any configuration τ' of D' will also approximate to τ.

Lemma B.2 *For compact subsets $\mathcal{B}_1, ..., \mathcal{B}_{k_1}, \mathcal{C}_1, ..., \mathcal{C}_{k_2} \subset B_{\mathbf{0}}^K(M)$, let*

$$\mathcal{B} = [\mathcal{B}_1, ..., \mathcal{B}_{k_1}, \mathcal{C}_1, ..., \mathcal{C}_{k_2}].$$

Suppose that for any $(\mathbf{x}_1, ..., \mathbf{x}_{k_1+k_2}) \in \mathcal{B}_1 \times \cdots \times \mathcal{B}_{k_1} \times \mathcal{C}_1 \times \cdots \times \mathcal{C}_{k_2}$, $\{\mathbf{x}_1, ..., \mathbf{x}_{k_1}\}$ and $\{\mathbf{x}_{k_1+1}, ..., \mathbf{x}_{k_1+k_2}\}$ are a collection of anchor points in $\mathbb{R}^K$. Then, for any $\epsilon_c > 0$, there exists $\epsilon_d > 0$ such that for any $\tau' \in \mathcal{H}_{k_1+k_2,K_+,M}$ and $\tau \in \mathcal{B}$ satisfying

$$\|\Phi_{k_1,k_2,K_+}(\tau') - \Phi_{k_1,k_2,K_+}(\tau)\|_F < \epsilon_d,$$

we have

$$d(\tau', \tau) < \epsilon_c.$$

We end this section by the following lemma, which will also be used in the proof of Theorem 1.

Lemma B.3 *Suppose $\{\mathbf{b}_1^*, ..., \mathbf{b}_{k_1}^*\}, \{\mathbf{c}_1^*, ..., \mathbf{c}_{k_2}^*\} \subset B_{\mathbf{0}}^K(C)$ are a collection of anchor points in $\mathbb{R}^K$. Then, for any $\mathbf{x} \in B_{\mathbf{0}}^K(C)$, the $\{\mathbf{x}, \mathbf{b}_1^*, ..., \mathbf{b}_{k_1}^*\}, \{\mathbf{c}_1^*, ..., \mathbf{c}_{k_2}^*\}$ are also a collection of anchor points in $\mathbb{R}^K$.*

B.2 Proof of Theorems

Proof of Theorem 1. We first show the proof of (5). For ϵ which is given in condition A3, there exist constant $p_\epsilon \in (0, 1)$, and balls of radius ϵ in $\mathbb{R}^K$, denoted by

$\tilde{B}_1(\epsilon)$, ..., $\tilde{B}_{k_1}(\epsilon)$, $\tilde{G}_1(\epsilon), ..., \tilde{G}_{k_2}(\epsilon)$, such that for N, J large enough,

$$\frac{\sum_{l=1}^{N} 1_{\{\boldsymbol{\theta}_l^* \in B_{\mathbf{b}_i^*}(\epsilon), \tilde{\boldsymbol{\theta}}_l \in \tilde{B}_i(\epsilon)\}}}{N} > p_\epsilon, i = 1, ..., k_1,$$
$$\frac{\sum_{l=1}^{J} 1_{\{\mathbf{a}_l^* \in B_{\mathbf{c}_i^*}(\epsilon), \tilde{\mathbf{a}}_l \in \tilde{G}_i(\epsilon)\}}}{J} > p_\epsilon, i = 1, ..., k_2.$$

This comes straightforwardly from condition A0 and requirement (2) of anchor points in condition A3. Note that the centers of $\tilde{B}_k(\epsilon)$ and $\tilde{G}_l(\epsilon)$ may vary through N, J. We also use $B_k^*(\epsilon)$ and $G_l^*(\epsilon)$ to denote $B_{\mathbf{b}_k^*}(\epsilon)$ and $B_{\mathbf{c}_l^*}(\epsilon)$, respectively.

We first focus on the set of person points

$$I_1(\epsilon) := \bigcup_{k=1}^{k_1} \{i \in [N] : \boldsymbol{\theta}_i^* \in B_k^*(\epsilon), \tilde{\boldsymbol{\theta}}_i \in \tilde{B}_k(\epsilon)\}$$

and the set of item points

$$I_2(\epsilon) := \bigcup_{l=1}^{k_2} \{j \in [J] : \mathbf{a}_j^* \in G_l^*(\epsilon), \tilde{\mathbf{a}}_j \in \tilde{G}_l(\epsilon)\}.$$

Let $\boldsymbol{\theta}_i^+ = \left((\boldsymbol{\theta}_i^*)^\top, \mathbf{0}^\top\right)^\top, \mathbf{a}_j^+ = \left((\mathbf{a}_j^*)^\top, \mathbf{0}^\top\right) \in \mathbb{R}^{K_+}$. We will show that there exists an isometry mapping $F_{N,J} \in \mathcal{A}_{K_+}$, under which $F_{N,J}(\tilde{\boldsymbol{\theta}}_i) \approx \boldsymbol{\theta}_i^+$ and $F_{N,J}(\tilde{\mathbf{a}}_j) \approx \mathbf{a}_j^+$, for all $i \in I_1(\epsilon)$ and $j \in I_2(\epsilon)$. This is formalized in the following lemma.

Lemma B.4 *For N, J large enough, there exists an isometry $F_{N,J} \in \mathcal{A}_{K_+}$, such that*

$$\|F_{N,J}(\mathbf{x})\| \leq 4M, \textit{ for all } \mathbf{x} \in B_{\mathbf{0}}^{K_+}(M),$$

and for all $i \in I_1(\epsilon)$ and for all $j \in I_2(\epsilon)$,

$$\|F_{N,J}(\tilde{\boldsymbol{\theta}}_i) - \boldsymbol{\theta}_i^+\| \leq 5\epsilon,$$

and

$$\|F_{N,J}(\tilde{\mathbf{a}}_j) - \mathbf{a}_j^+\| \leq 5\epsilon.$$

We then show that for most of the person points $i \notin I_1(\epsilon)$ and for most of the item points $j \notin I_2(\epsilon)$, we still have $F_{N,J}(\tilde{\boldsymbol{\theta}}_i) \approx \boldsymbol{\theta}_i^+$ and $F_{N,J}(\tilde{\mathbf{a}}_j) \approx \mathbf{a}_j^+$, under the same isometry mapping $F_{N,J}$ as in Lemma B.4. This is formalized in Lemma B.5 below.

Lemma B.5 *For N, J large enough, there exists a constant $\kappa > 0$, such that for the isometry mapping $F_{N,J}$ defined in Lemma B.4, the proportions*

$$\lambda_{1,N,J} = \frac{\sum_{i=1}^N 1_{\{\|F_{N,J}(\tilde{\boldsymbol{\theta}}_i) - \boldsymbol{\theta}_i^+\| > \kappa\epsilon\}}}{N}$$

and

$$\lambda_{2,N,J} = \frac{\sum_{j=1}^J 1_{\{\|F_{N,J}(\tilde{\mathbf{a}}_j) - \mathbf{a}_j^+\| > \kappa\epsilon\}}}{J}$$

satisfy

$$\lambda_{k,N,J} \to 0, \tag{B.3}$$

for $k = 1, 2,$ as N, J grow to infinity.

Since by Lemma B.4, we have $F_{N,J}$ maps $B_{\mathbf{0}}^{K_+}(M)$ to $B_{\mathbf{0}}^{K_+}(4M)$, then for all $\tilde{\boldsymbol{\theta}}_i$ and for all $\tilde{\mathbf{a}}_j$,

$$\|F_{N,J}(\tilde{\boldsymbol{\theta}}_i) - \boldsymbol{\theta}_i^+\| \leq 5M$$

and

$$\|F_{N,J}(\tilde{\mathbf{a}}_j) - \mathbf{a}_j^+\| \leq 5M.$$

Combining this with Lemma B.5, we have

$$
\begin{aligned}
&\min_{F\in\mathcal{A}_{K_+}}\left\{\frac{\sum_{i=1}^{N}\|F(\tilde{\boldsymbol{\theta}}_i)-\boldsymbol{\theta}_i^{+}\|^2}{N}+\frac{\sum_{j=1}^{J}\|F(\tilde{\mathbf{a}}_j)-\mathbf{a}_j^{+}\|^2}{J}\right\}\\
\leq&\frac{\sum_{i=1}^{N}\|F_{N,J}(\tilde{\boldsymbol{\theta}}_i)-\boldsymbol{\theta}_i^{+}\|^2}{N}+\frac{\sum_{j=1}^{J}\|F_{N,J}(\tilde{\mathbf{a}}_j)-\mathbf{a}_j^{+}\|^2}{J} \qquad\qquad (B.4)\\
\leq&\left(25(M)^2\lambda_{1,N,J}+\kappa^2\epsilon^2\right)+\left(25(M)^2\lambda_{2,N,J}+\kappa^2\epsilon^2\right)\\
\leq&25(M)^2(\lambda_{1,N,J}+\lambda_{2,N,J})+2\kappa^2\epsilon^2
\end{aligned}
$$

By (B.3), (5) holds. (6) holds if ϵ can be arbitrarily small. We complete the proof. ■

Proof of Theorem 2. Combining Theorem 1 and Proposition 3, we have the result. ■

Proof of Theorem 3. Theorem 3 is a special case of Proposition 5. See the proof of Proposition 5. ■

Proof of Theorem A.1. For simplicity of writing, we suppose $K_+ = K$ in this proof. We only prove the result for the respondents. The proof for the items is the same. Under the conditions of Theorem A.1, the result of Theorem 2 is satisfied and with a slight change in the proof, we can get

$$\max_{F\in\mathcal{A}_K}\frac{\sum_{i=1}^{N}\|\hat{\boldsymbol{\theta}}_i-F(\mathbf{b}^*_{\vartheta_i^*})\|^2}{N}=o_p(1).$$

Consequently, there exists isometry $F^*_{N,J}$, such that

$$\frac{\sum_{i=1}^{N}\|\hat{\boldsymbol{\theta}}_i-F^*_{N,J}(\mathbf{b}^*_{\vartheta_i^*})\|^2}{N}=o_p(1), \qquad (B.5)$$

noting that $\mathbf{b}^*_{\vartheta_i^*}=\boldsymbol{\theta}_i^*$.

Lemma B.6 *Under the same conditions as Theorem A.1, suppose that*

$$\frac{\sum_{i=1}^{N}\|\hat{\boldsymbol{\theta}}_i-F^*_{N,J}(\mathbf{b}^*_{\vartheta_i^*})\|^2}{N}=o_p(1).$$

Then we have

$$\max_{\zeta \in \mathcal{B}_{k_1}} \frac{\sum_{i=1}^{N} 1_{\{\vartheta_i^* = \zeta(\hat{\vartheta}_i)\}}}{N} = o_p(1).$$

With Lemma B.6, we complete the proof for the respondents. ∎

B.3 Proof of Propositions

Proof of Proposition 1. It suffices to prove in the case when $\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$ and $\sum_{i=1}^n \mathbf{y}_i = \mathbf{0}$. Denote $D = (d_{ij})_{n\times n}$, where $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2 = \|\mathbf{y}_i - \mathbf{y}_j\|^2$ and let $B = (b_{ij})_{n\times n} = -\frac{1}{2}JDJ$, where $J = I_n - 1_n 1_n^\top / n$. Then B is inner product matrix of both $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ and $\{\mathbf{y}_1, ..., \mathbf{y}_n\}$. That is, $b_{ij} = \mathbf{x}_i^\top \mathbf{x}_j = \mathbf{y}_i \mathbf{y}_j^\top$, for $1 \le i, j \le n$. We refer readers to Critchley (1988) for the relation between inner product matrix and distance matrix. So if we denote

$$P_1 = (\mathbf{x}_1, ..., \mathbf{x}_n)^\top, \quad P_2 = (\mathbf{y}_1, ..., \mathbf{y}_n)^\top,$$

then we have

$$P_1 P_1^\top = P_2 P_2^\top = B.$$

Let

$$P_1^\top = Q_1 R_1, \quad P_2^\top = Q_2 R_2$$

be the QR decomposition (see Cheney and Kincaid (2009)) of P_1, P_2, where Q_1, Q_2 are $k \times k$ orthogonal matrix and R_1, R_2 are $k \times n$ upper-triangular matrix with non-negative diagonal entries. Since $\mathbf{x}_i^\top \mathbf{x}_j = \mathbf{y}_i^\top \mathbf{y}_j$, for $1 \le i, j \le n$, it is not difficult to check that $R_1 = R_2$. If we define $O = Q_2 Q_1^\top$, then

$$OP_1^\top = OQ_1R_1 = Q_2Q_1^\top Q_1 R_1 = Q_2 R_1 = Q_2 R_2 = P_2^\top,$$

which means $O\mathbf{x}_i = \mathbf{y}_i$, for $1 \le i \le n$. We complete the proof. ∎

Proof of Proposition 2. We first introduce a lemma as following.

Lemma B.7 *There exists a collection of anchor points* $\{\mathbf{b}_1^*, ..., \mathbf{b}_{k_2}^*\}, \{\mathbf{c}_1^*, ..., \mathbf{c}_{k_2}^*\} \subset$ *int*(G)*, where* G *is the ball defined in Proposition 2.*

We fix such collection of anchor points. For any $\epsilon > 0$, we denote $B_k^*(\epsilon), G_l^*(\epsilon)$, for $1 \leq k \leq k_1$ and $1 \leq l \leq k_2$, as balls centered at $\mathbf{b}_k^*$ and $\mathbf{c}_l^*$, respectively. For sufficiently small $\epsilon > 0$, it is easy to see that for any

$$\mathbf{b}_1 \in B_1^*(\epsilon), ..., \mathbf{b}_{k_1} \in B_{k_1}^*(\epsilon), \mathbf{c}_1 \in G_1^*(\epsilon), ..., \mathbf{c}_{k_2} \in G_{k_2}^*(\epsilon),$$

the $\{\mathbf{b}_1, ..., \mathbf{b}_{k_1}\}, \{\mathbf{c}_1, ..., \mathbf{c}_{k_2}\}$ are a collection of anchor points in $\mathbb{R}^K$. Therefore, the (1) of A3 holds. We define

$$\beta_\epsilon := \frac{1}{2} \min_{\substack{1 \leq k \leq k_1 \\ 1 \leq l \leq k_2}} \{P_1 B_k^*(\epsilon), P_2 G_l^*(\epsilon)\}$$

and use $\mathcal{A}_{N,J}$ to denote the following event

$$\begin{aligned} \left| \frac{1}{N} \sum_{i=1}^{N} 1_{\{\boldsymbol{\theta}_i^* \in B_k^*(\epsilon)\}} - P_1 B_k^*(\epsilon) \right| &\leq \beta_\epsilon, \quad k = 1, ..., k_1, \\ \left| \frac{1}{J} \sum_{j=1}^{J} 1_{\{\mathbf{a}_j^* \in G_l^*(\epsilon)\}} - P_2 G_l^*(\epsilon) \right| &\leq \beta_\epsilon, \quad l = 1, ..., k_2, \end{aligned} \tag{B.6}$$

where $P_1 B_k^*(\epsilon), P_2 G_l^*(\epsilon)$ represent the probability measure of $B_k^*(\epsilon), G_l^*(\epsilon)$ with respect to P_1 and P_2, respectively. By Hoeffding's inequality, we have

$$\Pr\left((\text{B.6}) \text{ holds }\right) \geq 1 - 2k_1 \exp(-\frac{1}{2} N \beta_\epsilon^2) - 2k_2 \exp(-\frac{1}{2} J \beta_\epsilon^2). \tag{B.7}$$

So we have

$$\Pr(\mathcal{A}_{N,J}) \to 1$$

as N, J grow. On $\mathcal{A}_{N,J}$, we have

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} 1_{\{\boldsymbol{\theta}_i^* \in B_k^*(\epsilon)\}} \geq \beta_\epsilon, \quad 1 \leq k \leq k_1, \\
&\frac{1}{J}\sum_{j=1}^{J} 1_{\{\mathbf{a}_j^* \in G_l^*(\epsilon)\}} \geq \beta_\epsilon, \quad 1 \leq l \leq k_2.
\end{aligned} \tag{B.8}$$

On $\mathcal{A}_{N,J}$, (B.8) holds. Then, the (2) of A3 holds almost surely. ■

Proof of Proposition 3. Proposition 3 is a special case of Proposition 4. See the proof of Proposition 4. ■

Proof of Proposition 4. The proof of Proposition 4 is similar to Theorem 1 of Davenport et al. (2014). We only state the main steps.

We denote D as the partial distance matrix of $(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_N)$ and $(\mathbf{a}_1, ..., \mathbf{a}_J)$ (to simplify the notation, we ignore the subscripts N and J for D). Since the likelihood function depends on $(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_N)$ and $(\mathbf{a}_1, ..., \mathbf{a}_J)$ only through their partial distance matrix, we re-parameterize the likelihood function by D. We denote

$$l_{\Omega,Y}(D) = \log L^{\Omega}(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_N, \mathbf{a}_1, ..., \mathbf{a}_J),$$

where the subscripts $\Omega = (\omega_{ij})_{N\times J}$ and $Y = (Y_{ij})_{N\times J}$ indicate the random variables in the likelihood function and D contains the parameters.

Let

$$\bar{l}_{\Omega,Y}(D) = l_{\Omega,Y}(D) - l_{\Omega,Y}(\mathbf{0}), \tag{B.9}$$

where $\mathbf{0}$ represents an $N \times J$ matrix whose elements are all 0 and let

$$G = \left\{ D \in \mathbb{R}^{N\times J} : \|D\|_* \leq 4M^2\sqrt{(K_+ + 2)NJ} \right\}. \tag{B.10}$$

Lemma B.8 *Under the same conditions as Proposition 4, there exist constant* C_1

and C_2 *such that*

$$\Pr\left(\sup_{D\in G}|\bar{l}_{\Omega,Y}(D)-E\bar{l}_{\Omega,Y}(D)|\geq 4M^2C_1L_{4M^2}\sqrt{K_++2}\sqrt{n(N+J)+NJ\log(NJ)}\right)$$
$$\leq\frac{C_2}{N+J}.$$

Let $H=\{D:d_{ij}=\|\boldsymbol{\theta}_i-\mathbf{a}_j\|^2,\ \text{where }\|\boldsymbol{\theta}_i\|,\|\mathbf{a}_j\|\leq M, i=1,...,N, j=1,...,J\}$. It is easy to check that $H\subset G$. Consequently,

$$\begin{aligned}
&\Pr\left(\sup_{D\in H}|\bar{l}_{\Omega,Y}(D)-E\bar{l}_{\Omega,Y}(D)|\geq 4C_1M^2L_{4M^2}\sqrt{K_++2}\sqrt{n(N+J)+NJ\log(NJ)}\right)\\
\leq\ &\Pr\left(\sup_{D\in G}|\bar{l}_{\Omega,Y}(D)-E\bar{l}_{\Omega,Y}(D)|\geq 4C_1M^2L_{4M^2}\sqrt{K_++2}\sqrt{n(N+J)+NJ\log(NJ)}\right)\\
\leq\ &\frac{C_2}{N+J}.
\end{aligned}$$

Given the above development, Proposition 4 is implied by the following lemma.

Lemma B.9 *Under the same conditions as Proposition 4,*

$$\frac{1}{NJ}\|D^*_{N,J}-\hat{D}_{N,J}\|_F^2\leq\frac{16}{n}\beta_{4M^2}\sup_{D\in H}|\bar{l}_{\Omega,Y}(D)-E\bar{l}_{\Omega,Y}(D)|.$$

Therefore, with probability at least $1-C_2/(N+J)$,

$$\frac{1}{NJ}\|D^*_{N,J}-\hat{D}_{N,J}\|_F^2\leq 64C_1M^2L_{4M^2}\beta_{4M^2}\sqrt{K_++2}\sqrt{\frac{N+J}{n}}\sqrt{1+\frac{NJ\log(NJ)}{n(N+J)}}.$$

We complete the proof by absorbing $64\sqrt{K_++2}$ into C_1. ■

Proof of Proposition 5. We use $\mathcal{A}_{N,J}$ to denote the event that the result in Proposition 4 holds. By Theorem 1 and Proposition 2, on $\mathcal{A}_{N,J}$, we have

$$\min_{F\in\mathcal{A}_{K_+}}\frac{\sum_{i=1}^N\|\boldsymbol{\theta}_i^+-F(\hat{\boldsymbol{\theta}}_i^\Omega)\|^2}{N}+\frac{\sum_{j=1}^J\|\mathbf{a}_j^+-F(\hat{\mathbf{a}}_j^\Omega)\|^2}{J}$$

goes to 0, as N, J grow to infinity. Since $\Pr(\mathcal{A}_{N,J}) \to 0$, we complete the proof. ■

B.4 Proof of Lemmas

Proof of Lemma B.1. Let $\tau_1 = [\mathbf{x}_1, ..., \mathbf{x}_n], \tau_2 = [\mathbf{y}_1, ..., \mathbf{y}_n], \tau_3 = [\mathbf{z}_1, ..., \mathbf{z}_n]$. Define

$$\tilde{d}(\tau_1, \tau_2) := \min_{F \in \mathcal{A}_K} \max_i \|F(\mathbf{y}_i) - \mathbf{x}_i\|$$

and it is easy to check that

$$\tilde{d}(\tau_1, \tau_2) \leq d(\tau_1, \tau_2) \leq \sqrt{n}\tilde{d}(\tau_1, \tau_2).$$

So we just need to verify that function $\tilde{d}(\cdot, \cdot)$ satisfies the triangle inequality. Let isometries F_{21}, F_{31} satisfy

$$\tilde{d}(\tau_1, \tau_2) = \max_i \|F_{21}(\mathbf{y}_i) - \mathbf{x}_i\| = \|F_{21}(\mathbf{y}_l) - \mathbf{x}_l\|,$$
$$\tilde{d}(\tau_1, \tau_3) = \max_i \|F_{31}(\mathbf{z}_i) - \mathbf{x}_i\| = \|F_{31}(\mathbf{z}_m) - \mathbf{x}_m\|.$$

Then

$$\begin{aligned}
\tilde{d}(\tau_2, \tau_3) &\leq \max_i \{\|F_{31}(\mathbf{z}_i) - F_{21}(\mathbf{y}_i)\|\} \\
&\leq \max_i \{\|F_{31}(\mathbf{z}_i) - \mathbf{x}_i\| + \|F_{21}(\mathbf{y}_i) - \mathbf{x}_i\|\} \\
&\leq \|F_{21}(\mathbf{y}_l) - \mathbf{x}_l\| + \|F_{31}(\mathbf{z}_m) - \mathbf{x}_m\| \\
&= \tilde{d}(\tau_1, \tau_2) + \tilde{d}(\tau_1, \tau_3).
\end{aligned}$$

We complete the proof. ■

Proof of Lemma B.2. Otherwise there exist $\epsilon_0 > 0$ and sequences $\{\tau_1^{(n)}\}_{n=1}^{\infty} \subset$

$\mathcal{H}_{k_1+k_2,K_+,M}$, and $\{\tau_2^{(n)}\}_{n=1}^{\infty} \subset \mathcal{B}$ such that

$$\left\| \Phi_{k_1,k_2,K_+}(\tau_1^{(n)}) - \Phi_{k_1,k_2,K_+}(\tau_2^{(n)}) \right\|_F < \frac{1}{n}$$

and

$$d(\tau_1^{(n)}, \tau_2^{(n)}) > \epsilon_0.$$

Since both $\mathcal{H}_{k_1+k_2,K_+,M}$ and $\mathcal{B}$ are compact, there exists a subsequence $\{n_k\}_{k=1}^{\infty} \subset \mathbb{N}^+$, such that $\lim_{k\to\infty} \tau_1^{(n_k)} = \tilde{\tau} \in \mathcal{H}_{k_1+k_2,K_+,M}$ and $\lim_{k\to\infty} \tau_2^{(n_k)} = \tau_0 \in \mathcal{B}$. The two configurations $\tilde{\tau}$ and τ_0 have the same partial distance matrix but $d(\tilde{\tau}, \tau_0) > \epsilon_0$. This makes a contradiction because $\tau_0 \in \mathcal{B}$ is the only configuration of its partial distance matrix, by the requirement of $\mathcal{B}$. ■

Proof of Lemma B.3. For a collection of points $\{\mathbf{x}, \mathbf{b}_1^*, ..., \mathbf{b}_{k_1}^*\}, \{\mathbf{c}_1^*, ..., \mathbf{c}_{k_2}^*\}$, it is not difficult to verify that condition A2 holds. So we only need to verify A1.

To verify A1, it suffices to show that if $\{\mathbf{b}_1, ..., \mathbf{b}_{k_1}\}, \{\mathbf{c}_1, ..., \mathbf{c}_{k_2}\}$ is a collection of anchor points, then for any $\mathbf{x} \in B_{\mathbf{0}}^K(C)$, $[\mathbf{x}, \mathbf{b}_1, ..., \mathbf{b}_{k_1}, \mathbf{c}_1, ..., \mathbf{c}_{k_2}]$ is the unique configuration corresponding to its $(k_1+1) \times k_2$ partial distance matrix.

Suppose that $\tau = [\mathbf{x}, \mathbf{b}_1, ..., \mathbf{b}_{k_1}, \mathbf{c}_1, ...\mathbf{c}_{k_2}]$ and $\tau' = [\mathbf{x}', \mathbf{b}_1', ..., \mathbf{b}_{k_1}', \mathbf{c}_1', ...\mathbf{c}_{k_2}']$ satisfy

$$\Phi_{k_1+1,k_2,K}(\tau) = \Phi_{k_1+1,k_2,K}(\tau').$$

Then

$$\Phi_{k_1,k_2,K}([\mathbf{b}_1, ..., \mathbf{b}_{k_2}, \mathbf{c}_1, ..., \mathbf{c}_{k_2}]) = \Phi_{k_1,k_2,K}([\mathbf{b}_1', ..., \mathbf{b}_{k_1}', \mathbf{c}_1', ...\mathbf{c}_{k_2}']).$$

Since $\{\mathbf{b}_1, ..., \mathbf{b}_{k_1}\}, \{\mathbf{c}_1, ...\mathbf{c}_{k_2}\}$ are a collection of anchor points, then $[\mathbf{b}_1, ..., \mathbf{b}_{k_1}, \mathbf{c}_1, ...\mathbf{c}_{k_2}] = [\mathbf{b}_1', ..., \mathbf{b}_{k_1}', \mathbf{c}_1', ...\mathbf{c}_{k_2}']$. Without loss of generality, we suppose $\mathbf{b}_l = \mathbf{b}_l'$ and $\mathbf{c}_m = \mathbf{c}_m'$. Then, the two configurations, $[\mathbf{x}, \mathbf{c}_1, ..., \mathbf{c}_{k_2}]$ and $[\mathbf{x}, \mathbf{c}_1', ..., \mathbf{c}_{k_2}']$, have the same complete distance matrix, which further leads that

$$[\mathbf{x}, \mathbf{c}_1, ..., \mathbf{c}_{k_2}] = [\mathbf{x}', \mathbf{c}_1, ..., \mathbf{c}_{k_2}].$$

Since $\mathbf{c}_1, ..., \mathbf{c}_{k_2}$ can affine span $\mathbb{R}^K$, it is not difficult to see that $\mathbf{x} = \mathbf{x}'$. Then, we get $\tau = \tau'$, and A1 has been verified. ■

Proof of Lemma B.4. We define

$$\begin{aligned} S^*_{N,J}(\epsilon) &= \left[B^*_1(\epsilon), ..., B^*_{k_1}(\epsilon), G^*_1(\epsilon), ..., G^*_{k_2}(\epsilon)\right] \subset \mathcal{H}_{k_1+k_2,K,M}, \\ \tilde{S}_{N,J}(\epsilon) &= \left[\tilde{B}_1(\epsilon), ..., \tilde{B}_{k_1}(\epsilon), \tilde{G}_1(\epsilon), ..., \tilde{G}_{k_2}(\epsilon)\right] \subset \mathcal{H}_{k_1+k_2,K_+,M}, \end{aligned}$$

where $B^*_k(\epsilon), \tilde{B}_k(\epsilon), G^*_l(\epsilon), \tilde{G}_l(\epsilon)$ are defined in the proof of Theorem 1. Let

$$\sigma_{N,J} := d(\tilde{S}_{N,J}(\epsilon), S^*_{N,J}(\epsilon)) \tag{B.11}$$

By (B.1) and triangle inequality, there exists an iosmetry $F_{N,J} \in \mathcal{A}_{K_+}$, such that for all $\mathbf{x}^*_k \in B^*_k(\epsilon), \mathbf{y}^*_l \in G^*_l(\epsilon), \tilde{\mathbf{x}}_k \in \tilde{B}_k(\epsilon), \tilde{\mathbf{y}}_l \in \tilde{G}_l(\epsilon)$,

$$\begin{aligned} \|F_{N,J}(\tilde{\mathbf{x}}_k) - \mathbf{x}^+_k\| &\leq 4\epsilon + \sigma_{N,J}, \quad 1 \leq k \leq k_1, \\ \|F_{N,J}(\tilde{\mathbf{y}}_l) - \mathbf{y}^+_l\| &\leq 4\epsilon + \sigma_{N,J}, \quad 1 \leq l \leq k_2. \end{aligned} \tag{B.12}$$

In what follows, we will show that $\sigma_{N,J} \leq \epsilon$ for N, J large enough. We first define

$$\gamma_{N,J} = \inf\{\|\Phi_{k_1,k_2,K_+}(\tilde{\tau}) - \Phi_{k_1,k_2,K_+}(\tau^*)\|_F : \tilde{\tau} \in \tilde{S}_{N,J}(\epsilon), \tau^* \in S^*_{N,J}(\epsilon)\} \tag{B.13}$$

and we have

$$\gamma^2_{N,J}(p_\epsilon N)(p_\epsilon J) \leq \|\tilde{D}_{N,J} - D^*_{N,J}\|^2_F = o(NJ),$$

which leads to

$$\gamma_{N,J} = o(1). \tag{B.14}$$

By (B.11), there exist $\tilde{\tau} \in \tilde{S}_{N,J}(\epsilon)$ and $\tau^* \in S^*_{N,J}(\epsilon)$ such that

$$\|\Phi_{k_1,k_2,K_+}(\tilde{\tau}) - \Phi_{k_1,k_2,K_+}(\tau^*)\|_F \leq 2\gamma_{N,J}.$$

Then by (B.11), we have

$$\sigma_{N,J} = d(\tilde{S}_{N,J}(\epsilon), S^*_{N,J}(\epsilon)) \leq d(\tau^*, \tilde{\tau}). \tag{B.15}$$

As shown in the beginning of proof for Theorem 1 and according to Definition 1, the τ^* is the unique configuration corresponding to its $k_1 \times k_2$ partial distance matrix. Since $\tilde{\tau} \in \tilde{S}_{N,J}(\epsilon) \subset \mathcal{H}_{k_1+k_2,K_+,M}$, by Lemma B.2, we know $d(\tau^*, \tilde{\tau}) \to 0$ as N, J grow to infinity, and thus

$$d(\tau^*, \tilde{\tau}) < \epsilon \tag{B.16}$$

for N, J large enough.

Finally, since

$$B^*_k(\epsilon), G^*_l(\epsilon) \subset B^K_{\mathbf{0}}(M), \quad \tilde{B}_k(\epsilon), \tilde{G}_l(\epsilon) \subset B^{K_+}_{\mathbf{0}}(M),$$

we have, for N, J large enough,

$$\|F_{N,J}(\mathbf{x})\| \leq 4M, \text{ for } \mathbf{x} \in B^{K_+}_{\mathbf{0}}(M).$$

To see this, if there exists $\mathbf{x} \in B^{K_+}_{\mathbf{0}}(M)$ such that $\|F_{N,J}(\mathbf{x})\| > 4M$, then by simple geometry,

$$\min_{\mathbf{x} \in B^{K_+}_{\mathbf{0}}(M)} \|F_{N,J}(\mathbf{x}) - \mathbf{x}\| > M.$$

According to (B.12) and (B.16), we will get

$$M < \|F_{N,J}(\tilde{\mathbf{x}}_k) - \mathbf{x}^+_k\| \leq 4\epsilon + \sigma_{N,J} \leq 5\epsilon,$$

which contradicts with the fact that $\epsilon < \frac{1}{10}M \leq \frac{1}{10}M$.

■

Proof of Lemma B.5. Let $\tilde{\mathbf{c}}_1, ..., \tilde{\mathbf{c}}_{k_2}$ denote the centers of $\tilde{G}_1(\epsilon), ..., \tilde{G}_{k_2}(\epsilon)$ and $\tilde{\mathbf{c}}_l^+ = (\tilde{\mathbf{c}}_l^\top, \mathbf{0}^\top)^\top \in \mathbb{R}^{K+}$. We first give the following lemma.

Lemma B.10 *For any*

$$\tau_1 = [\mathbf{x}, \mathbf{x}_1, ..., \mathbf{x}_{k_2}] \in [B_{\mathbf{0}}^K(M), G_1^*(\epsilon), ..., G_{k_2}^*(\epsilon)],$$
$$\tau_2 = [\mathbf{y}, \mathbf{y}_1, ..., \mathbf{y}_{k_2}] \in [B_{\mathbf{0}}^{K+}(M), B_{\tilde{\mathbf{c}}_1^+}(\epsilon), ..., B_{\tilde{\mathbf{c}}_{k_2}^+}(\epsilon)],$$

we have

$$\|\mathbf{x}^+ - \mathbf{y}\| \leq c \max\left\{ d(\tau_1, \tau_2), \sqrt{\sum_{l=1}^{k_2} \|\mathbf{x}_l^+ - \mathbf{y}_l\|^2} \right\},$$

for a constant c, which only depends on the set $\{\mathbf{c}_1^*, ..., \mathbf{c}_{k_2}^*\}$ and M.

Define

$$H_1(\epsilon) := \{i \in [N] : \|F_{N,J}(\tilde{\boldsymbol{\theta}}_i) - \boldsymbol{\theta}_i^+\| > 5\max(c, 1)\sqrt{k_1 + k_2}\epsilon\} \tag{B.17}$$

and

$$H_2(\epsilon) := \{j \in [J] : \|F_{N,J}(\tilde{\mathbf{a}}_j) - \mathbf{a}_j^+\| > 5\max(c, 1)\sqrt{k_1 + k_2}\epsilon\}, \tag{B.18}$$

where c is the constant in Lemma B.10. We set the constant κ in Lemma B.5 to be $5\max(c, 1)\sqrt{k_1 + k_2}\epsilon$. and then we have $|H_1(\epsilon)| = N\lambda_{1,N,J}, |H_2(\epsilon)| = J\lambda_{2,N,J}$. Note that $I_1(\epsilon) \cap H_1(\epsilon) = \emptyset, I_2(\epsilon) \cap H_2(\epsilon) = \emptyset$ for N, J large.

We choose $i_1, ..., i_{k_1} \in I_1(\epsilon)$ and $j_1, ..., j_{k_2} \in I_2(\epsilon)$ such that

$$\boldsymbol{\theta}_{i_k}^* \in B_k^*(\epsilon), \quad \tilde{\boldsymbol{\theta}}_{i_k} \in \tilde{B}_k(\epsilon),$$
$$\mathbf{a}_{j_l}^* \in G_l^*(\epsilon), \quad \tilde{\mathbf{a}}_{j_l} \in \tilde{G}_l(\epsilon)$$

for $1 \leq k \leq k_1$ and $1 \leq l \leq k_2$. For any $i \in H_1(\epsilon)$, we consider the following configurations

$$\tau^* = [\boldsymbol{\theta}_i^*, \boldsymbol{\theta}_{i_1}^*, ..., \boldsymbol{\theta}_{i_{k_1}}^*, \mathbf{a}_{j_1}^*, ..., \mathbf{a}_{j_{k_2}}^*] \in \mathcal{H}_{k_1+k_2+1,K,M},$$
$$\tilde{\tau} = [\tilde{\boldsymbol{\theta}}_i, \tilde{\boldsymbol{\theta}}_{i_1}, ..., \tilde{\boldsymbol{\theta}}_{i_{k_1}}, \tilde{\mathbf{a}}_{j_1}, ..., \tilde{\mathbf{a}}_{j_{k_2}}] \in \mathcal{H}_{k_1+k_2+1,K_+,M}$$

and

$$\tau_1^* = [\boldsymbol{\theta}_i^*, \mathbf{a}_{j_1}^*, ..., \mathbf{a}_{j_{k_2}}^*] \in [B_{\mathbf{0}}^K(M), G_1^*(\epsilon), ..., G_{k_2}^*(\epsilon)],$$
$$\tilde{\tau}_1 = [\tilde{\boldsymbol{\theta}}_i, \tilde{\mathbf{a}}_{j_1}, ..., \tilde{\mathbf{a}}_{j_{k_2}}] \in [B_{\mathbf{0}}^{K_+}(M), \tilde{G}_1(\epsilon), ..., \tilde{G}_{k_2}(\epsilon)].$$

It is obvious that

$$d(\tilde{\tau}, \tau^*) \geq d\left(\tilde{\tau}_1, \tau_1^*\right).$$

By Lemma B.4, we have

$$\sqrt{\sum_{l=1}^{k_2} \|F_{N,J}(\tilde{\mathbf{a}}_{j_l}) - \mathbf{a}_{j_l}^*\|^2} \leq 5\sqrt{k_2}\epsilon \leq 5\sqrt{k_1+k_2}\epsilon.$$

Combining it with (B.17) and Lemma (B.10), we have

$$d(\tilde{\tau}_1, \tau_1^*) > 5\sqrt{k_1+k_2}\epsilon,$$

which leads to

$$d(\tilde{\tau}, \tau^*) > 5\sqrt{k_1+k_2}\epsilon. \tag{B.19}$$

According to Lemma B.3, $\{\boldsymbol{\theta}_i^*, \boldsymbol{\theta}_{i_1}^*, ..., \boldsymbol{\theta}_{i_{k_1}}^*\}, \{\mathbf{a}_{j_1}^*, ..., \mathbf{a}_{j_{k_2}}^*\}$ are a collection of anchor points. Let $\tilde{D}, D \in \mathcal{P}_{k_1+1,k_2,K_+,M}$ be the partial distance matrix of $\tilde{\tau}$ and τ^*, respectively. Combining (B.19) and Lemma B.2, there exists a constant $\delta_\epsilon > 0$ such that

$$\|\tilde{D} - D\|_F \geq \delta_\epsilon. \tag{B.20}$$

For each $i \in H_1(\epsilon)$, we choose $i_1, ..., i_{k_1} \in I_1(\epsilon)$ to form a group $\{i, i_1, ..., i_{k_1}\} \subset [N]$

such that

$$(\boldsymbol{\theta}_i^*, \boldsymbol{\theta}_{i_1}^*, ..., \boldsymbol{\theta}_{i_{k_1}}^*) \in B_{\mathbf{0}}^K(M) \times B_1^*(\epsilon) \times \cdots \times B_{k_1}^*(\epsilon)$$

and

$$(\tilde{\boldsymbol{\theta}}_i, \tilde{\boldsymbol{\theta}}_{i_1}, ..., \tilde{\boldsymbol{\theta}}_{i_{k_1}}) \in B_{\mathbf{0}}^{K_+}(M) \times \tilde{B}_1(\epsilon) \times \cdots \times \tilde{B}_{k_1}(\epsilon).$$

We could find at least $\min\{\lambda_{1,N,J}, p_\epsilon\} \times N$ such groups which are mutually exclusive. We could also find at least $p_\epsilon J$ mutually exclusive groups of $\{j_1, ..., j_{k_2}\} \subset [J]$ such that

$$(\mathbf{a}_{j_1}^*, ..., \mathbf{a}_{j_{k_2}}^*) \in G_1^*(\epsilon) \times \cdots \times G_{k_2}^*(\epsilon)$$

and

$$(\tilde{\mathbf{a}}_{j_1}, ..., \tilde{\mathbf{a}}_{j_{k_2}}) \in \tilde{G}_1(\epsilon) \times \cdots \times \tilde{G}_{k_2}(\epsilon).$$

By (4) and (B.20), we have

$$\min\{\lambda_{1,N,J}, p_\epsilon\} N p_\epsilon J \delta_\epsilon^2 \leq o(NJ).$$

So

$$\min\{\lambda_{1,N,J}, p_\epsilon\} = o(1),$$

which means $\lambda_{1,N,J} \to 0$, as N, J grow to infinity. Similar result holds for $\lambda_{2,N,J}$ and we do not repeat it. ■

Proof of Lemma B.6. Consider the K-means clustering of the person points in Algorithm A.1. We define a loss function

$$\mathcal{L}(\vartheta_1, ..., \vartheta_N) = \frac{1}{N} \sum_{i=1}^{N} \|\hat{\boldsymbol{\theta}}_i - \boldsymbol{\mu}_{\vartheta_i}\|^2,$$

as the loss function for K-means clustering, where $\vartheta_i \in \{1, ..., k_1\}$ represents the cluster membership of person i and

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^N \hat{\boldsymbol{\theta}}_i 1_{\{\vartheta_i = k\}}}{\sum_{i=1}^N 1_{\{\vartheta_i = k\}}}$$

denotes the centroid of the kth cluster. Under the conditions of Theorem A.1, the K-means clustering converges to the global optima, which implies that

$$\mathcal{L}(\hat{\vartheta}_1, ..., \hat{\vartheta}_N) = \min_{\vartheta_i \in \{1,...,k_1\}, i=1,...,N} \mathcal{L}(\vartheta_1, ..., \vartheta_N). \tag{B.21}$$

So for any isometry $F \in \mathcal{A}_K$,

$$\sum_{i=1}^N \|\hat{\boldsymbol{\theta}}_i - \boldsymbol{\mu}_{\hat{\vartheta}_i}\|^2 \leq \sum_{i=1}^N \|\hat{\boldsymbol{\theta}}_i - F(\mathbf{b}^*_{\vartheta_i^*})\|^2.$$

By triangle inequality,

$$\begin{aligned}\left(\sum_{i=1}^N \|\boldsymbol{\mu}_{\hat{\vartheta}_i} - F(\mathbf{b}^*_{\vartheta_i^*})\|^2\right)^{\frac{1}{2}} &\leq \left(\sum_{i=1}^N (\|\boldsymbol{\mu}_{\hat{\vartheta}_i} - \hat{\boldsymbol{\theta}}_i\|^2\right)^{\frac{1}{2}} + \left(\sum_{i=1}^N \|\hat{\boldsymbol{\theta}}_i - F(\mathbf{b}^*_{\vartheta_i^*})\|^2\right)^{\frac{1}{2}}, \\ &\leq 2\left(\sum_{i=1}^N \|\hat{\boldsymbol{\theta}}_i - F(\mathbf{b}^*_{\vartheta_i^*})\|^2\right)^{\frac{1}{2}}.\end{aligned}$$

Define $d = \min_{i \neq j} \|\mathbf{b}_i^* - \mathbf{b}_j^*\|$ and for $F \in \mathcal{A}_K$, define

$$A_F := \{1 \leq i \leq N : \|\boldsymbol{\mu}_{\hat{\vartheta}_i} - F(\mathbf{b}^*_{\vartheta_i^*})\| < \frac{d}{2}\},$$

and denote $A_F^{\mathsf{c}} := \{1, ..., N\}/A_F$.

Then

$$
\begin{aligned}
\frac{\sum_{i \in A_{F^*_{N,J}}} 1}{N} &= 1 - \frac{\sum_{i \in A^{\mathrm{c}}_{F^*_{N,J}}} 1}{N} \\
&\geq 1 - \frac{4}{d^2} \frac{\sum_{i \in A^{\mathrm{c}}_{F^*_{N,J}}} \|\boldsymbol{\mu}_{\hat{\vartheta}_i} - F^*_{N,J}(\mathbf{b}^*_{\vartheta^*_i})\|^2}{N} \\
&\geq 1 - \frac{4}{d^2} \frac{\sum_{i=1}^{N} \|\boldsymbol{\mu}_{\hat{\vartheta}_i} - F^*_{N,J}(\mathbf{b}^*_{\vartheta^*_i})\|^2}{N} \\
&\geq 1 - \frac{16}{d^2} \frac{\sum_{i=1}^{N} \|\hat{\boldsymbol{\theta}}_i - F^*_{N,J}(\mathbf{b}^*_{\vartheta^*_i})\|^2}{N} \\
&\stackrel{pr}{\to} 1
\end{aligned}
\tag{B.22}
$$

Lemma B.11 *Under the same conditions as Lemma B.6, if there exists $\zeta_1 \in \mathcal{B}_{k_1}$ satisfying*

$$
\|\boldsymbol{\mu}_{\zeta_1(l)} - F^*_{N,J}(\mathbf{b}^*_l)\| < \frac{d}{2},
$$

where $\boldsymbol{\mu}_l$ is the centroid of the lth cluster, $F^*_{N,J}$ is defined in (B.5) *and d is defined above, then there exists $\zeta_2 \in \mathcal{B}_{k_1}$, such that for all $i \in A_{F^*_{N,J}}, \hat{\vartheta}_i = \zeta_2(\vartheta^*_i)$.*

Let $\Omega_{N,J} := \{\omega : \exists \zeta \in \mathcal{B}_{k_1},\ \text{s.t.}\ \|\boldsymbol{\mu}_{\zeta(l)}(\omega) - F^*_{N,J}(\mathbf{b}^*_l)\| < \frac{d}{2},\ i = 1, ..., k_1\}$. Notice that $\Omega_{N,J}$ is a subset of the whole probability space. By Lemma B.11, for any $\omega \in \Omega_{N,J}$, there exists $\zeta_{N,J} \in \mathcal{B}_{k_1}$, which corresponds to ζ_2 in Lemma B.11, such that

$$
\max_{\zeta \in \mathcal{B}_{k_1}} \frac{\sum_{i=1}^{N} 1_{\{\vartheta^*_i = \zeta(\hat{\vartheta}_i(\omega))\}}}{N} \geq \frac{\sum_{i=1}^{N} 1_{\{\hat{\vartheta}_i = \zeta_{N,J}(\vartheta^*_i)\}}}{N} \geq \frac{\sum_{i \in A_{F^*_{N,J}}} 1}{N}
$$

Lemma B.12 *Under the same conditions as Lemma B.6, we have*

$$
\lim_{N,J \to \infty} \Pr\left(\Omega_{N,J}\right) = 1,
$$

where $\Omega_{N,J}$ is defined above.

By Lemma B.12 and (B.22), we complete the proof. ■

Proof of Lemma B.7. Without loss of generality, we suppose that the ball $G \subset \mathbb{R}^K$ has center at orgin. By Theorem 2.1 of Alfakih (2005), we know there exist $k_1, k_2 \geq K+1$ and two sets of points, $\{\mathbf{b}_1^*, ..., \mathbf{b}_{k_1}^*\}, \{\mathbf{c}_1^*, ..., \mathbf{c}_{k_2}^*\} \subset \text{int}(G)$, satisfying condition A2 whose partial distance matrix D^* has unique configuration. Furthermore, points near $\mathbf{b}_i^*, \mathbf{c}_j^*$ also have this property. Specifically, there exists $\epsilon > 0$ such that for

$$\mathbf{b}_i \in B_{\mathbf{b}_i^*}^K(\epsilon) \subset G, \quad \mathbf{c}_j \in B_{\mathbf{c}_j^*}^K(\epsilon) \subset G,$$

the $\{\mathbf{b}_1, ..., \mathbf{b}_{k_1}\}, \{\mathbf{c}_1, ..., \mathbf{c}_{k_2}\}$ satisfy condition A2 and their partial distance matrix D has unique configuration. Then, by Lemma B.2, condition A1 holds and $\{\mathbf{b}_1^*, ..., \mathbf{b}_{k_1}^*\}, \{\mathbf{c}_1^*, ..., \mathbf{c}_{k_2}^*\}$ are anchor points in $\mathbb{R}^K$. ■

Proof of Lemma B.8. The proof of Lemma B.8 is similar to Lemma A.1 of Davenport et al. (2014). ■

Proof of Lemma B.9. We have

$$\begin{aligned} 0 \leq \bar{l}_{\Omega,Y}(\hat{D}_{N,J}) - \bar{l}_{\Omega,Y}(D_{N,J}^*) =& \bar{l}_{\Omega,Y}(\hat{D}_{N,J}) - \mathbb{E}\bar{l}_{\Omega,Y}(\hat{D}_{N,J}) + \mathbb{E}\bar{l}_{\Omega,Y}(\hat{D}_{N,J}) - \mathbb{E}\bar{l}_{\Omega,Y}(D_{N,J}^*) \\ &+ \mathbb{E}\bar{l}_{\Omega,Y}(D_{N,J}^*) - \bar{l}_{\Omega,Y}(D_{N,J}^*) \\ \leq& \left(\mathbb{E}\bar{l}_{\Omega,Y}(\hat{D}_{N,J}) - \mathbb{E}\bar{l}_{\Omega,Y}(D_{N,J}^*)\right) + 2 \sup_{D \in H} |\bar{l}_{\Omega,Y}(D) - \bar{l}_{\Omega,Y}(D)|. \end{aligned}$$

So

$$\mathbb{E}\left(\bar{l}_{\Omega,Y}(D_{N,J}^*) - \bar{l}_{\Omega,Y}(\hat{D}_{N,J})\right) \leq 2 \sup_{D \in H} |\bar{l}_{\Omega,Y}(D) - \bar{l}_{\Omega,Y}(D)|.$$

Notice that

$$\begin{aligned} \mathbb{E}\left(\bar{l}_{\Omega,Y}(D_{N,J}^*) - \bar{l}_{\Omega,Y}(\hat{D}_{N,J})\right) &= \mathbb{E}\left(l_{\Omega,Y}(D_{N,J}^*) - l_{\Omega,Y}(\hat{D}_{N,J})\right) \\ &= \frac{n}{NJ} \sum_{i,j} f(d_{ij}^*) \log(\frac{f(d_{ij}^*)}{f(\hat{d}_{ij})}) + (1 - f(d_{ij}^*)) \log(\frac{1 - f(d_{ij}^*)}{1 - f(\hat{d}_{ij})}) \end{aligned}$$

For two distributions $\mathcal{P}$ and $\mathcal{Q}$, let $D_{KL}(\mathcal{P}\|\mathcal{Q})$ denote the Kullback-Leibler divergence

$$D_{KL}(\mathcal{P}\|\mathcal{Q}) := \int p(x)\log\left(\frac{p(x)}{q(x)}\right)dx,$$

where $p(x)$ and $q(x)$ are the density functions for $\mathcal{P}$ and $\mathcal{Q}$, respectively. For $0 < p, q < 1$, we use

$$D_{KL}(p\|q) := p\log(\frac{p}{q}) + (1-p)\log(\frac{1-p}{1-q})$$

to denote the Kullback-Leibler divergence between two Bernoulli distributions with parameter p and q, respectively. For $P, Q \in (0,1)^{N\times J}$, we define

$$D_{KL}(P\|Q) := \frac{1}{NJ}\sum_{i,j} D_{KL}(P_{ij}\|Q_{ij}).$$

For a partial distance matrix $D_{N,J}$, denote $f(D_{N,J})$ as the matrix $(f(d_{ij}))_{N\times J}$. So from above, we know that

$$nD_{KL}(f(D^*_{N,J})\|f(\hat{D}_{N,J})) \leq 2\sup_{D\in H}|\bar{l}_{\Omega,Y}(D) - \bar{l}_{\Omega,Y}(D)|.$$

Still for $0 < p, q < 1$, let

$$d_H^2(p,q) := (\sqrt{p}-\sqrt{q})^2 + (\sqrt{1-p}-\sqrt{1-q})^2$$

denote the Hellinger distance between two Bernoulli distributions with parameters p and q, respectively. For $P, Q \in (0,1)^{N\times J}$, we define

$$d_H^2(P\|Q) := \frac{1}{NJ}\sum_{i,j} d_H^2(P_{ij}, Q_{ij}).$$

It is easy to check that $d_H^2(p,q) \le D_{KL}(p\|q)$. So

$$d_H^2(f(D_{N,J}^*), f(\hat{D}_{N,J})) \le \frac{2}{n} \sup_{D\in H} |\bar{l}_{\Omega,Y}(D) - \bar{l}_{\Omega,Y}(D)|$$

By Lemma A.2 of Davenport et al. (2014), we have

$$\begin{aligned}\frac{1}{NJ}\|\hat{D}_{N,J} - D_{N,J}^*\|_F^2 &\le 8\beta_{4M^2} d_H^2(f(D_{N,J}^*), f(\hat{D}_{N,J})) \\ &\le \frac{16}{n}\beta_{4M^2} \sup_{D\in H} |\bar{l}_{\Omega,Y}(D) - \bar{l}_{\Omega,Y}(D)|.\end{aligned}$$

■

Proof of Lemma B.10. Denote

$$\eta := \max\left\{ d(\tau_1,\tau_2), \sqrt{\sum_{l=1}^{k_2} \|\mathbf{x}_l^+ - \mathbf{y}_l\|^2} \right\}$$

and then $d(\tau_1,\tau_2) \le \eta$ and

$$\|\mathbf{x}_l^+ - \mathbf{y}_l\| \le \eta, \quad l = 1, ..., k_2. \tag{B.23}$$

Therefore there exist $A \in \mathcal{O}_{K_+}$ and $\mathbf{b} \in \mathbb{R}^{K_+}$ such that

$$\sqrt{\|A\mathbf{x}^+ + \mathbf{b} - \mathbf{y}\|^2 + \sum_{l=1}^{k_2} \|A\mathbf{x}_l^+ + \mathbf{b} - \mathbf{y}_l\|^2} \le \eta,$$

which leads that

$$\|A\mathbf{x}^+ + \mathbf{b} - \mathbf{y}\| \le \eta \tag{B.24}$$

and

$$\|A\mathbf{x}_l^+ + \mathbf{b} - \mathbf{y}_l\| \le \eta, \quad l = 1, ..., k_2. \tag{B.25}$$

Combining (B.23) and (B.25), we get

$$\|A\mathbf{x}_l^+ + \mathbf{b} - \mathbf{x}_l^+\| \leq 2\eta, \quad l = 1, ..., k_2.$$

According to condition A2, $\mathbf{x}_1, ..., \mathbf{x}_{k_2}$ can affine span $\mathbb{R}^K$. Then there exists $\alpha_1, ..., \alpha_{k_1}$ satisfying $\sum_{l=1}^{k_2} \alpha_l = 1$, such that $\mathbf{x} = \sum_{l=1}^{k_2} \alpha_l \mathbf{x}_l$. So we have

$$\|A\mathbf{x}^+ + \mathbf{b} - \mathbf{x}^+\| = \left\| \sum_{l=1}^{k_2} \alpha_l (A\mathbf{x}_l^+ + \mathbf{b} - \mathbf{x}_l^+) \right\| \leq 2(\sum_{j=1}^{k_2} |\alpha_j|)\eta.$$

Combining it with (B.24), we have $\|\mathbf{x}^+ - \mathbf{y}\| \leq (2 \sum_{j=1}^{k_1} |\alpha_j| + 1)\eta$. We complete the proof by setting the constant c in Lemma B.10 to be

$$\max_{\substack{\mathbf{x} \in B_{\mathbf{0}}^K(M) \\ \mathbf{x}_l \in G_l^*(\epsilon)}} \inf_{\substack{\sum_l \alpha_l = 1 \\ \mathbf{x} = \sum_l \alpha_l \mathbf{x}_l}} 2 \sum_{l=1}^{k_2} |\alpha_l|.$$

■

Proof of Lemma B.11. For $i, j \in A_{F_{N,J}^*}$, if $\vartheta_i^* = \vartheta_j^*$ and suppose they are both equal to k, then $\|\boldsymbol{\mu}_{\hat{\vartheta}_i} - F_{N,J}^*(\mathbf{b}_k^*)\| < d/2$ and $\|\boldsymbol{\mu}_{\hat{\vartheta}_j} - F_{N,J}^*(\mathbf{b}_k^*)\| < d/2$. Given the condition in Lemma B.11, it is easy to check that there is only one $\boldsymbol{\mu}_l$ among $\{\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_{k_1}\}$ satisfying

$$\|\boldsymbol{\mu}_l - F_{N,J}^*(\mathbf{b}_k^*)\| < d/2,$$

then $\hat{\vartheta}_i = \hat{\vartheta}_j$. If $\vartheta_i^* \neq \vartheta_j^*$, then $\|\boldsymbol{\mu}_{\hat{\vartheta}_i} - F_{N,J}^*(\mathbf{b}_{\vartheta_i^*}^*)\| < d/2$ and $\|\boldsymbol{\mu}_{\hat{\vartheta}_j} - F_{N,J}^*(\mathbf{b}_{\hat{\vartheta}_j}^*)\| < d/2$. So

$$\begin{aligned}\|\boldsymbol{\mu}_{\hat{\vartheta}_i} - \boldsymbol{\mu}_{\hat{\vartheta}_j}\| &\geq \|F_{N,J}^*(\mathbf{b}_{\vartheta_i^*}^*) - F_{N,J}^*(\mathbf{b}_{\vartheta_j^*}^*)\| - \|\boldsymbol{\mu}_{\hat{\vartheta}_i} - F_{N,J}^*(\mathbf{b}_{\vartheta_i^*}^*)\| - \|\boldsymbol{\mu}_{\hat{\vartheta}_j} - F_{N,J}^*(\mathbf{b}_{\hat{\vartheta}_j}^*)\| \\ &> d - \frac{d}{2} - \frac{d}{2} > 0,\end{aligned}$$

which means $\hat{\vartheta}_i \neq \hat{\vartheta}_j$. So there exists ζ_2 such that for $i \in A_{F_{N,J}^*}, \hat{\vartheta}_i = \zeta_2(\vartheta_i^*)$. ■

Proof of Lemma B.12. Let

$$\Gamma_{N,J}^{(\epsilon')} := \{\omega : \frac{\sum_{i \in A_{F^*_{N,J}}} 1}{N} \geq 1 - \epsilon'\},$$

which is a subset of the whole probability space. By (B.22), for any $\epsilon' > 0$, we have

$$\lim_{N,J \to \infty} \Pr(\Gamma_{N,J}^{(\epsilon')}) = 1.$$

For any $\omega \notin \Omega_{N,J}$, there exists l such that $\|\boldsymbol{\mu}_m(\omega) - F^*_{N,J}(\mathbf{b}_l^*)\| \geq d/2$, for $m = 1, ..., k_1$. So for i satisfying $\vartheta_i^* = l$, we have $\|\boldsymbol{\mu}_{\hat{\vartheta}_i}(\omega) - F^*_{N,J}(\mathbf{b}^*_{\vartheta_i^*})\| \geq d/2$. According to (2) of condition A3, for sufficiently small ϵ', if N, J are sufficiently large, then $\omega \notin \Gamma_{N,J}^{(\epsilon')}$, which means $\Gamma_{N,J}^{(\epsilon')} \subset \Omega_{N,J}$. By (B.22), for sufficiently small ϵ',

$$\lim_{N,J \to \infty} \Pr(\Omega_{N,J}) \geq \lim_{N,J \to \infty} \Pr(\Gamma_{N,J}^{(\epsilon')}) = 1.$$

We complete the proof. ■

C Algorithm-based MDU: Real Data Examples

To compare the proposed method with classical algorithm-based MDU methods, we apply ordinal MDU (Busing et al., 2005) to both real datasets analyzed in the paper. The application is based on the implementation in R package *smacof* (de Leeuw and Mair, 2009). For both examples, the latent dimension is set to two, and all the tuning parameters are set to be the default ones. The results below show that the ordinal MDU approach provides similar visualization results as the proposed one, especially for the roll call voting data due to its unidimensional nature. The results for the movie rating dataset are also similar for the two methods, but the interpretable patterns from the ordinal MDU approach is not as clear as the proposed one.

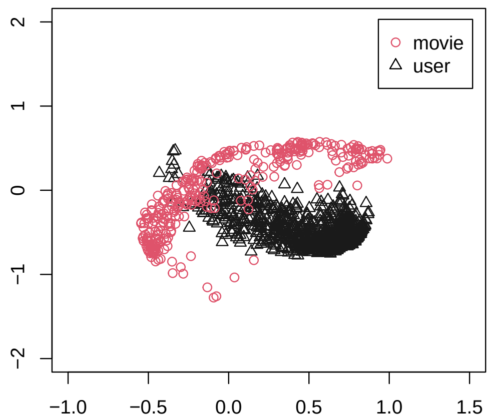
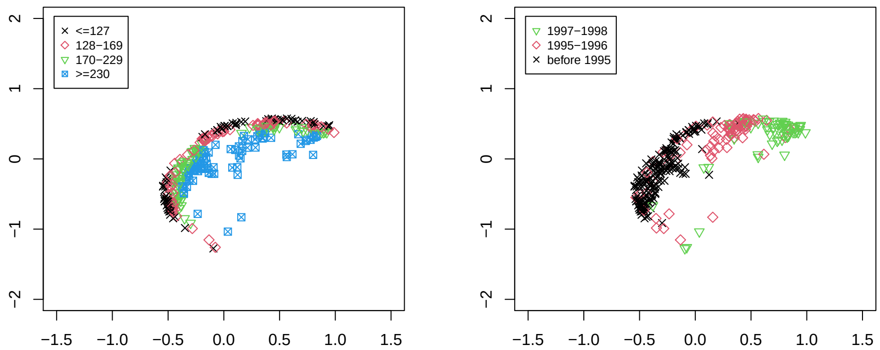


Figure C.1: Analysis of movie rating data: Simultaneous visualization of the estimated movie and user points.

Figures C.1 through C.3 show the same plots as in Figures 4 through 6 in Section 5.1, respectively, for the movie rating dataset. Figure C.1 provides the simultaneous visualization of the movie and user points. Similar to the plot in Figure 4 given by our method, the movies and the users tend to form two giant clusters that only slightly overlap.

Figure C.2 is similar to Figure 5, where the two panels show the same scatter plot for the movie points. In the left panel, the movies are stratified by the the numbers of ratings that they received, where different stratums are marked by different colors. In the right panel, the movies are stratified by their release time. Recall that the patterns of popularity and release time are captured by the proposed method as shown in Figure 5. Figure C.2 seems also to capture these patterns, but not as clear as those in Figure 5. According to panel (a) of Figure C.2, the more popular movies tend to be located near the origin, while the less popular movies tend to be located away from the origin. According to panel (b) of Figure C.2, the clustering patten of the movies can be largely explained by the three categories of release dates. From the left to the right of the space, the points correspond to movies from the relatively



(a) (b)

Figure C.2: Analysis of movie rating data. Panel (a): Visualization of movie points, with movies stratified into four equal-size categories based on the numbers of rating. Movies with numbers of rating less than 127, 128-169, 170-229 and more than 230 are indicated by black, red, green and blue points, respectively. Panel (b): Visualization of movie points, with movies stratified into three categories based on their release time. Movies released in 1997-1998, 1995-1996, and before 1995 are indicated by green, red and black points, respectively.

older ones to the relatively more recent ones.

Figure C.3 shows the same plots as in Figure 6. Similar pattern is shown that the shorter the average distance from a user point to the movies points, the more active the user is. In Figure C.3, users are classified into four equal-size groups depending on the numbers of movies they rated. These groups of users, from the most active one to the least active one, lie from the top left to the bottom right.

Figures C.4 through C.6 show the same plots as in Figures 7 through 9 in Section 5.2, respectively, for the roll call voting dataset. Figure C.4 provides the simultaneous visualization of senators and roll calls. Similar to the plot in Figure 7, most of the points tend to lie on a straight line.

Figure C.5 provides a scatter plot of the senator points. Similar to Figure 8, most of the senator points tend to locate around a straight line, with the Democrats on one side and the Republicans on the other side. Also similar to Figure 8, the

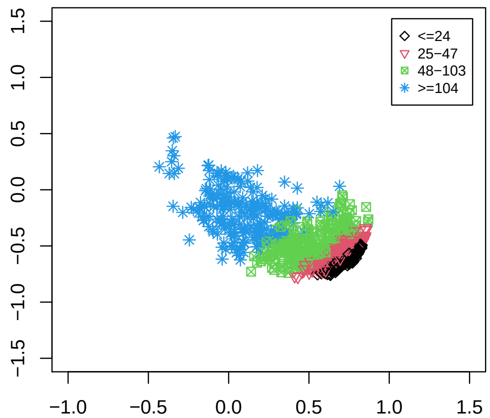


Figure C.3: Analysis of movie rating data: Visualization of user points, with users classified into four equal-size categories based on the numbers of rating. Users who rated less than 24, , 25-47, 48-103 and more than 104 movies are indicated by black, red, green and blue points, respectively.

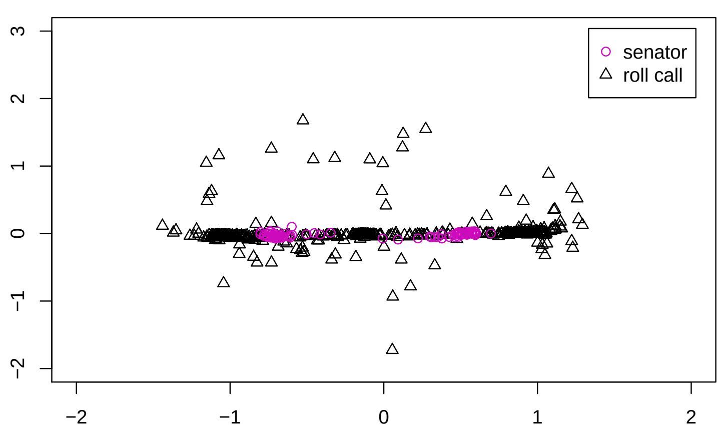


Figure C.4: Analysis of senator roll call data: Simultaneous visualization of the estimated senator and roll call ideal points.

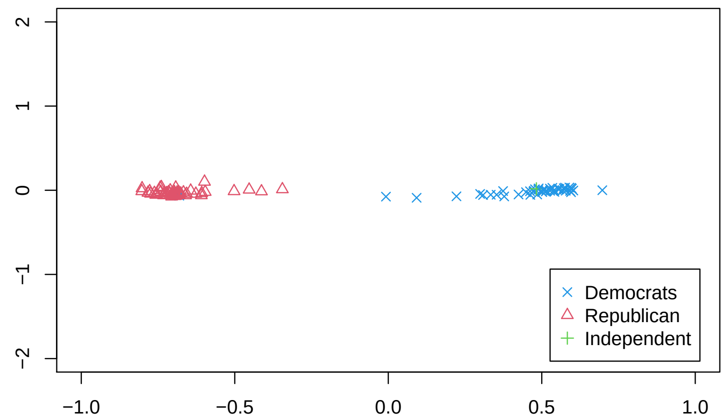
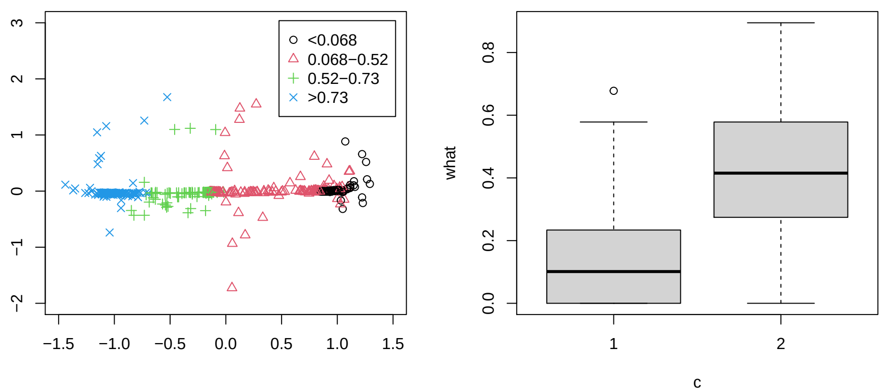


Figure C.5: Analysis of senator roll call data: Visualization of senator points, where senators are classified by their party membership. Specifically, The Democrats, Republicans and an independent politician are indicated by blue, red, and green, respectively.

independent senator, Jim Jeffords from the state of Vermont, is mixed together with the Democrats, while the Democrat senator, Zell Miller from the state of Georgia, is mixed together with the Republicans.

Finally, Figure C.6 shows the unfolding results for the roll calls. The pattern in panel (a) of Figure C.6 is similar to that of Figure 9, where from the right to the left, the proportion of "Yeas" from the Republicans increases. Also similar to Figure 9, although most of the roll calls lie near the x-axis, there are still quite a few of them spreading out along the y-axis. According to panel (b) of Figure C.6 based on the cross entropy measure, the voting behavior on these roll calls tends to be heterogeneous within both parties. This result is similar to that given in panel (b) of Figure 9.



(a) (b)

Figure C.6: Analysis of senator roll call data. Panel (a): Visualization of roll call points, where roll calls are classified by the proportion of Yeas from Republicans. Specifically, roll calls who have the proportions less than 0.068, 0.068-0.52,0.52-0.73 and larger than 0.73 are indicated by black, red, green and blue points, respectively. Panel (b): Box plots of $\min\{\mathrm{CE}_j^{(1)}, \mathrm{CE}_j^{(2)}\}$, for roll calls lying near the x-axis ($|\hat{a}_{j2}| \leq 0.05$) one the left and for those spreading out along the y-axis ($|\hat{a}_{j2}| > 0.05$) on the right.

References

Alfakih, A. Y. (2005). On the uniqueness of euclidean distance matrix completions: the case of points in general position. *Linear Algebra and its Applications*, 397:265–277.

Borg, I. and Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications*. Springer, New York, NY.

Busing, F. M., Groenen, P. J., and Heiser, W. J. (2005). Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*, 70:71–98.

Cheney, W. and Kincaid, D. (2009). Linear algebra: Theory and applications. *The Australian Mathematical Society*, 110:544–550.

Choi, D. and Wolfe, P. J. (2014). Co-clustering separately exchangeable network data. *The Annals of Statistics*, 42:29–63.

Critchley, F. (1988). On certain linear mappings between inner-product and squared-distance matrices. *Linear Algebra and its Applications*, 105:91–107.

Davenport, M. A., Plan, Y., van den Berg, E., and Wootters, M. (2014). 1-bit matrix completion. *Information and Inference: A Journal of the IMA*, 3:189–223.

de Leeuw, J. and Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, 31:1–30.

Dhillon, I. S. (2001). Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 269–274.

Hartigan, J. A. (1972). Direct clustering of a data matrix. *Journal of the American Statistical Association*, 67:123–129.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, 5:109–137.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional scaling*. Sage, Beverly Hills, CA.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Cam, L. M. L. and Neyman, J., editors, *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, pages 281–297. University of California Press, Berkeley, CA.

Rohe, K., Qin, T., and Yu, B. (2016). Co-clustering directed graphs to discover asymmetries and directional communities. *Proceedings of the National Academy of Sciences*, 113:12679–12684.