

Supplementary Material for “Delaunay Weighted Two-sample Test for High-dimensional Data by Incorporating Geometric Information”

Jiaqi Gu

Department of Mathematics and Statistics, University of South Florida
and

Ruoxu Tan

School of Mathematical Sciences and
School of Economics and Management, Tongji University
and

Guosheng Yin

Department of Statistics and Actuarial Science, University of Hong Kong

S1 Proof of Uniqueness of the Delaunay Weight Vector

Given that $\mathbb{Z}$ is generic, the Delaunay triangulation $\mathcal{DT}_{\mathcal{M}}(\mathbb{Z})$ is unique ([Leibon and Letscher, 2000](#)). Thus, for each $\mathbf{p} \in \mathcal{M}$, the Delaunay simplex $\mathcal{S}(\mathbf{p}; \mathbb{Z})$ containing $\mathbf{p}^*$, the projection of $\mathbf{p}$ on the convex hull $\mathcal{H}_{\mathcal{M}}(\mathbb{Z})$, is unique. By condition (a) of Definition 2, the number of nonzero components of $\mathbf{w}(\mathbf{p}; \mathbb{Z})$ is not larger than $d+1$. Without loss of generality, let $\mathbb{V}(\mathcal{S}(\mathbf{p}; \mathbb{Z})) = \{\mathbf{z}_1, \dots, \mathbf{z}_{d+1}\}$. As a result, conditions (b) and (c[†]) of Definition 2 imply that $(w_1(\mathbf{p}; \mathbb{Z}), \dots, w_{d+1}(\mathbf{p}; \mathbb{Z}))^\top$ is the solution of the following equation,

$$\begin{pmatrix} 1 & \cdots & 1 \\ \log_{\mathbf{p}^*}(\mathbf{z}_1) & \cdots & \log_{\mathbf{p}^*}(\mathbf{z}_{d+1}) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{d+1} \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}. \quad (\text{S1})$$

Since the intrinsic dimension of $\mathcal{M}$ is d and $\log_{\mathbf{p}^*}(\mathbf{z}_1), \dots, \log_{\mathbf{p}^*}(\mathbf{z}_{d+1}) \in \mathbf{T}_{\mathbf{p}^*}\mathcal{M}$, which is a d -dimensional Euclidean space, we deduce that $\log_{\mathbf{p}^*}(\mathbf{z}_1), \dots, \log_{\mathbf{p}^*}(\mathbf{z}_{d+1})$ are linearly dependent. Without loss of generality, for $k = 0, \dots, d$, let $\{\log_{\mathbf{p}^*}(\mathbf{z}_1), \dots, \log_{\mathbf{p}^*}(\mathbf{z}_{k+1})\}$ be the smallest subset of linearly dependent log maps.

When $k = 0$ (i.e., $\log_{\mathbf{p}^*}(\mathbf{z}_1)$ itself is linearly dependent), we have $\log_{\mathbf{p}^*}(\mathbf{z}_1) = 0$ and thus (S1) has a unique solution $(1, 0, \dots, 0)^\top$.

When $k = 1, \dots, d-1$, we have $\mathbf{p}^*$ being in the k -simplex formed by $\mathbf{z}_1, \dots, \mathbf{z}_{k+1}$. By the convexity of $\mathcal{S}(\mathbf{p}; \mathbb{Z})$, all points $\mathbf{z}_{k+2}, \dots, \mathbf{z}_{d+1}$ are on the same side of the k -simplex formed by $\mathbf{z}_1, \dots, \mathbf{z}_{k+1}$. In other words, the geodesic path between any pair of points in $\mathbf{z}_{k+2}, \dots, \mathbf{z}_{d+1}$ does not intersect the k -simplex formed by $\mathbf{z}_1, \dots, \mathbf{z}_{k+1}$. Thus, the solution of (S1) has $\beta_{k+2} = \dots = \beta_{d+1} = 0$.

In conclusion, for $k = 1, \dots, d$, the number of solutions of (S1) is the same as that of

$$\begin{pmatrix} 1 & \dots & 1 \\ \log_{\mathbf{p}^*}(\mathbf{z}_1) & \dots & \log_{\mathbf{p}^*}(\mathbf{z}_{k+1}) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{k+1} \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}, \quad (\text{S2})$$

and $\beta_{k+2} = \dots = \beta_{d+1} = 0$, when $k = 1, \dots, d-1$.

Since

$$\begin{aligned} \mathbf{A}_k &= \begin{pmatrix} 1 & \dots & 1 \\ \log_{\mathbf{p}^*}(\mathbf{z}_1) & \dots & \log_{\mathbf{p}^*}(\mathbf{z}_{k+1}) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & \dots & 0 \\ \log_{\mathbf{p}^*}(\mathbf{z}_1) & \log_{\mathbf{p}^*}(\mathbf{z}_2) - \log_{\mathbf{p}^*}(\mathbf{z}_1) & \dots & \log_{\mathbf{p}^*}(\mathbf{z}_{k+1}) - \log_{\mathbf{p}^*}(\mathbf{z}_1) \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \end{aligned}$$

and $\log_{\mathbf{p}^*}(\mathbf{z}_2) - \log_{\mathbf{p}^*}(\mathbf{z}_1), \dots, \log_{\mathbf{p}^*}(\mathbf{z}_{k+1}) - \log_{\mathbf{p}^*}(\mathbf{z}_1)$ are linearly independent, $\mathbf{A}_k$ has a full-column rank and thus (S2) has a unique solution. It follows that (S1) has a unique solution and the Delaunay weight vector $\mathbf{w}(\mathbf{p}; \mathbb{Z})$ is unique for all $\mathbf{p} \in \mathcal{M}$ and generic $\mathbb{Z}$.

S2 Formulas of Kernel Density Estimate in Figure 4

To characterize the direction information captured by different graphs, we adopt the kernel density estimate (KDE) of directions $\mathbf{v}_1, \dots, \mathbf{v}_n \in \partial\mathcal{B}_{\mathcal{R}^d}(\mathbf{0}, 1)$,

$$h(\mathbf{u}) \propto \sum_{i=1}^n K\left(\frac{\mathbf{v}_i^\top \mathbf{u}}{\xi}\right), \quad \mathbf{u} \in \partial\mathcal{B}_{\mathcal{R}^d}(\mathbf{0}, 1),$$

proposed by [Hall et al. \(1987\)](#) and [Bai et al. \(1988\)](#). Under the case that $d = 2$, the above KDE can also be expressed as

$$h(\theta) \propto \sum_{i=1}^n K\left(\frac{1}{\xi} \cdot \mathbf{v}_i^\top \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}\right), \quad \theta \in [-\pi, \pi].$$

Here, a larger value of $h(\theta)$ indicates that more directions are concentrated around the direction θ .

In Section 3.1, we generate a dataset $\mathbb{Z}$ in $\mathcal{R}^2$, where $\mathbb{Z} = \mathbb{X} \cup \mathbb{Y}$, $\mathbb{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_{20}\}$ contains 20 points from $\text{MVN}(\mathbf{0}, \text{diag}(4, 1))$, and $\mathbb{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_{20}\}$ is obtained by rotating $\mathbb{X}$ by $\pi/2$ counterclockwise with respect to the origin. Thus, we have 40 points in $\mathbb{Z}$. Without loss of generality, we let $\mathbf{z}_i = \mathbf{x}_i$ for $i = 1, \dots, 20$ and $\mathbf{z}_i = \mathbf{y}_{i-20}$ for $i = 21, \dots, 40$. Based on the neighborhood graphs of $\mathbb{Z}$ induced by k -NN, k -MST and $\Gamma_{\mathbb{Z}}$, we derive the direction of any edge $\mathbf{z}_i \rightarrow \mathbf{z}_j$ as

$$\mathbf{v}_{ij} = \frac{(\mathbf{z}_j - \mathbf{z}_i)^\top}{\|\mathbf{z}_j - \mathbf{z}_i\|_2},$$

and compute KDE of direction among edges pointing $\mathbf{x}_i$'s and $\mathbf{y}_i$'s as follows.

- For the neighborhood graph $\mathcal{G}$ induced by either k -NN or k -MST, the KDE of direc-

tion among edges pointing $\mathbf{x}_i$'s and $\mathbf{y}_i$'s are respectively

$$h_{\mathcal{G},\mathbb{X}}(\theta) \propto \sum_{i=1}^{40} \sum_{j=1}^{20} \frac{I((\mathbf{z}_i, \mathbf{z}_j) \in \mathcal{G})}{d_i} \cdot \exp \left\{ \frac{1}{\xi} \cdot \mathbf{v}_{ij}^\top \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \right\},$$

$$\text{and } h_{\mathcal{G},\mathbb{Y}}(\theta) \propto \sum_{i=1}^{40} \sum_{j=21}^{40} \frac{I((\mathbf{z}_i, \mathbf{z}_j) \in \mathcal{G})}{d_i} \cdot \exp \left\{ \frac{1}{\xi} \cdot \mathbf{v}_{ij}^\top \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \right\},$$

for $\theta \in [-\pi, \pi]$.

- For the neighborhood graph induced by $\mathbf{\Gamma}_{\mathbb{Z}} = (\gamma_{ij})_{40 \times 40}$, the KDE of direction among edges pointing $\mathbf{x}_i$'s and $\mathbf{y}_i$'s are respectively

$$h_{\text{DW},\mathbb{X}}(\theta) \propto \sum_{i=1}^{40} \sum_{j=1}^{20} \gamma_{ij} \cdot \exp \left\{ \frac{1}{\xi} \cdot \mathbf{v}_{ij}^\top \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \right\},$$

$$\text{and } h_{\text{DW},\mathbb{Y}}(\theta) \propto \sum_{i=1}^{40} \sum_{j=21}^{40} \gamma_{ij} \cdot \exp \left\{ \frac{1}{\xi} \cdot \mathbf{v}_{ij}^\top \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \right\},$$

for $\theta \in [-\pi, \pi]$.

In Figure 4 (d)-(f) of Section 3.1, we visualize the KDEs with the bandwidth $\xi = 1/3$. By doing so, $h_{\text{DW},\mathbb{X}}(\theta)$ measures the proportion of weights assigned to edges pointing $\mathbf{x}_i$'s with direction close to θ . If most of the density in $h_{\text{DW},\mathbb{X}}(\theta)$ concentrates at the principal direction of $\mathbb{X}$, we can say that the neighborhood graphs of $\mathbb{Z}$ induced by the Delaunay weight matrix $\mathbf{\Gamma}_{\mathbb{Z}}$ capture the information of $\mathbb{X}$'s principal direction. This is also true for $h_{\text{DW},\mathbb{Y}}(\theta)$.

S3 Robust Manifold Learning Algorithm

We first provide the intrinsic dimension estimation method proposed by [Facco et al. \(2017\)](#) in Algorithm S1. We use $\varepsilon = 0.9$ in our numerical experiments. The computational complexity of Algorithm S1 is dominated by step 2 with the cost $O(n^2 D)$ where D is the dimension of the ambient space $\mathcal{R}^D$.

Algorithm S1 Estimation of intrinsic dimension (Facco et al., 2017).

- 1: **Input:** Dataset $\mathbb{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ and a tuning parameter $\varepsilon \in (0, 1)$.
- 2: Compute the distances $\|\mathbf{z}_i - \mathbf{z}_j\|$ for all $i \neq j$.
- 3: **for** $i = 1, \dots, n$ **do**
- 4: Obtain r_{i1}, r_{i2} as the smallest and the second smallest values among $\|\mathbf{z}_i - \mathbf{z}_1\|, \dots, \|\mathbf{z}_i - \mathbf{z}_{i-1}\|, \|\mathbf{z}_i - \mathbf{z}_{i+1}\|, \dots, \|\mathbf{z}_i - \mathbf{z}_n\|$.
- 5: Compute $\mu_i = r_{i2}/r_{i1}$.
- 6: **end for**
- 7: Obtain the ECDF of μ_i 's as

$$\hat{F}(\mu) = \frac{1}{n} \sum_{i=1}^n I\{\mu \geq \mu_i\}.$$

- 8: Discard all the records $(\log(\mu_i), \log(1 - \hat{F}(\mu_i)))$ with $\hat{F}(\mu_i) > \varepsilon$.
- 9: Fit a linear model,

$$\log\{1 - \hat{F}(\mu_i)\} = d \times \log(\mu_i) + \epsilon_i, \quad d \in \{1, 2, \dots\},$$

to all the remaining records $(\log(\mu_i), \log\{1 - \hat{F}(\mu_i)\})$ and obtain the estimated integer $\hat{d}$ of d .

- 10: **Output:** Estimated intrinsic dimension $\hat{d}$.
-

Based on the estimated intrinsic dimension $\hat{d}$, we follow the idea of Isomap (Tenenbaum et al., 2000) to develop a robust manifold learning algorithm as shown in Algorithm S2. A key feature is the use of the union of a k -NN graph and a MST graph as the adjacency graph, which ensures connectivity and computational efficiency. Because the computational complexity of MDS algorithm to embed $\mathbb{Z}$ in $\mathcal{R}^{\hat{d}}$ is $O(n^2\hat{d})$, together with the computational complexity $O(n^2D)$ required by estimating the intrinsic dimension, we conclude that the overall computational complexity of performing the nonlinear dimension reduction is $O(n^2D)$.

S4 Stereographic Projected DELAUNAYSPARSE Algorithm

Algorithm S3 details the procedure to find the Delaunay simplex $\mathcal{S}(\tilde{\mathbf{z}}_i; \tilde{\mathbb{Z}} \setminus \{\tilde{\mathbf{z}}_i\})$, for $i = 1, \dots, n$, where we suggest $\eta = 10$ in practice. Compared with the original DELAU-

Algorithm S2 Robust manifold learning algorithm.

- 1: **Input:** Dataset $\mathbb{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ and the $\hat{d}$ obtained by Algorithm S1.
- 2: Compute the MST $\mathcal{G}_{\text{MST}}(\mathbb{Z}, \mathcal{E}_{\text{MST}})$ with respect to the Euclidean distance among $\mathbb{Z}$.
- 3: Compute the symmetrized k -NN graph $\mathcal{G}_{k\text{-NN}}(\mathbb{Z}, \mathcal{E}_{k\text{-NN}})$ with respect to the Euclidean distance among $\mathbb{Z}$, where $k = 4 \cdot \hat{d}$ (Budninskiy et al., 2019) and the edge $(i, j) \in \mathcal{E}_{k\text{-NN}}$ if either $\mathbf{z}_i$ is one of the k nearest neighbors of $\mathbf{z}_j$ or vice versa.
- 4: Construct the extended graph $\mathcal{G}_{\text{Ext}}(\mathbb{Z}, \mathcal{E}_{\text{Ext}})$ where $\mathcal{E}_{\text{Ext}} = \mathcal{E}_{\text{MST}} \cup \mathcal{E}_{k\text{-NN}}$.
- 5: **for** $1 \leq i < j \leq n$ **do**
- 6: Compute the shortest path, i.e., the estimated geodesic distance $\tilde{d}_{\mathcal{M}}(\mathbf{z}_i, \mathbf{z}_j)$, by applying Dijkstra’s algorithm to $\mathcal{G}_{\text{Ext}}$.
- 7: **end for**
- 8: Perform the metric multidimensional scaling (MDS) to compute the low-dimensional Euclidean representation $\tilde{\mathbb{Z}} = \{\tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_n\} \subset \mathcal{R}^{\hat{d}}$ as the solution of

$$\min_{\tilde{\mathbb{Z}}} \sum_{1 \leq i < j \leq n} \left(\tilde{d}_{\mathcal{M}}(\mathbf{z}_i, \mathbf{z}_j) - \|\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j\|_2 \right)^2,$$

- 9: **Output:** The low-dimensional Euclidean representation $\tilde{\mathbb{Z}} = \{\tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_n\} \subset \mathcal{R}^{\hat{d}}$.
-

NAYSPARSE algorithm (Chang et al., 2020) which finds $\mathcal{S}(\tilde{\mathbf{z}}_i; \tilde{\mathbb{Z}} \setminus \{\tilde{\mathbf{z}}_i\})$ directly in $\mathcal{R}^d$, the stereographic projected DELAUNAYSPARSE algorithm performs the same task in a round-about way by projecting data from $\mathcal{R}^d$ to $\partial\mathcal{B}_{\eta r_{\max}} \setminus \{(\mathbf{0}_d, \eta r_{\max})^\top\}$ (steps 2-3 of Algorithm S3), finding the Delaunay simplex $\mathcal{S}^*$ on $\partial\mathcal{B}_{\eta r_{\max}} \setminus \{(\mathbf{0}_d, \eta r_{\max})^\top\}$ (steps 6-17 of Algorithm S3) and translating $\mathcal{S}^*$ back to $\mathcal{S}(\tilde{\mathbf{z}}_i; \tilde{\mathbb{Z}} \setminus \{\tilde{\mathbf{z}}_i\})$ on $\mathcal{R}^d$ (steps 18-24 of Algorithm S3). As the computational complexities of projection from $\mathcal{R}^d$ to $\partial\mathcal{B}_{\eta r_{\max}} \setminus \{(\mathbf{0}_d, \eta r_{\max})^\top\}$ ($O(d)$ for each $\tilde{\mathbf{z}}_i$) and translation of $\mathcal{S}^*$ to $\mathcal{S}(\tilde{\mathbf{z}}_i; \tilde{\mathbb{Z}} \setminus \{\tilde{\mathbf{z}}_i\})$ ($O(nd^2)$ for each $\tilde{\mathbf{z}}_i$) are dominated by finding $\mathcal{S}^*$, whose computational complexity is $O(n^{1+1/d}d^2)$ for each $\tilde{\mathbf{z}}_i$ (Chang et al., 2020), the overall computational complexity of the stereographic projected DELAUNAYSPARSE algorithm is $n \times O(n^{1+1/d}d^2) = O(n^{2+1/d}d^2)$.

To empirically evaluate the impact of the stereographic projection parameter η on the final Delaunay weight matrix, we perform an experiment. Here, we generate 1000 Gaussian datasets $\tilde{\mathbb{Z}}$ containing $n = 200$ i.i.d. individuals from the multivariate normal distribution $\text{MVN}(\mathbf{0}_d, \mathbf{I}_{d \times d})$ ($d = 20$), compute the Delaunay weight matrix $\mathbf{\Gamma}_{\tilde{\mathbb{Z}}, \eta}$ with different η ($\eta = 5, 10, 15, 20$) and compute the average Frobenius norm $\|\mathbf{\Gamma}_{\tilde{\mathbb{Z}}, \eta_1} - \mathbf{\Gamma}_{\tilde{\mathbb{Z}}, \eta_2}\|_2$ over 1000 datasets. As shown in the Table S1, the Delaunay weight matrix $\mathbf{\Gamma}_{\tilde{\mathbb{Z}}, \eta}$ only differs due to floating-

Algorithm S3 Stereographic projected DELAUNAYSPARSE algorithm

- 1: **Input:** Euclidean representation $\tilde{\mathbb{Z}} = \{\tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_n\} \subset \mathcal{R}^d$ and scaling parameter $\eta > 0$.
 - 2: Compute $r_{\max} = \max_{i=1, \dots, n} \|\tilde{\mathbf{z}}_i\|$, and apply the inverse stereographic projection $\phi_{\eta r_{\max}}^{-1}$ in (6) to obtain $\check{\mathbf{z}}_i = \phi_{\eta r_{\max}}^{-1}(\tilde{\mathbf{z}}_i)$, for $i = 1, \dots, n$.
 - 3: Set $\check{\mathbb{Z}} = \{\check{\mathbf{z}}_0, \dots, \check{\mathbf{z}}_n\}$ by augmenting the reference point $\check{\mathbf{z}}_0 = (\mathbf{0}_d, \eta r_{\max})^\top$.
 - 4: Initialize $\mathbb{S} = \emptyset$.
 - 5: **for** $i = 1, \dots, n$ **do**
 - 6: Generate a random seed Delaunay simplex $\mathcal{S}_{\text{seed}}$ on the d -hypersphere $\partial\mathcal{B}_{\eta r_{\max}}$ with respect to the set $\check{\mathbb{Z}} \setminus \{\check{\mathbf{z}}_i\}$.
 - 7: Let $\mathcal{S}_{\text{current}} = \mathcal{S}_{\text{seed}}$, $\mathbb{A}_{\text{Frontier}} = \{\mathcal{S}_{\text{seed}}\}$, $\mathbb{A}_{\text{Explored}} = \emptyset$.
 - 8: **while** $\check{\mathbf{z}}_i \notin \mathcal{S}_{\text{current}}$ **do**
 - 9: Compute $\mathbb{F}_{\check{\mathbf{z}}_i}(\mathcal{S}_{\text{current}})$ as the set of facets of $\mathcal{S}_{\text{current}}$ that are visible to $\check{\mathbf{z}}_i$. A facet $\mathcal{F}$ of the simplex $\mathcal{S}_{\text{current}}$ is visible to $\check{\mathbf{z}}_i$ if there exists an internal point $\check{\mathbf{z}}$ of $\mathcal{S}_{\text{current}}$ such that the geodesic path $\ell_{\partial\mathcal{B}_{\eta r_{\max}}}(\check{\mathbf{z}}_i, \check{\mathbf{z}})$ intersects $\mathcal{F}$.
 - 10: **for** each facet $\mathcal{F} \in \mathbb{F}_{\check{\mathbf{z}}_i}(\mathcal{S}_{\text{current}})$ **do**
 - 11: Grow a new Delaunay simplex $\mathcal{S}_{\text{new}} \neq \mathcal{S}_{\text{current}}$ on the facet $\mathcal{F}$.
 - 12: $\mathbb{A}_{\text{Frontier}} \leftarrow \mathbb{A}_{\text{Frontier}} \cup \{\mathcal{S}_{\text{new}}\}$ if $\mathcal{S}_{\text{new}} \notin \mathbb{A}_{\text{Explored}} \cup \mathbb{A}_{\text{Frontier}}$.
 - 13: **end for**
 - 14: Update $\mathbb{A}_{\text{Explored}} \leftarrow \mathbb{A}_{\text{Explored}} \cup \{\mathcal{S}_{\text{current}}\}$ and $\mathbb{A}_{\text{Frontier}} \leftarrow \mathbb{A}_{\text{Frontier}} \setminus \{\mathcal{S}_{\text{current}}\}$.
 - 15: Update $\mathcal{S}_{\text{current}}$ as the simplex that enters $\mathbb{A}_{\text{Frontier}}$ earliest among all remaining simplices in $\mathbb{A}_{\text{Frontier}}$.
 - 16: **end while**
 - 17: Let $\mathcal{S}^* = \mathcal{S}_{\text{current}}$.
 - 18: **if** $\check{\mathbf{z}}_0 \notin \mathbb{V}(\mathcal{S}^*)$ **then**
 - 19: Obtain $\mathcal{S}(\check{\mathbf{z}}_i; \tilde{\mathbb{Z}} \setminus \{\tilde{\mathbf{z}}_i\})$ as the simplex formed by $\{\tilde{\mathbf{z}}_j : \check{\mathbf{z}}_j \in \mathbb{V}(\mathcal{S}^*)\}$.
 - 20: **else**
 - 21: Compute $\mathcal{S}^\dagger$ as the neighbor simplex of $\mathcal{S}^*$ opposite to $\check{\mathbf{z}}_0$.
 - 22: Obtain $\mathcal{S}(\check{\mathbf{z}}_i; \tilde{\mathbb{Z}} \setminus \{\tilde{\mathbf{z}}_i\})$ as the simplex formed by $\{\tilde{\mathbf{z}}_j : \check{\mathbf{z}}_j \in \mathbb{V}(\mathcal{S}^\dagger)\}$.
 - 23: **end if**
 - 24: Update $\mathbb{S} = \mathbb{S} \cup \{\mathcal{S}(\check{\mathbf{z}}_i; \tilde{\mathbb{Z}} \setminus \{\tilde{\mathbf{z}}_i\})\}$
 - 25: **end for**
 - 26: **Output:** The set of Delaunay simplices $\mathbb{S}$.
-

point errors of when $\eta \geq 15$. Considering that setting η too large would make projections $\tilde{\mathbb{Z}}$ close to the same hyperplane and reduce the numerical stability, we set $\eta = 15$ in all our experiments.

Table S1: Average Frobenius norm $\|\mathbf{\Gamma}_{\tilde{\mathbb{Z}},\eta_1} - \mathbf{\Gamma}_{\tilde{\mathbb{Z}},\eta_2}\|_2$ over 1000 datasets for different pairs of (η_1, η_2) .

η_1	η_2	$\sum_{i=1}^{1000} (\ \mathbf{\Gamma}_{\tilde{\mathbb{Z}},\eta_1} - \mathbf{\Gamma}_{\tilde{\mathbb{Z}},\eta_2}\ _2) / 1000$
5	10	1.381×10^{-4}
5	15	1.474×10^{-4}
5	20	1.474×10^{-4}
10	15	1.614×10^{-5}
10	20	1.614×10^{-5}
15	20	3.166×10^{-15}

S5 Proof of Theorem 1

Given the observed pooled sample $\mathbb{Z}$ and sample sizes (n_1, n_0) , it is clear that $\mathbf{\Gamma}_{\mathbb{Z}}$ is fixed and thus the conditional distribution of T_{DW} can be deduced from the null distribution of $\boldsymbol{\delta}$. Recall that $\boldsymbol{\delta}$ contains n_1 ones and n_0 zeros. Therefore, under the null hypothesis $H_0 : F = G$, $\boldsymbol{\delta}$ is independent of $\mathbb{Z}$, and the probability of any given realization of $\boldsymbol{\delta}$ is

$$\frac{1}{\binom{n}{n_1}} = \frac{n_1!n_0!}{n!}.$$

For any distinct values i, j, k and l , among all $\binom{n}{n_1}$ realizations of $\boldsymbol{\delta}$, there are

- $\binom{n-4}{n_1-4}$ realizations satisfying $\delta_i = \delta_j = \delta_k = \delta_l = 1$;
- $\binom{n-4}{n_0-4}$ realizations satisfying $\delta_i = \delta_j = \delta_k = \delta_l = 0$;
- $\binom{n-4}{n_1-2}$ realizations satisfying $\delta_i = \delta_j = 1$ and $\delta_k = \delta_l = 0$;
- $\binom{n-4}{n_1-2}$ realizations satisfying $\delta_i = \delta_j = 0$ and $\delta_k = \delta_l = 1$.

Therefore, we have

$$\begin{aligned}
\mathcal{P}_0 &\equiv \Pr(\delta_i = \delta_j, \delta_k = \delta_l | \mathbb{Z}) \\
&= \Pr(\delta_i = \delta_j, \delta_k = \delta_l) \\
&= \Pr(\delta_i = \delta_j = \delta_k = \delta_l = 1) + \Pr(\delta_i = \delta_j = \delta_k = \delta_l = 0) \\
&\quad + \Pr(\delta_i = \delta_j = 1, \delta_k = \delta_l = 0) + \Pr(\delta_i = \delta_j = 0, \delta_k = \delta_l = 1) \\
&= \frac{\binom{n-4}{n_1-4} + \binom{n-4}{n_0-4} + 2\binom{n-4}{n_1-2}}{\binom{n}{n_1}} \\
&= \frac{n_1(n_1-1)(n_1-2)(n_1-3) + n_0(n_0-1)(n_0-2)(n_0-3) + 2n_1(n_1-1)n_0(n_0-1)}{n(n-1)(n-2)(n-3)} \\
&= 1 - \frac{4n_1n_0}{n(n-1)} + \frac{4n_1(n_1-1)n_0(n_0-1)}{n(n-1)(n-2)(n-3)}. \tag{S3}
\end{aligned}$$

For any distinct values i, j and k , among all $\binom{n}{n_1}$ realizations of $\boldsymbol{\delta}$, there are

- $\binom{n-3}{n_1-3}$ realizations satisfying $\delta_i = \delta_j = \delta_k = 1$;
- $\binom{n-3}{n_0-3}$ realizations satisfying $\delta_i = \delta_j = \delta_k = 0$.

Therefore, we have

$$\begin{aligned}
\mathcal{P}_1 &\equiv \Pr(\delta_i = \delta_j = \delta_k | \mathbb{Z}) \\
&= \Pr(\delta_i = \delta_j = \delta_k) \\
&= \Pr(\delta_i = \delta_j = \delta_k = 1) + \Pr(\delta_i = \delta_j = \delta_k = 0) \\
&= \frac{\binom{n-3}{n_1-3} + \binom{n-3}{n_0-3}}{\binom{n}{n_1}} \\
&= \frac{n_1(n_1-1)(n_1-2) + n_0(n_0-1)(n_0-2)}{n(n-1)(n-2)} \\
&= 1 - \frac{3n_1n_0}{n(n-1)}. \tag{S4}
\end{aligned}$$

For any distinct values i and j , among all $\binom{n}{n_1}$ realizations of $\boldsymbol{\delta}$, there are

- $\binom{n-2}{n_1-2}$ realizations satisfying $\delta_i = \delta_j = 1$;
- $\binom{n-2}{n_0-2}$ realizations satisfying $\delta_i = \delta_j = 0$.

Therefore, we have

$$\begin{aligned}
\mathcal{P}_2 &\equiv \Pr(\delta_i = \delta_j | \mathbb{Z}) \\
&= \Pr(\delta_i = \delta_j) \\
&= \Pr(\delta_i = \delta_j = 1) + \Pr(\delta_i = \delta_j = 0) \\
&= \frac{\binom{n-2}{n_1-2} + \binom{n-2}{n_0-2}}{\binom{n}{n_1}} \\
&= \frac{n_1(n_1-1) + n_0(n_0-1)}{n(n-1)} \\
&= 1 - \frac{2n_1n_0}{n(n-1)}.
\end{aligned} \tag{S5}$$

Recalling the notation $\check{\gamma}_{ij} = \gamma_{ij} + \gamma_{ji}$, since

$$T_{\text{DW}} = \frac{1}{n} \sum_{j=1}^n \sum_{i \neq j} \gamma_{ij} \times I(\delta_i = \delta_j) = \frac{1}{n} \sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} \times I(\delta_i = \delta_j)$$

and $\sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} = n$, we have that

$$\begin{aligned}
E_{H_0}(T_{\text{DW}} | \mathbb{Z}) &= \sum_{1 \leq i < j \leq n} n^{-1} \check{\gamma}_{ij} \times \Pr(\delta_i = \delta_j | \mathbb{Z}) \\
&= \sum_{1 \leq i < j \leq n} n^{-1} \check{\gamma}_{ij} \mathcal{P}_2 \\
&= \mathcal{P}_2 \\
&= \frac{n_1(n_1-1) + n_0(n_0-1)}{n(n-1)}
\end{aligned} \tag{S6}$$

and

$$\begin{aligned}
&\text{Var}_{H_0}(T_{\text{DW}} | \mathbb{Z}) \\
&= \sum_{1 \leq i < j \leq n} n^{-2} \check{\gamma}_{ij}^2 \times \text{Var}_{H_0}\{I(\delta_i = \delta_j)\} \\
&\quad + \sum_{1 \leq i < j < k \leq n} 2n^{-2} \check{\gamma}_{ij} \check{\gamma}_{ik} \times \text{Cov}_{H_0}\{I(\delta_i = \delta_j), I(\delta_i = \delta_k)\}
\end{aligned}$$

$$\begin{aligned}
& + \sum_{1 \leq i < j < k \leq n} 2n^{-2} \check{\gamma}_{ij} \check{\gamma}_{jk} \times \text{Cov}_{H_0} \{I(\delta_i = \delta_j), I(\delta_j = \delta_k)\} \\
& + \sum_{1 \leq i < j < k \leq n} 2n^{-2} \check{\gamma}_{ik} \check{\gamma}_{jk} \times \text{Cov}_{H_0} \{I(\delta_i = \delta_k), I(\delta_j = \delta_k)\} \\
& + \sum_{1 \leq i < j < k < l \leq n} 2n^{-2} \check{\gamma}_{ij} \check{\gamma}_{kl} \times \text{Cov}_{H_0} \{I(\delta_i = \delta_j), I(\delta_k = \delta_l)\} \\
& + \sum_{1 \leq i < j < k < l \leq n} 2n^{-2} \check{\gamma}_{ik} \check{\gamma}_{jl} \times \text{Cov}_{H_0} \{I(\delta_i = \delta_k), I(\delta_j = \delta_l)\} \\
& + \sum_{1 \leq i < j < k < l \leq n} 2n^{-2} \check{\gamma}_{il} \check{\gamma}_{jk} \times \text{Cov}_{H_0} \{I(\delta_i = \delta_l), I(\delta_j = \delta_k)\} \\
& = \sum_{1 \leq i < j \leq n} n^{-2} \check{\gamma}_{ij}^2 \times (\mathcal{P}_2 - \mathcal{P}_2^2) + \sum_{1 \leq i < j < k \leq n} 2n^{-2} (\check{\gamma}_{ij} \check{\gamma}_{ik} + \check{\gamma}_{ij} \check{\gamma}_{jk} + \check{\gamma}_{ik} \check{\gamma}_{jk}) \times (\mathcal{P}_1 - \mathcal{P}_2^2) \\
& + \sum_{1 \leq i < j < k < l \leq n} 2n^{-2} (\check{\gamma}_{ij} \check{\gamma}_{kl} + \check{\gamma}_{ik} \check{\gamma}_{jl} + \check{\gamma}_{il} \check{\gamma}_{jk}) \times (\mathcal{P}_0 - \mathcal{P}_2^2). \tag{S7}
\end{aligned}$$

Using (S3)–(S5), we have

$$\begin{aligned}
\mathcal{P}_0 - \mathcal{P}_2^2 &= \frac{4n_1(n_1 - 1)n_0(n_0 - 1)}{n(n - 1)(n - 2)(n - 3)} - \frac{4n_1^2 n_0^2}{n^2(n - 1)^2}, \\
\mathcal{P}_1 - \mathcal{P}_2^2 &= \frac{n_1 n_0}{n(n - 1)} - \frac{4n_1^2 n_0^2}{n^2(n - 1)^2}, \\
\mathcal{P}_2 - \mathcal{P}_2^2 &= \frac{2n_1 n_0}{n(n - 1)} - \frac{4n_1^2 n_0^2}{n^2(n - 1)^2},
\end{aligned}$$

and thus

$$\begin{aligned}
& \text{Var}_{H_0}(T_{\text{DW}} | \mathbb{Z}) \\
&= \sum_{1 \leq i < j \leq n} n^{-2} \check{\gamma}_{ij}^2 \times \frac{2n_1 n_0}{n(n - 1)} + \sum_{1 \leq i < j < k \leq n} 2n^{-2} (\check{\gamma}_{ij} \check{\gamma}_{ik} + \check{\gamma}_{ij} \check{\gamma}_{jk} + \check{\gamma}_{ik} \check{\gamma}_{jk}) \times \frac{n_1 n_0}{n(n - 1)} \\
&+ \sum_{1 \leq i < j < k < l \leq n} 2n^{-2} (\check{\gamma}_{ij} \check{\gamma}_{kl} + \check{\gamma}_{ik} \check{\gamma}_{jl} + \check{\gamma}_{il} \check{\gamma}_{jk}) \times \frac{4n_1(n_1 - 1)n_0(n_0 - 1)}{n(n - 1)(n - 2)(n - 3)} \\
&- \frac{4n_1^2 n_0^2}{n^4(n - 1)^2} \times \left(\sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} \right)^2 \\
&= \sum_{1 \leq i < j \leq n} \frac{\check{\gamma}_{ij}^2}{n(n - 1)/2} \times \frac{n_1 n_0}{n^2} + \sum_{1 \leq i < j < k \leq n} \frac{(\check{\gamma}_{ij} \check{\gamma}_{ik} + \check{\gamma}_{ij} \check{\gamma}_{jk} + \check{\gamma}_{ik} \check{\gamma}_{jk})}{n(n - 1)(n - 2)/6} \times \frac{n_1 n_0(n - 2)}{3n^2}
\end{aligned}$$

$$\begin{aligned}
& + \sum_{1 \leq i < j < k < l \leq n} \frac{(\check{\gamma}_{ij}\check{\gamma}_{kl} + \check{\gamma}_{ik}\check{\gamma}_{jl} + \check{\gamma}_{il}\check{\gamma}_{jk})}{n(n-1)(n-2)(n-3)/24} \times \frac{n_1(n_1-1)n_0(n_0-1)}{3n^2} - \frac{4n_1^2n_0^2}{(n-1)^2n^2} \\
& = \frac{n_1(n_1-1)n_0(n_0-1)}{3n^2} V_{0,\mathbb{Z}} + \frac{n_1n_0(n-2)}{3n^2} V_{1,\mathbb{Z}} + \frac{n_1n_0}{n^2} V_{2,\mathbb{Z}} - \frac{4n_1^2n_0^2}{(n-1)^2n^2}, \tag{S8}
\end{aligned}$$

given that $\sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} = n$.

S6 Proof of Theorem 2

Recalling from (4) where

$$T_{\text{DW}} = \frac{1}{n} \sum_{j=1}^n \sum_{i \neq j} \gamma_{ij} \times I(\delta_i = \delta_j),$$

we define an analogous statistic,

$$T_{\text{DW}}^* = \frac{1}{n} \sum_{j=1}^n \sum_{i \neq j} \gamma_{ij} \times I(\delta_i^* = \delta_j^*) = \frac{1}{n} \sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} \times I(\delta_i^* = \delta_j^*), \tag{S9}$$

where $\delta^* = \{\delta_1^*, \dots, \delta_n^*\}$ and

$$\delta_i^* = \begin{cases} 1, & \text{with probability } n_1/n, \\ 0, & \text{with probability } n_0/n, \end{cases} \tag{S10}$$

for $i = 1, \dots, n$. Let $n_1^* = \sum_{i=1}^n \delta_i^*$ and let $\stackrel{D}{=}$ denote equality in distribution, it is clear that when H_0 is true,

$$(T_{\text{DW}}^* | n_1^* = n_1, \mathbb{Z}) \stackrel{D}{=} (T_{\text{DW}} | \mathbb{Z}), \tag{S11}$$

for all $\mathbb{Z}$. To prove the asymptotic null distribution (7), we study the asymptotic distribution of (T_{DW}^*, n_1^*) conditional on $\mathbb{Z}$. We first introduce several lemmas in S6.1–S6.3 and then prove Theorem 2 in S6.4.

S6.1 First Two Moments of (T_{DW}^*, n_1^*)

Lemma 1. *For the first two moments of (T_{DW}^*, n_1^*) , we have*

$$\begin{aligned}
E_{H_0}(n_1^*/n) &= \frac{n_1}{n}, \\
\text{Var}_{H_0}(n_1^*/n) &= \frac{n_1 n_0}{n^3}, \\
E_{H_0}(T_{\text{DW}}^* | \mathbb{Z}) &= \frac{n_1^2 + n_0^2}{n^2}, \\
\text{Var}_{H_0}(T_{\text{DW}}^* | \mathbb{Z}) &= \frac{n_1^2 n_0^2 (n-1)(n-2)(n-3)}{3n^5} V_{0,\mathbb{Z}} + \frac{n_1 n_0 (n-1)(n-2)}{3n^3} V_{1,\mathbb{Z}} \\
&\quad + \frac{n_1 n_0 (n-1)}{n^3} V_{2,\mathbb{Z}} - \frac{4n_1^2 n_0^2}{n^4}, \\
\text{Cov}_{H_0}(T_{\text{DW}}^*, n_1^*/n | \mathbb{Z}) &= \frac{2n_1 n_0 (n_1 - n_0)}{n^4}.
\end{aligned}$$

Proof of Lemma 1: The δ_i^* 's are defined to be i.i.d. Bernoulli variables with probability n_1/n , and thus we have for any distinct values i, j, k and l ,

$$\begin{aligned}
\mathcal{P}_0^* &= \Pr(\delta_i^* = \delta_j^*, \delta_k^* = \delta_l^*) \\
&= \Pr(\delta_i^* = \delta_j^* = \delta_k^* = \delta_l^* = 1) + \Pr(\delta_i^* = \delta_j^* = \delta_k^* = \delta_l^* = 0) \\
&\quad + \Pr(\delta_i^* = \delta_j^* = 1, \delta_k^* = \delta_l^* = 0) + \Pr(\delta_i^* = \delta_j^* = 0, \delta_k^* = \delta_l^* = 1) \\
&= \frac{n_1^4}{n^4} + \frac{n_0^4}{n^4} + \frac{n_1^2 n_0^2}{n^4} + \frac{n_1^2 n_0^2}{n^4} \\
&= 1 - \frac{4n_1 n_0}{n^2} + \frac{4n_1^2 n_0^2}{n^4}, \tag{S12}
\end{aligned}$$

$$\begin{aligned}
\mathcal{P}_1^* &= \Pr(\delta_i^* = \delta_j^* = \delta_k^*) \\
&= \Pr(\delta_i^* = \delta_j^* = \delta_k^* = 1) + \Pr(\delta_i^* = \delta_j^* = \delta_k^* = 0) \\
&= \frac{n_1^3}{n^3} + \frac{n_0^3}{n^3} \\
&= 1 - \frac{3n_1 n_0}{n^2}, \tag{S13}
\end{aligned}$$

$$\begin{aligned}
\mathcal{P}_2^* &= \Pr(\delta_i^* = \delta_j^*) \\
&= \Pr(\delta_i^* = \delta_j^* = 1) + \Pr(\delta_i^* = \delta_j^* = 0) \\
&= \frac{n_1^2}{n^2} + \frac{n_0^2}{n^2}
\end{aligned}$$

$$= 1 - \frac{2n_1n_0}{n^2}. \quad (\text{S14})$$

Since n_1^* follows the binomial distribution with parameters n and n_1/n , we have

$$\begin{aligned} E_{H_0}(n_1^*/n) &= n \times \frac{n_1}{n} \times \frac{1}{n} = \frac{n_1}{n}, \\ \text{Var}_{H_0}(n_1^*/n) &= n \times \frac{n_1}{n} \times \left(1 - \frac{n_1}{n}\right) \times \frac{1}{n^2} = \frac{n_1n_0}{n^3}. \end{aligned}$$

Based on equations (S12)–(S14), we have

$$\begin{cases} \mathcal{P}_0^* - (\mathcal{P}_2^*)^2 = 0, \\ \mathcal{P}_1^* - (\mathcal{P}_2^*)^2 = \frac{n_1n_0}{n^2} - \frac{4n_1^2n_0^2}{n^4}, \\ \mathcal{P}_2^* - (\mathcal{P}_2^*)^2 = \frac{2n_1n_0}{n^2} - \frac{4n_1^2n_0^2}{n^4}. \end{cases} \quad (\text{S15})$$

Thus, similar to (S6)–(S8), we have

$$\begin{aligned} E_{H_0}(T_{\text{DW}}^*|\mathbb{Z}) &= \sum_{1 \leq i < j \leq n} n^{-1} \check{\gamma}_{ij} \times \Pr(\delta_i^* = \delta_j^*) = \sum_{1 \leq i < j \leq n} n^{-1} \check{\gamma}_{ij} \times \mathcal{P}_2^* = \mathcal{P}_2^* = \frac{n_1^2 + n_0^2}{n^2}, \\ \text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z}) &= \sum_{1 \leq i < j \leq n} n^{-2} \check{\gamma}_{ij}^2 \times \{\mathcal{P}_2^* - (\mathcal{P}_2^*)^2\} \\ &\quad + \sum_{1 \leq i < j < k \leq n} 2n^{-2} (\check{\gamma}_{ij}\check{\gamma}_{ik} + \check{\gamma}_{ij}\check{\gamma}_{jk} + \check{\gamma}_{ik}\check{\gamma}_{jk}) \times \{\mathcal{P}_1^* - (\mathcal{P}_2^*)^2\} \\ &\quad + \sum_{1 \leq i < j < k < l \leq n} 2n^{-2} (\check{\gamma}_{ij}\check{\gamma}_{kl} + \check{\gamma}_{ik}\check{\gamma}_{jl} + \check{\gamma}_{il}\check{\gamma}_{jk}) \times \{\mathcal{P}_0^* - (\mathcal{P}_2^*)^2\} \\ &= \sum_{1 \leq i < j \leq n} \check{\gamma}_{ij}^2 \times \frac{2n_1n_0}{n^4} + \sum_{1 \leq i < j < k \leq n} 2(\check{\gamma}_{ij}\check{\gamma}_{ik} + \check{\gamma}_{ij}\check{\gamma}_{jk} + \check{\gamma}_{ik}\check{\gamma}_{jk}) \times \frac{n_1n_0}{n^4} \\ &\quad - \frac{4n_1^2n_0^2}{n^6} \times \left\{ \left(\sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} \right)^2 - \sum_{1 \leq i < j < k < l \leq n} 2(\check{\gamma}_{ij}\check{\gamma}_{kl} + \check{\gamma}_{ik}\check{\gamma}_{jl} + \check{\gamma}_{il}\check{\gamma}_{jk}) \right\} \\ &= \sum_{1 \leq i < j \leq n} \frac{\check{\gamma}_{ij}^2}{n(n-1)/2} \times \frac{n_1n_0(n-1)}{n^3} \\ &\quad + \sum_{1 \leq i < j < k \leq n} \frac{\check{\gamma}_{ij}\check{\gamma}_{ik} + \check{\gamma}_{ij}\check{\gamma}_{jk} + \check{\gamma}_{ik}\check{\gamma}_{jk}}{n(n-1)(n-2)/6} \times \frac{n_1n_0(n-1)(n-2)}{3n^3} \end{aligned}$$

$$\begin{aligned}
& -\frac{4n_1^2n_0^2}{n^4} \left\{ 1 - \frac{\sum_{1 \leq i < j < k < l \leq n} 2(\check{\gamma}_{ij}\check{\gamma}_{kl} + \check{\gamma}_{ik}\check{\gamma}_{jl} + \check{\gamma}_{il}\check{\gamma}_{jk})}{n^2} \right\} \\
& = \frac{n_1^2n_0^2(n-1)(n-2)(n-3)}{3n^5} V_{0,\mathbb{Z}} + \frac{n_1n_0(n-1)(n-2)}{3n^3} V_{1,\mathbb{Z}} \\
& + \frac{n_1n_0(n-1)}{n^3} V_{2,\mathbb{Z}} - \frac{4n_1^2n_0^2}{n^4},
\end{aligned} \tag{S16}$$

with $\sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} = n$.

Because δ_i^* , δ_j^* and δ_k^* are i.i.d. Bernoulli variables with probability n_1/n , we have

$$\text{Cov}_{H_0}\{I(\delta_i^* = \delta_j^*), \delta_k^*\} = 0, \tag{S17}$$

$$\begin{aligned}
\text{Cov}_{H_0}\{I(\delta_i^* = \delta_j^*), \delta_i^* + \delta_j^*\} &= \text{E}_{H_0}\{I(\delta_i^* = \delta_j^*) \times (\delta_i^* + \delta_j^*)\} \\
&\quad - \text{E}_{H_0}\{I(\delta_i^* = \delta_j^*)\} \text{E}_{H_0}(\delta_i^* + \delta_j^*) \\
&= 2 \times \text{Pr}(\delta_i^* = \delta_j^* = 1) \\
&\quad - 2 \times \text{Pr}(\delta_i^* = \delta_j^*) \times \text{Pr}(\delta_i^* = 1) \\
&= 2 \times \text{Pr}(\delta_i^* = \delta_j^* = 1) - 2 \times \mathcal{P}_2^* \times \text{Pr}(\delta_i^* = 1) \\
&= \frac{2n_1^2}{n^2} - \frac{2(n_1^3 + n_1n_0^2)}{n^3} \\
&= \frac{2(n_1^3 + n_1^2n_0 - n_1^3 - n_1n_0^2)}{n^3} \\
&= \frac{2n_1n_0(n_1 - n_0)}{n^3}.
\end{aligned} \tag{S18}$$

Using (S17)–(S18) and $\sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} = n$, we have

$$\begin{aligned}
\text{Cov}_{H_0}(T_{\text{DW}}^*, n_1^*/n | \mathbb{Z}) &= \sum_{1 \leq i < j \leq n} n^{-2} \check{\gamma}_{ij} \text{Cov}_{H_0}\{I(\delta_i^* = \delta_j^*), \delta_i^* + \delta_j^*\} \\
&\quad + \sum_{1 \leq i < j \leq n} \sum_{k \neq i, j} n^{-2} \check{\gamma}_{ij} \text{Cov}_{H_0}\{I(\delta_i^* = \delta_j^*), \delta_k^*\} \\
&= \sum_{1 \leq i < j \leq n} n^{-2} \check{\gamma}_{ij} \times \frac{2n_1n_0(n_1 - n_0)}{n^3} \\
&= \frac{2n_1n_0(n_1 - n_0)}{n^4}.
\end{aligned}$$

S6.2 Proximity Graph Induced by $\Gamma_{\mathbb{Z}}$

For any events A and B , let “ $A \implies B$ ” denote the statement “ A is sufficient for B ” and let “ $A \iff B$ ” denote the statement “ A is necessary and sufficient for B ”.

We first define the proximity graph $\mathcal{G}(\mathbb{Z}, \mathcal{E})$ induced by $\Gamma_{\mathbb{Z}}$ by letting

$$\mathcal{E} = \{(i, j) : \check{\gamma}_{ij} > 0\} \tag{S19}$$

and the weight of edge (i, j) be $\check{\gamma}_{ij}$. Recalling that $\check{\gamma}_{ij} = \gamma_{ij} + \gamma_{ji}$, we have $(i, j) \in \mathcal{E} \iff \{\gamma_{ij} > 0 \text{ or } \gamma_{ji} > 0\}$. To investigate the asymptotic property of degrees $d_1, \dots, d_n$ of $\mathcal{G}$, where $d_i = \sum_{j \neq i} I(\check{\gamma}_{ij} > 0)$, we first deduce a necessary condition for any pair of (i, j) to be linked in $\mathcal{G}$.

Lemma 2. *Let $\mathbb{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ be an i.i.d. sample from a distribution H with Lipschitz continuous density h with respect to the measure on $\mathcal{M}$ induced by the Lebesgue measure on the ambient Euclidean space. The proximity graph $\mathcal{G}(\mathbb{Z}, \mathcal{E})$ defined by (S19) is a subgraph of the auxiliary graph $\mathcal{G}^*(\mathbb{Z}, \mathcal{E}^*)$ a.s. as $n \rightarrow \infty$, where $\mathcal{E}^* = \{(i, j) : \exists \mathcal{S} \in \mathcal{DT}_{\mathcal{M}}(\mathbb{Z}) \text{ s.t. } \mathbf{z}_i, \mathbf{z}_j \in \mathbb{V}(\mathcal{S})\}$.*

Although the proximity graph $\mathcal{G}(\mathbb{Z}, \mathcal{E})$ and the auxiliary graph $\mathcal{G}^*(\mathbb{Z}, \mathcal{E}^*)$ are both defined in terms of Delaunay triangulation, they are different because $\mathcal{G}$ depicts the relation between each $\mathbf{z}_i$ to the Delaunay triangulation $\mathcal{DT}_{\mathcal{M}}(\mathbb{Z} \setminus \{\mathbf{z}_i\})$ of the leave-one-out dataset while $\mathcal{G}^*$ is directly constructed on the Delaunay triangulation $\mathcal{DT}_{\mathcal{M}}(\mathbb{Z})$ of the whole dataset.

Proof of Lemma 2:

For all i, j , we intend to show that

$$\gamma_{ij} > 0 \implies \exists \mathcal{S} \in \mathcal{DT}_{\mathcal{M}}(\mathbb{Z}) \text{ s.t. } \mathbf{z}_i, \mathbf{z}_j \in \mathbb{V}(\mathcal{S}), \text{ a.s.} \tag{S20}$$

as $n \rightarrow \infty$.

For fixed i , since $\mathcal{M}$ is geodesically convex, we project all the $\mathbf{z}_j$'s ($j \neq i$) on $\mathcal{M}$ to $\log_{\mathbf{z}_i}(\mathbf{z}_j)$ on the tangent space $\mathbf{T}_{\mathbf{z}_i}\mathcal{M}$. Let $\mathbb{Z}^{(i)} = \{\log_{\mathbf{z}_i}(\mathbf{z}_j) : j = 1, \dots, n\} \subset \mathbf{T}_{\mathbf{z}_i}\mathcal{M}$ and

define $\mathcal{DT}_{\mathbf{T}_{\mathbf{z}_i}\mathcal{M}}(\mathbb{Z}^{(i)})$ as the Delaunay triangulation of projected dataset $\mathbb{Z}^{(i)}$ on the tangent space $\mathbf{T}_{\mathbf{z}_i}\mathcal{M}$. For any fixed i , we define

$$\mathbf{N}(\mathbf{z}_i, \mathcal{DT}_{\mathcal{M}}(\mathbb{Z})) = \{\mathbf{z}_j : \exists \mathcal{S} \in \mathcal{DT}_{\mathcal{M}}(\mathbb{Z}) \text{ s.t. } \mathbf{z}_i, \mathbf{z}_j \in \mathbb{V}(\mathcal{S})\},$$

as the set of all neighbors of $\mathbf{z}_i$ in $\mathcal{DT}_{\mathcal{M}}(\mathbb{Z})$. Analogously, we also define

$$\mathbf{N}(\mathbf{z}_i, \mathcal{DT}_{\mathbf{T}_{\mathbf{z}_i}\mathcal{M}}(\mathbb{Z}^{(i)})) = \{\log_{\mathbf{z}_i}(\mathbf{z}_j) : \exists \mathcal{S} \in \mathcal{DT}_{\mathbf{T}_{\mathbf{z}_i}\mathcal{M}}(\mathbb{Z}^{(i)}) \text{ s.t. } \mathbf{0}, \log_{\mathbf{z}_i}(\mathbf{z}_j) \in \mathbb{V}(\mathcal{S})\},$$

as the set of all neighbors of $\mathbf{z}_i$ in $\mathcal{DT}_{\mathbf{T}_{\mathbf{z}_i}\mathcal{M}}(\mathbb{Z}^{(i)})$. Here, with slight abuse of notation, $\log_{\mathbf{z}_i}(\mathbf{z}_j)$ denotes the endpoint of the vector $\log_{\mathbf{z}_i}(\mathbf{z}_j)$. Since $\mathcal{M}$ is locally Euclidean and H is a continuous distribution on $\mathcal{M}$, we have

$$\mathbf{z}_j \in \mathbf{N}(\mathbf{z}_i, \mathcal{DT}_{\mathcal{M}}(\mathbb{Z})) \iff \log_{\mathbf{z}_i}(\mathbf{z}_j) \in \mathbf{N}(\mathbf{z}_i, \mathcal{DT}_{\mathbf{T}_{\mathbf{z}_i}\mathcal{M}}(\mathbb{Z}^{(i)})), \forall j \neq i, \text{ a.s.} \quad (\text{S21})$$

as $n \rightarrow \infty$.

Let $\mathbf{\Gamma}_{\mathbb{Z}^{(i)}} = (\gamma_{kj}^{(i)})_{n \times n}$ be the Delaunay weight matrix of $\mathbb{Z}^{(i)} \subset \mathbf{T}_{\mathbf{z}_i}\mathcal{M}$. Since $\mathcal{S}(\mathbf{z}_i; \mathbb{Z} \setminus \{\mathbf{z}_i\})$ on $\mathcal{M}$ and $\mathcal{S}(\mathbf{z}_i; \mathbb{Z}^{(i)} \setminus \{\mathbf{z}_i\})$ on $\mathbf{T}_{\mathbf{z}_i}\mathcal{M}$ depict neighborhoods of $\mathbf{z}_i$ in $\mathcal{DT}_{\mathcal{M}}(\mathbb{Z} \setminus \{\mathbf{z}_i\})$ and $\mathcal{DT}_{\mathbf{T}_{\mathbf{z}_i}\mathcal{M}}(\mathbb{Z}^{(i)} \setminus \{\mathbf{z}_i\})$ respectively, we have

$$\gamma_{ij} > 0 \iff \gamma_{ij}^{(i)} > 0, \forall j \neq i, \text{ a.s.} \quad (\text{S22})$$

as $n \rightarrow \infty$.

Using (S21)–(S22), it is clear that proving (S20) is equivalent to proving that, for all $i \neq j$,

$$\gamma_{ij}^{(i)} > 0 \implies \log_{\mathbf{z}_i}(\mathbf{z}_j) \in \mathbf{N}(\mathbf{z}_i, \mathcal{DT}_{\mathbf{T}_{\mathbf{z}_i}\mathcal{M}}(\mathbb{Z}^{(i)})), \text{ a.s.} \quad (\text{S23})$$

as $n \rightarrow \infty$.

Recalling the property (1) in the Euclidean space (Abadie and L'Hour, 2021), we have

$$(\gamma_{i1}^{(i)}, \dots, \gamma_{i,i-1}^{(i)}, \gamma_{i,i+1}^{(i)}, \dots, \gamma_{in}^{(i)})^\top = \mathbf{w}(\mathbf{z}_i; \mathbb{Z}^{(i)} \setminus \{\mathbf{z}_i\}) = \lim_{\lambda \rightarrow 0} \mathbf{w}^{(\lambda)}(\mathbf{z}_i; \mathbb{Z}^{(i)} \setminus \{\mathbf{z}_i\}), \quad (\text{S24})$$

where $\mathbf{w}^{(\lambda)}(\mathbf{z}_i; \mathbb{Z}^{(i)} \setminus \{\mathbf{z}_i\}) = (w_1^{(\lambda)}, \dots, w_{i-1}^{(\lambda)}, w_{i+1}^{(\lambda)}, \dots, w_n^{(\lambda)})^\top$ is the solution of the optimization problem,

$$\min \left\| \sum_{j \neq i} w_j \log_{\mathbf{z}_i}(\mathbf{z}_j) \right\|^2 + \lambda \sum_{j \neq i} w_j \|\log_{\mathbf{z}_i}(\mathbf{z}_j)\|^2,$$

subject to $\mathbf{w} = (w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_n)^\top \in [0, 1]^{n-1}$ and $\sum_{j \neq i} w_j = 1$. By Theorem 3 of [Abadie and L'Hour \(2021\)](#) that for all $\log_{\mathbf{z}_i}(\mathbf{z}_j) \notin \mathbf{N}(\mathbf{z}_i, \mathcal{DT}_{\mathbf{z}_i, \mathcal{M}}(\mathbb{Z}^{(i)}))$, the corresponding $w_j^{(\lambda)}$ in $\mathbf{w}^{(\lambda)}(\mathbf{z}_i; \mathbb{Z}^{(i)} \setminus \{\mathbf{z}_i\})$ are all 0, we have

$$w_j^{(\lambda)} > 0 \implies \log_{\mathbf{z}_i}(\mathbf{z}_j) \in \mathbf{N}(\mathbf{z}_i, \mathcal{DT}_{\mathbf{z}_i, \mathcal{M}}(\mathbb{Z}^{(i)})), \quad \forall j \neq i. \quad (\text{S25})$$

Letting $\lambda \rightarrow 0$, we conclude (S23) from (S24)–(S25) and, as a result, (S20) is shown.

Similarly, we also have

$$\gamma_{ji} > 0 \implies \exists \mathcal{S} \in \mathcal{DT}_{\mathcal{M}}(\mathbb{Z}), \text{ s.t. } \mathbf{z}_i, \mathbf{z}_j \in \mathbb{V}(\mathcal{S}), \text{ a.s.}$$

as $n \rightarrow \infty$, and thus

$$\begin{aligned} (i, j) \in \mathcal{E} &\iff \gamma_{ij} > 0 \text{ or } \gamma_{ji} > 0 \\ &\implies \exists \mathcal{S} \in \mathcal{DT}_{\mathcal{M}}(\mathbb{Z}), \text{ s.t. } \mathbf{z}_i, \mathbf{z}_j \in \mathbb{V}(\mathcal{S}) \iff (i, j) \in \mathcal{E}^*, \text{ a.s.} \end{aligned}$$

as $n \rightarrow \infty$. Hence, Lemma 2 is proven.

We then provide the asymptotic property of $d_1, \dots, d_n$ in Lemma 3.

Lemma 3. *Let $\mathbb{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ be an i.i.d. sample from distribution H with Lipschitz continuous density h with respect to the measure on $\mathcal{M}$ induced by the Lebesgue measure on the ambient Euclidean space. Assume that there exists $\nu < 9/4$ such that $\sum_{i=1}^n |\mathcal{N}_i|^4 = o(n^\nu)$, where $\mathcal{N}_i = \cup_{\mathcal{S} \in \mathcal{DT}_{\mathcal{M}}(\mathbb{Z}): \mathbf{z}_i \in \mathcal{S}} \mathbb{V}(\mathcal{S})$. Then, for $\eta = 1, 2, 3, 4$, the degrees d_i of the graph $\mathcal{G}(\mathbb{Z}, \mathcal{E})$ defined by (S19) satisfy*

$$\sum_{i=1}^n d_i^\eta = o(n^{1+\eta(\nu-1)/4}). \quad (\text{S26})$$

Proof of Lemma 3: By Lemma 2 which suggests that $\mathcal{G}$ is a subgraph of $\mathcal{G}^*$ a.s. as $n \rightarrow \infty$, we have for $\eta = 1, 2, 3, 4$,

$$\sum_{i=1}^n d_i^\eta \leq \sum_{i=1}^n |\mathcal{N}_i|^\eta.$$

Thus, (S26) holds when $\eta = 4$ because $\sum_{i=1}^n |\mathcal{N}_i|^4 = o(n^\nu)$. In addition, by the Cauchy–Schwartz inequality, we have

- when $\eta = 2$,

$$\sum_{i=1}^n d_i^2 \leq \sum_{i=1}^n |\mathcal{N}_i|^2 \leq \sqrt{n \times \sum_{i=1}^n |\mathcal{N}_i|^4} = \sqrt{o(n \times n^\nu)} = o(n^{1+(\nu-1)/2});$$

- when $\eta = 1$,

$$\sum_{i=1}^n d_i \leq \sum_{i=1}^n |\mathcal{N}_i| \leq \sqrt{n \times \sum_{i=1}^n |\mathcal{N}_i|^2} = \sqrt{o(n \times n^{1+(\nu-1)/2})} = o(n^{1+(\nu-1)/4});$$

- when $\eta = 3$,

$$\sum_{i=1}^n d_i^3 \leq \sum_{i=1}^n |\mathcal{N}_i|^3 \leq \sqrt{\sum_{i=1}^n |\mathcal{N}_i|^2 \times \sum_{i=1}^n |\mathcal{N}_i|^4} = \sqrt{o(n^{1+(\nu-1)/2} \times n^\nu)} = o(n^{1+3(\nu-1)/4}).$$

Hence, Lemma 3 is proven.

S6.3 Asymptotic Distribution of (T_{DW}^*, n_1^*) Conditional on $\mathbb{Z}$

To deduce the asymptotic distribution of (T_{DW}^*, n_1^*) conditional on $\mathbb{Z}$, we first introduce a lemma obtained from the proof of Theorem 5.1 of [Chen et al. \(2018\)](#).

Lemma 4. ([Chen et al., 2018](#)) *Consider a proximity graph $\mathcal{G}(\mathbb{Z}, \mathcal{E})$ of the data $\mathbb{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$. For each edge $e = (i, j)$ of $\mathcal{G}$, we define two sets,*

$$A_e = \{e\} \cup \{e' \in \mathcal{G} : e \cap e' \neq \emptyset\}, \quad B_e = A_e \cup \{e'' \in \mathcal{G} : \exists e' \in A_e, e' \cap e'' \neq \emptyset\}, \quad (\text{S27})$$

and the weight $w_e \in [0, 1]$. Recalling the definition of δ^* in (S10), we define two random variables

$$U_w^* = \frac{T_w^* - \mathbb{E}(T_w^*|\mathbb{Z})}{\sqrt{\text{Var}(T_w^*|\mathbb{Z})}}, \quad U_{n_1}^* = \frac{n_1^* - n_1}{\sqrt{n_1 n_0/n}}$$

where $T_w^* = \sum_{1 \leq i < j \leq n} w_{(i,j)} I(\delta_i^* = \delta_j^*)$. If (i) the edge weights $\{w_e | e \in \mathcal{G}\}$ satisfy

$$\frac{\max_{e \in \mathcal{G}} w_e}{\sum_{e \in \mathcal{G}} w_e} = o(1);$$

(ii) there exists $1 \leq \beta < 1.5$ such that the number of edges in $\mathcal{G}$ is $|\mathcal{G}| = O(n^\beta)$, $\sum_{e \in \mathcal{G}} |A_e|^2 = o(n^{0.5+\beta})$, and $\sum_{e \in \mathcal{G}} |A_e| \cdot |B_e| = o(n^{1.5\beta})$; then $(U_w^*, U_{n_1}^*)|\mathbb{Z}$ converges in distribution to a bivariate Gaussian random variable as $n \rightarrow \infty$.

By applying Lemma 4, we prove the asymptotic normality of (T_{DW}^*, n_1^*) as follows.

Lemma 5. Let $\mathbb{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ be an i.i.d. sample from distribution H with Lipschitz continuous density h with respect to the measure on $\mathcal{M}$ induced by the Lebesgue measure on the ambient Euclidean space. Assume that there exists $\nu < 9/4$ such that $\sum_{i=1}^n |\mathcal{N}_i|^4 = o(n^\nu)$, where $\mathcal{N}_i = \cup_{S \in \mathcal{DT}_{\mathcal{M}}(\mathbb{Z}) : \mathbf{z}_i \in S} \mathbb{V}(\mathcal{S})$. Let

$$U_{\text{DW}}^* = \frac{T_{\text{DW}}^* - \mathbb{E}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}}, \quad U_{n_1}^* = \frac{n_1^*/n - n_1/n}{\sqrt{n_1 n_0/n^3}},$$

where T_{DW}^* is defined by (S9). Then, $\{(U_{\text{DW}}^*, U_{n_1}^*)|\mathbb{Z}\}$ converges in distribution to a bivariate Gaussian random variable as $n \rightarrow \infty$.

Proof of Lemma 5: Consider the proximity graph $\mathcal{G}(\mathbb{Z}, \mathcal{E})$ constructed on the basis of $\Gamma_{\mathbb{Z}}$ where $\mathcal{E} = \{(i, j) : \check{\gamma}_{ij} > 0\}$ and the weight of edge (i, j) is $\check{\gamma}_{ij}$. We can verify that $\mathcal{G}$ satisfies all the assumptions of Lemma 4 for $\beta = \max\{2\nu/3, 1\} \in [1, 1.5)$ as follows.

- Since $\sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} = n$ while $\check{\gamma}_{ij} \in [0, 2]$ for all $1 \leq i < j \leq n$, we have

$$\frac{\max_{1 \leq i < j \leq n} \check{\gamma}_{ij}}{\sum_{1 \leq i < j \leq n} \check{\gamma}_{ij}} \leq \frac{2}{n} = o(1). \quad (\text{S28})$$

- Since the number of nonzero components in each row of $\Gamma_{\mathbb{Z}}$ is not larger than $d + 1$,

we conclude that

$$|\mathcal{G}| = \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) \leq \sum_{1 \leq i < j \leq n} \{I(\gamma_{ij} > 0) + I(\gamma_{ji} > 0)\} \leq n(d+1) = O(n), \quad (\text{S29})$$

and thus $|\mathcal{G}| = O(n^\beta)$.

- By Lemma 3, we have

$$\begin{aligned} \sum_{e \in \mathcal{G}} |A_e|^2 &= \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) \left[I(\check{\gamma}_{ij} > 0) + \sum_{k \notin \{i, j\}} \{I(\check{\gamma}_{ik} > 0) + I(\check{\gamma}_{jk} > 0)\} \right]^2 \\ &= \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) (d_i + d_j - 1)^2 \\ &= \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) (d_i^2 + d_j^2 + 2d_i d_j - 2d_i - 2d_j + 1) \\ &\leq \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) (2d_i^2 + 2d_j^2 - 2d_i - 2d_j + 1) \\ &= \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) (2d_i^2 + 2d_j^2 - 2d_i - 2d_j) + \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) \\ &= \sum_{i=1}^n \sum_{j \neq i} I(\check{\gamma}_{ij} > 0) (d_i^2 + d_j^2 - d_i - d_j) + \sum_{i=1}^n d_i/2 \\ &= \sum_{i=1}^n \sum_{j \neq i} I(\check{\gamma}_{ij} > 0) (d_i^2 - d_i) + \sum_{j=1}^n \sum_{i \neq j} I(\check{\gamma}_{ij} > 0) (d_j^2 - d_j) + \sum_{i=1}^n d_i/2 \\ &= \sum_{i=1}^n d_i (d_i^2 - d_i) + \sum_{j=1}^n d_j (d_j^2 - d_j) + \sum_{i=1}^n d_i/2 \\ &= \sum_{i=1}^n (2d_i^3 - 2d_i^2 + d_i/2) \\ &= o(n^{1+3(\nu-1)/4}) + o(n^{1+(\nu-1)/2}) + o(n^{1+(\nu-1)/4}) \\ &= o(n^{1+3(\nu-1)/4}) = o(n^{0.5+\beta}), \end{aligned} \quad (\text{S30})$$

since $0.5 + \beta \geq 0.5 + 2\nu/3 > 1 + 3(\nu - 1)/4$ for all $\nu < 9/4$.

- By the definitions of A_e and B_e in (S27), we have $B_e = \cup_{e' \in A_e} A_{e'}$ and thus

$$\begin{aligned}
\sum_{e \in \mathcal{G}} |A_e| \cdot |B_e| &= \sum_{e \in \mathcal{G}} |A_e| \cdot |\cup_{e' \in A_e} A_{e'}| \\
&\leq \sum_{e \in \mathcal{G}} |A_e| \left\{ \sum_{e' \in A_e} |A_{e'}| \right\} \\
&= \sum_{e \in \mathcal{G}} \sum_{e' \in A_e} |A_e| \cdot |A_{e'}| \\
&\leq \sum_{e \in \mathcal{G}} \sum_{e' \in A_e} (|A_e|^2 + |A_{e'}|^2)/2 \\
&= \sum_{e \in \mathcal{G}} \sum_{e' \in A_e} |A_e|^2/2 + \sum_{e' \in \mathcal{G}} \sum_{e \in A_{e'}} |A_{e'}|^2/2 \\
&= \sum_{e \in \mathcal{G}} |A_e|^3/2 + \sum_{e' \in \mathcal{G}} |A_{e'}|^3/2 \\
&= \sum_{e \in \mathcal{G}} |A_e|^3 \\
&= \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) \left[I(\check{\gamma}_{ij} > 0) + \sum_{k \notin \{i, j\}} \{I(\check{\gamma}_{ik} > 0) + I(\check{\gamma}_{jk} > 0)\} \right]^3 \\
&= \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) (d_i + d_j - 1)^3 \\
&\leq \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) (d_i + d_j)^3 \\
&= \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) \{d_i^3 + 3(d_i + d_j)d_i d_j + d_j^3\}.
\end{aligned}$$

Since both d_i and d_j are positive, we have

$$(d_i + d_j)d_i d_j \leq (d_i + d_j)(d_i^2 + d_j^2)/2 = (d_i^3 + d_j^3)/2 + (d_i + d_j)d_i d_j/2,$$

and thus $(d_i + d_j)d_i d_j \leq d_i^3 + d_j^3$. As a result, by Lemma 3, we have

$$\begin{aligned}
\sum_{e \in \mathcal{G}} |A_e| \cdot |B_e| &\leq \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) \{d_i^3 + 3(d_i + d_j)d_i d_j + d_j^3\} \\
&\leq \sum_{1 \leq i < j \leq n} I(\check{\gamma}_{ij} > 0) (4d_i^3 + 4d_j^3)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \sum_{j \neq i} I(\tilde{\gamma}_{ij} > 0)(2d_i^3 + 2d_j^3) \\
&= \sum_{i=1}^n \sum_{j \neq i} I(\tilde{\gamma}_{ij} > 0)2d_i^3 + \sum_{j=1}^n \sum_{i \neq j} I(\tilde{\gamma}_{ij} > 0)2d_j^3 \\
&= \sum_{i=1}^n d_i \times 2d_i^3 + \sum_{j=1}^n d_j \times 2d_j^3 \\
&= \sum_{i=1}^n 4d_i^4 \\
&= o(n^\nu) = o(n^{1.5\beta}).
\end{aligned} \tag{S31}$$

By Lemma 4 and (S28)–(S31), we conclude that $(U_{\text{DW}}^*, U_{n_1^*}^*)|\mathbb{Z}$ converges in distribution to a bivariate Gaussian random variable as $n \rightarrow \infty$.

Thus, we conclude that the conditional distribution $\{U_{\text{DW}}^*|U_{n_1^*}^*, \mathbb{Z}\}$ converges to a Gaussian distribution as $n \rightarrow \infty$ or equivalently,

$$\left\{ \frac{T_{\text{DW}}^* - E_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}} \middle| n_1^*, \mathbb{Z} \right\} \xrightarrow{D} N(\mu, \sigma^2) \tag{S32}$$

as $n \rightarrow \infty$, for some μ and $\sigma^2 > 0$.

S6.4 Asymptotic Null Distribution of T_{DW}

Finally, we derive the asymptotic null distribution of T_{DW} . We first need to prove

$$\frac{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})}{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})} \rightarrow 1, \text{ a.s.} \tag{S33}$$

$$\frac{E_{H_0}(T_{\text{DW}}^*|\mathbb{Z}) - E_{H_0}(T_{\text{DW}}|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}} \rightarrow 0, \text{ a.s.} \tag{S34}$$

as $n \rightarrow \infty$. Since

$$\frac{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})}{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})} = 1 + \frac{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z}) - \text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})},$$

(S33) is equivalent to

$$\frac{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z}) - \text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})} \rightarrow 0, \text{ a.s.} \quad (\text{S35})$$

as $n \rightarrow \infty$.

By Theorem 1 and Lemma 1, we have

$$\begin{aligned} & \text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z}) - \text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z}) \\ &= \left\{ \frac{n_1^2 n_0^2 (n_1 - 1)(n_0 - 1)}{3n_1 n_0 n^2} - \frac{n_1^2 n_0^2 (n - 1)(n - 2)(n - 3)}{3n^5} \right\} V_{0,\mathbb{Z}} \\ & \quad + \left\{ \frac{n_1 n_0 (n - 2)}{3n^2} - \frac{n_1 n_0 (n - 1)(n - 2)}{3n^3} \right\} V_{1,\mathbb{Z}} \\ & \quad + \left\{ \frac{n_1 n_0}{n^2} - \frac{n_1 n_0 (n - 1)}{n^3} \right\} V_{2,\mathbb{Z}} - \left\{ \frac{4n_1^2 n_0^2}{(n - 1)^2 n^2} - \frac{4n_1^2 n_0^2}{n^4} \right\} \\ &= \frac{n_1 n_0 (n_1 - 1)(n_0 - 1)n^3 - n_1^2 n_0^2 (n - 1)(n - 2)(n - 3)}{3n^5} V_{0,\mathbb{Z}} \\ & \quad + \frac{n_1 n_0 (n - 2)}{3n^3} V_{1,\mathbb{Z}} + \frac{n_1 n_0}{n^3} V_{2,\mathbb{Z}} - \frac{4n_1^2 n_0^2 (2n - 1)}{(n - 1)^2 n^4} \\ &= \frac{n_1 n_0 n^3 - n_1 n_0 n^4 + n_1^2 n_0^2 (6n^2 - 11n + 6)}{3n^5} V_{0,\mathbb{Z}} \\ & \quad + \frac{n_1 n_0 (n - 2)}{3n^3} V_{1,\mathbb{Z}} + \frac{n_1 n_0}{n^3} V_{2,\mathbb{Z}} - \frac{4n_1^2 n_0^2 (2n - 1)}{(n - 1)^2 n^4}. \end{aligned}$$

Because $\lim_{n \rightarrow \infty} n_1/n = \kappa$, $\lim_{n \rightarrow \infty} n_0/n = (1 - \kappa)$, $V_{2,\mathbb{Z}}, V_{1,\mathbb{Z}}, V_{0,\mathbb{Z}} \in (0, 1)$,

$$\begin{aligned} \text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z}) &= \frac{n_1^2 n_0^2 (n - 1)(n - 2)(n - 3)}{3n^5} V_{0,\mathbb{Z}} + \frac{n_1 n_0 (n - 1)(n - 2)}{3n^3} V_{1,\mathbb{Z}} \\ & \quad + \frac{n_1 n_0 (n - 1)}{n^3} V_{2,\mathbb{Z}} - \frac{4n_1^2 n_0^2}{n^4}, \end{aligned}$$

and

$$\begin{cases} \frac{n_1 n_0 n^3 - n_1 n_0 n^4 + n_1^2 n_0^2 (6n^2 + 11n - 6)}{n_1^2 n_0^2 (n-1)(n-2)(n-3)} \rightarrow 0, \\ \frac{\frac{n_1 n_0 (n-2)}{3n^3}}{\frac{n_1 n_0 (n-1)(n-2)}{3n^3}} \rightarrow 0, \\ \frac{\frac{n_1 n_0}{n^3}}{\frac{n_1 n_0 (n-1)}{n^3}} \rightarrow 0, \\ \frac{\frac{4n_1^2 n_0^2 (2n-1)}{(n-1)^2 n^4}}{\frac{4n_1^2 n_0^2}{n^4}} \rightarrow 0, \end{cases}$$

as $n \rightarrow \infty$, (S35) holds, and thus (S33) is shown.

In addition, by Lemma 1, we have

$$\begin{aligned} E_{H_0}(T_{\text{DW}}^*|\mathbb{Z}) - E_{H_0}(T_{\text{DW}}|\mathbb{Z}) &= \frac{n_1^2 + n_0^2}{n^2} - \frac{n_1(n_1 - 1) + n_0(n_0 - 1)}{n(n-1)} \\ &= \frac{n_1^2(n-1) + n_0^2(n-1) - n_1^2 n - n_0^2 n + n^2}{n^2(n-1)} \\ &= \frac{n^2 - n_1^2 - n_0^2}{n^2(n-1)} \\ &= \frac{2n_1 n_0}{n^2(n-1)}. \end{aligned} \tag{S36}$$

By (S15), if $\lim_{n \rightarrow \infty} n_1/n = \kappa = 1/2$, we have $\mathcal{P}_0^* - (\mathcal{P}_2^*)^2 = 0$, $\mathcal{P}_2^* - (\mathcal{P}_2^*)^2 = 1/4 + o(1)$,

and

$$\begin{aligned} \mathcal{P}_1^* - (\mathcal{P}_2^*)^2 &= \frac{n_1 n_0}{n^2} - \frac{4n_1^2 n_0^2}{n^4} \\ &= \frac{n_1 n_0 (n^2 - 4n_1 n_0)}{n^4} \\ &= \frac{n_1 n_0 \{(n_1 + n_0)^2 - 4n_1 n_0\}}{n^4} \\ &= \frac{n_1 n_0 (n_1 - n_0)^2}{n^4} \geq 0. \end{aligned}$$

As there are at most $d + 1$ nonzero values in each row of $\mathbf{\Gamma}_{\mathbb{Z}}$ with the sum equal to 1, we have

$$\sum_{j \neq i} I(\gamma_{ij} > 0) \in \{1, \dots, d + 1\}, \text{ and } \sum_{j \neq i} \gamma_{ij} = \sum_{j \neq i} \gamma_{ij} \times I(\gamma_{ij} > 0) = 1,$$

for all $i = 1, \dots, n$. Also, using the Cauchy–Schwartz inequality, we have

$$\left\{ \sum_{j \neq i} \gamma_{ij}^2 \right\} \left\{ \sum_{j \neq i} I(\gamma_{ij} > 0) \right\} = \left\{ \sum_{j \neq i} \gamma_{ij}^2 \right\} \left\{ \sum_{j \neq i} I(\gamma_{ij} > 0)^2 \right\} \geq \left\{ \sum_{j \neq i} \gamma_{ij} \times I(\gamma_{ij} > 0) \right\}^2,$$

for all $i = 1, \dots, n$. Therefore, using (S16) and the results above, if $\lim_{n \rightarrow \infty} n_1/n = \kappa = 1/2$, we have

$$\begin{aligned} \text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z}) &= \sum_{1 \leq i < j \leq n} n^{-2} \tilde{\gamma}_{ij}^2 \times \{\mathcal{P}_2^* - (\mathcal{P}_2^*)^2\} \\ &\quad + \sum_{1 \leq i < j < k \leq n} 2n^{-2} (\tilde{\gamma}_{ij} \tilde{\gamma}_{ik} + \tilde{\gamma}_{ij} \tilde{\gamma}_{jk} + \tilde{\gamma}_{ik} \tilde{\gamma}_{jk}) \times \{\mathcal{P}_1^* - (\mathcal{P}_2^*)^2\} \\ &\quad + \sum_{1 \leq i < j < k < l \leq n} 2n^{-2} (\tilde{\gamma}_{ij} \tilde{\gamma}_{kl} + \tilde{\gamma}_{ik} \tilde{\gamma}_{jl} + \tilde{\gamma}_{il} \tilde{\gamma}_{jk}) \times \{\mathcal{P}_0^* - (\mathcal{P}_2^*)^2\} \\ &\geq \sum_{1 \leq i < j \leq n} n^{-2} \tilde{\gamma}_{ij}^2 \times \{1/4 + o(1)\} \\ &= \sum_{1 \leq i < j \leq n} \left\{ \frac{1}{4n^2} + o(n^{-2}) \right\} \tilde{\gamma}_{ij}^2 \\ &= \sum_{1 \leq i < j \leq n} \left\{ \frac{1}{4n^2} + o(n^{-2}) \right\} (\gamma_{ij} + \gamma_{ji})^2 \\ &\geq \left\{ \frac{1}{4n^2} + o(n^{-2}) \right\} \sum_{1 \leq i < j \leq n} (\gamma_{ij}^2 + \gamma_{ji}^2) \\ &= \left\{ \frac{1}{4n^2} + o(n^{-2}) \right\} \sum_{i=1}^n \sum_{j \neq i} \gamma_{ij}^2 \\ &\geq \left\{ \frac{1}{4n^2} + o(n^{-2}) \right\} \sum_{i=1}^n \frac{\{\sum_{j \neq i} \gamma_{ij} \times I(\gamma_{ij} > 0)\}^2}{\sum_{j \neq i} I(\gamma_{ij} > 0)} \\ &= \left\{ \frac{1}{4n^2} + o(n^{-2}) \right\} \sum_{i=1}^n \frac{1}{\sum_{j \neq i} I(\gamma_{ij} > 0)} \\ &\geq \left\{ \frac{1}{4n^2} + o(n^{-2}) \right\} \sum_{i=1}^n \frac{1}{d+1} \\ &= \left\{ \frac{1}{4n^2} + o(n^{-2}) \right\} \times \frac{n}{d+1} \\ &= \frac{1}{4n(d+1)} + o(n^{-1}). \end{aligned} \tag{S37}$$

If $\lim_{n \rightarrow \infty} n_1/n = \kappa \neq 1/2$, we have

$$\begin{aligned}\mathcal{P}_0^* - (\mathcal{P}_2^*)^2 &= 0, \\ \mathcal{P}_1^* - (\mathcal{P}_2^*)^2 &= \frac{n_1 n_0}{n^2} - \frac{4n_1^2 n_0^2}{n^4} = \kappa(1 - \kappa)\{1 - 4\kappa(1 - \kappa)\} + o(1), \\ \mathcal{P}_2^* - (\mathcal{P}_2^*)^2 &= \frac{2n_1 n_0}{n^2} - \frac{4n_1^2 n_0^2}{n^4} \geq 2\{\mathcal{P}_1^* - (\mathcal{P}_2^*)^2\}.\end{aligned}$$

As $\sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} = n$, using the Cauchy–Schwartz inequality, we have

$$\begin{aligned}n \left[\sum_{i=1}^n \left\{ \sum_{j \neq i} \check{\gamma}_{ij} \right\}^2 \right] &= \left[\sum_{i=1}^n 1^2 \right] \left[\sum_{i=1}^n \left\{ \sum_{j \neq i} \check{\gamma}_{ij} \right\}^2 \right] \\ &\geq \left\{ \sum_{i=1}^n \sum_{j \neq i} \check{\gamma}_{ij} \right\}^2 = \left\{ 2 \sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} \right\}^2 = 4n^2.\end{aligned}$$

Using (S16) and the results above, if $\lim_{n \rightarrow \infty} n_1/n = \kappa \neq 1/2$, we have

$$\begin{aligned}\text{Var}_{H_0}(T_{\text{DW}}^* | \mathbb{Z}) &= \sum_{1 \leq i < j \leq n} n^{-2} \check{\gamma}_{ij}^2 \times \{\mathcal{P}_2^* - (\mathcal{P}_2^*)^2\} \\ &\quad + \sum_{1 \leq i < j < k \leq n} 2n^{-2} (\check{\gamma}_{ij} \check{\gamma}_{ik} + \check{\gamma}_{ij} \check{\gamma}_{jk} + \check{\gamma}_{ik} \check{\gamma}_{jk}) \times \{\mathcal{P}_1^* - (\mathcal{P}_2^*)^2\} \\ &\quad + \sum_{1 \leq i < j < k < l \leq n} 2n^{-2} (\check{\gamma}_{ij} \check{\gamma}_{kl} + \check{\gamma}_{ik} \check{\gamma}_{jl} + \check{\gamma}_{il} \check{\gamma}_{jk}) \times \{\mathcal{P}_0^* - (\mathcal{P}_2^*)^2\} \\ &\geq \sum_{1 \leq i < j \leq n} n^{-2} \check{\gamma}_{ij}^2 \times 2\{\mathcal{P}_1^* - (\mathcal{P}_2^*)^2\} \\ &\quad + \sum_{1 \leq i < j < k \leq n} 2n^{-2} (\check{\gamma}_{ij} \check{\gamma}_{ik} + \check{\gamma}_{ij} \check{\gamma}_{jk} + \check{\gamma}_{ik} \check{\gamma}_{jk}) \times \{\mathcal{P}_1^* - (\mathcal{P}_2^*)^2\} \\ &= n^{-2} \left\{ \sum_{1 \leq i < j \leq n} 2\check{\gamma}_{ij}^2 + \sum_{1 \leq i < j < k \leq n} 2(\check{\gamma}_{ij} \check{\gamma}_{ik} + \check{\gamma}_{ij} \check{\gamma}_{jk} + \check{\gamma}_{ik} \check{\gamma}_{jk}) \right\} \\ &\quad \times [\kappa(1 - \kappa)\{1 - 4\kappa(1 - \kappa)\} + o(1)] \\ &= n^{-2} \sum_{i=1}^n \left\{ \sum_{j \neq i} \check{\gamma}_{ij} \right\}^2 \times [\kappa(1 - \kappa)\{1 - 4\kappa(1 - \kappa)\} + o(1)] \\ &\geq n^{-2} \times 4n \times [\kappa(1 - \kappa)\{1 - 4\kappa(1 - \kappa)\} + o(1)]\end{aligned}$$

$$= \frac{4\kappa(1-\kappa)\{1-4\kappa(1-\kappa)\}}{n} + o(n^{-1}). \quad (\text{S38})$$

By (S36)–(S38), we conclude (S34). By (S11), we have under H_0 ,

$$\left\{ \frac{T_{\text{DW}} - E_{H_0}(T_{\text{DW}}|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})}} \middle| \mathbb{Z} \right\} \stackrel{D}{=} \left\{ \frac{T_{\text{DW}}^* - E_{H_0}(T_{\text{DW}}|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})}} \middle| n_1^* = n_1, \mathbb{Z} \right\},$$

where the right-hand side equals

$$\begin{aligned} & \left\{ \frac{T_{\text{DW}}^* - E_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})}} \middle| n_1^* = n_1, \mathbb{Z} \right\} + \frac{E_{H_0}(T_{\text{DW}}^*|\mathbb{Z}) - E_{H_0}(T_{\text{DW}}|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})}} \\ &= \sqrt{\frac{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})}} \times \left\{ \frac{T_{\text{DW}}^* - E_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}} \middle| n_1^* = n_1, \mathbb{Z} \right\} \\ &+ \sqrt{\frac{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})}} \times \frac{E_{H_0}(T_{\text{DW}}^*|\mathbb{Z}) - E_{H_0}(T_{\text{DW}}|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}}, \end{aligned}$$

which, according to (S33) and (S34), has the same asymptotic distribution as

$$\left\{ \frac{T_{\text{DW}}^* - E_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}^*|\mathbb{Z})}} \middle| n_1^* = n_1, \mathbb{Z} \right\}$$

as $n \rightarrow \infty$. The quantity above converges to a Gaussian distribution according to (S32).

Hence,

$$\left\{ \frac{T_{\text{DW}} - E_{H_0}(T_{\text{DW}}|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})}} \middle| \mathbb{Z} \right\} \xrightarrow{D} N(0, 1), \quad (\text{S39})$$

as $n \rightarrow \infty$. Finally, using the law of total probability and noting that $N(0, 1)$ is independent of $\mathbb{Z}$, we conclude the result of Theorem 2.

S7 Proof of Theorem 3

Lemma 6. *Under the same conditions as in Theorem 2, if $\mathbb{X}$, $\mathbb{Y}$ and $\{\mathbf{z}_{n+1}\}$ are disjoint sets in $\mathcal{M}$ such that $\mathbb{X} \cup \mathbb{Y} \cup \{\mathbf{z}_{n+1}\}$ is generic, we have*

$$E_{H_1}(|T_{\text{DW}}^\Delta - T_{\text{DW}}|) = o(n^{-11/16}), \quad (\text{S40})$$

as $n \rightarrow \infty$, where T_{DW}^Δ is the Delaunay weighted test statistic with respect to $\mathbb{X} \cup \mathbb{Y}$ and $\{(\mathbf{z}_{n+1}, \delta_{n+1})\}$.

Proof of Lemma 6: By Lemma 2 that the proximity graph $\mathcal{G}(\mathbb{Z} = \mathbb{X} \cup \mathbb{Y}, \mathcal{E})$ defined by (S19) is a subgraph of the auxiliary graph $\mathcal{G}^*(\mathbb{Z}, \mathcal{E}^*)$ a.s. as $n \rightarrow \infty$, we consider the auxiliary graph $\mathcal{G}^*(\mathbb{Z}, \mathcal{E}^*)$.

For any fixed i , by the duality between the Voronoi diagram and the Delaunay triangulation, we have $(i, j) \in \mathcal{E}^*$ if and only if $\mathcal{V}_i(\mathbb{Z}) \cap \mathcal{V}_j(\mathbb{Z}) \neq \emptyset$, where

$$\mathcal{V}_i(\mathbb{Z}) = \{\mathbf{z} \in \mathcal{M} : d_{\mathcal{M}}(\mathbf{z}, \mathbf{z}_i) \leq \min_{j=1, \dots, n; j \neq i} d_{\mathcal{M}}(\mathbf{z}, \mathbf{z}_j)\}$$

is the Voronoi cell with respect to $\mathbf{z}_i \in \mathbb{Z}$. Thus, $\mathbf{z}_i$'s neighbors in $\mathcal{G}^*(\mathbb{Z}, \mathcal{E}^*)$ are $\{\mathbf{z}_j : \mathcal{V}_i(\mathbb{Z}) \cap \mathcal{V}_j(\mathbb{Z}) \neq \emptyset\}$. If $\mathcal{V}_i(\mathbb{Z} \cup \{\mathbf{z}_{n+1}\}) \cap \mathcal{V}_{n+1}(\mathbb{Z} \cup \{\mathbf{z}_{n+1}\}) = \emptyset$, we have $\mathcal{V}_i(\mathbb{Z} \cup \{\mathbf{z}_{n+1}\})$ remains the same as $\mathcal{V}_i(\mathbb{Z})$, and thus $\{\mathbf{z}_j : \mathcal{V}_i(\mathbb{Z} \cup \{\mathbf{z}_{n+1}\}) \cap \mathcal{V}_j(\mathbb{Z} \cup \{\mathbf{z}_{n+1}\}) \neq \emptyset\} = \{\mathbf{z}_j : \mathcal{V}_i(\mathbb{Z}) \cap \mathcal{V}_j(\mathbb{Z}) \neq \emptyset\}$.

In other words, if $\mathbf{z}_i$ is not $\mathbf{z}_{n+1}$'s neighbor in $\mathcal{G}^*(\mathbb{Z} \cup \{\mathbf{z}_{n+1}\}, \mathcal{E}_{n+1}^*)$ (or equivalently, $(i, n+1) \notin \mathcal{E}_{n+1}^*$), its neighbors in $\mathcal{G}^*(\mathbb{Z} \cup \{\mathbf{z}_{n+1}\}, \mathcal{E}_{n+1}^*)$ remain the same as its neighbors in $\mathcal{G}^*(\mathbb{Z}, \mathcal{E}^*)$, where $\mathcal{E}_{n+1}^* = \{(i, j) : \exists \mathcal{S} \in \mathcal{DT}_{\mathcal{M}}(\mathbb{Z} \cup \{\mathbf{z}_{n+1}\}) \text{ s.t. } \mathbf{z}_i, \mathbf{z}_j \in \mathbb{V}(\mathcal{S})\}$. Therefore, if $\mathbf{z}_i$ is not $\mathbf{z}_{n+1}$'s neighbor in $\mathcal{G}^*(\mathbb{Z} \cup \{\mathbf{z}_{n+1}\}, \mathcal{E}_{n+1}^*)$, all nonzero elements $\{\gamma_{ij} > 0 | j \neq i\}$ in the i -th row of $\mathbf{\Gamma}_{\mathbb{Z} \cup \{\mathbf{z}_{n+1}\}}$ remain the same as the i -th row of $\mathbf{\Gamma}_{\mathbb{Z}}$ a.s. as $n \rightarrow \infty$. According to the statement that $\sum_{i=1}^n |\mathcal{N}_i| = o(n^{1+(\nu-1)/4})$ in proof of Lemma 3, the expected number of $\mathbf{z}_{n+1}$'s neighbors in the graph $\mathcal{G}^*(\mathbb{Z} \cup \{\mathbf{z}_{n+1}\}, \mathcal{E}_{n+1}^*)$ is $o(n^{(\nu-1)/4}) = o(n^{5/16})$, due to $\nu < 9/4$. As a result, on average, only nonzero elements in $o(n^{5/16})$ rows among the first n rows of $\mathbf{\Gamma}_{\mathbb{Z} \cup \{\mathbf{z}_{n+1}\}}$ can be different from $\mathbf{\Gamma}_{\mathbb{Z}}$. Because the sum of each row of $\mathbf{\Gamma}_{\mathbb{Z}}$ is fixed as 1, we

conclude (S40), and Lemma 6 is proven.

Proof of Theorem 3: To prove the consistency of our Delaunay weighted test, we need to show that under $H_1 : F \neq G$,

$$\frac{T_{\text{DW}} - E_{H_0}(T_{\text{DW}}|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})}} \rightarrow \infty, \text{ a.s.} \quad (\text{S41})$$

as $n \rightarrow \infty$.

In Theorem 2 of Henze and Penrose (1999), the authors showed that

$$\begin{aligned} n^{-1}R_n &= n^{-1} \sum_{j=1}^n \sum_{i \neq j} I(\text{MST of } \mathbb{Z} \text{ contains an edge of } \mathbf{z}_i \text{ and } \mathbf{z}_j) I(\delta_i \neq \delta_j) \\ &\rightarrow \int \frac{2\kappa(1-\kappa)f(\mathbf{z})g(\mathbf{z})}{\kappa f(\mathbf{z}) + (1-\kappa)g(\mathbf{z})} d\mathbf{z}, \text{ a.s.} \end{aligned}$$

as $n \rightarrow \infty$. We intend to follow their Poissonization approach to show that

$$\begin{aligned} T_{\text{DW}} &= n^{-1} \sum_{j=1}^n \sum_{i \neq j} \gamma_{ij} \times I(\delta_i = \delta_j) \\ &\rightarrow \int_{\mathbf{z} \in \mathcal{M}} \frac{\kappa^2 f^2(\mathbf{z}) + (1-\kappa)^2 g^2(\mathbf{z})}{\kappa f(\mathbf{z}) + (1-\kappa)g(\mathbf{z})} d\mathbf{z}, \text{ a.s.} \end{aligned} \quad (\text{S42})$$

as $n \rightarrow \infty$.

Specifically, let N_1 and N_0 be Poisson random variables with means n_1 and n_0 respectively, which are independent of each other and $\mathbb{X}^\dagger = \{\mathbf{x}_1, \dots, \mathbf{x}_{N_1}\}$ and $\mathbb{Y}^\dagger = \{\mathbf{y}_1, \dots, \mathbf{y}_{N_0}\}$ on $\mathcal{M}$. As a result, $\mathbb{X}^\dagger$ and $\mathbb{Y}^\dagger$ can be treated as observations from nonhomogeneous Poisson processes of rates $n_1 f(\cdot)$ and $n_0 g(\cdot)$ respectively. Analogous to the derivation in Henze and Penrose (1999), when $T_{\text{DW}}^\dagger$ is deduced from the Deluanay weight matrix $\mathbf{\Gamma}_{\mathbb{Z}^\dagger}$ of the pooled sample $\mathbb{Z}^\dagger = \mathbb{X}^\dagger \cup \mathbb{Y}^\dagger$ and the group indicator $\boldsymbol{\delta}^\dagger$ in the same way as (4), we have

$$\begin{aligned} E_{H_1}(|T_{\text{DW}}^\dagger - T_{\text{DW}}|) &= E_{H_1}\{E_{H_1}(|T_{\text{DW}}^\dagger - T_{\text{DW}}| | N_1, N_0)\} \\ &\leq E_{H_1}\{E_{H_1}(|T_{\text{DW}}^\Delta - T_{\text{DW}}|) \cdot (|N_1 - n_1| + |N_0 - n_0|)\} \\ &= E_{H_1}\{(|N_1 - n_1| + |N_0 - n_0|)\} \cdot o(n^{-11/16}) \\ &= O(n^{1/2}) \cdot o(n^{-11/16}) \rightarrow 0, \end{aligned} \quad (\text{S43})$$

as $n \rightarrow \infty$ with $\lim_{n \rightarrow \infty} n_1/n = \kappa$, where the second equality follows from Lemma 6.

Let $\mathbb{Z}^\dagger = \{\mathbf{z}_1^\dagger, \dots, \mathbf{z}_N^\dagger\}$ be an i.i.d. sample from the distribution $\{n_1 F(\cdot) + n_0 G(\cdot)\}/n$ on $\mathcal{M}$, and let $\delta_1^\dagger, \dots, \delta_N^\dagger$ be independent Bernoulli random variables defined by

$$\Pr(\delta_i^\dagger = 1 | \mathbf{z}_i^\dagger) = \frac{n_1 f(\mathbf{z}_i^\dagger)}{n_1 f(\mathbf{z}_i^\dagger) + n_0 g(\mathbf{z}_i^\dagger)}, \quad i = 1, \dots, N, \quad (\text{S44})$$

where N is a Poisson variable that is independent of all the $\mathbf{z}_i^\dagger$'s with mean n . Similar to $T_{\text{DW}}^\dagger$, we define $T_{\text{DW}}^\dagger$ based on $\mathbb{Z}^\dagger$ and $\delta_1^\dagger, \dots, \delta_N^\dagger$. Following the same argument as in the proof of Theorem 2 of Henze and Penrose (1999) that $T_{\text{DW}}^\dagger$ and $T_{\text{DW}}^\dagger$ follow the same distribution and (S43), it suffices to prove (S42) with T_{DW} replaced by $T_{\text{DW}}^\dagger$.

Using (S44), conditional on $\mathbb{Z}^\dagger$, the probability of $\{\delta_i^\dagger = \delta_j^\dagger\}$ (for any $i \neq j$) is

$$\begin{aligned} \psi_{n_1, n_0}(\mathbf{z}_i^\dagger, \mathbf{z}_j^\dagger) &\equiv \Pr(\delta_i^\dagger = \delta_j^\dagger | \mathbf{z}_i^\dagger, \mathbf{z}_j^\dagger) \\ &= \Pr(\delta_i^\dagger = 1 | \mathbf{z}_i^\dagger) \Pr(\delta_j^\dagger = 1 | \mathbf{z}_j^\dagger) + \Pr(\delta_i^\dagger = 0 | \mathbf{z}_i^\dagger) \Pr(\delta_j^\dagger = 0 | \mathbf{z}_j^\dagger) \\ &= \frac{n_1^2 f(\mathbf{z}_i^\dagger) f(\mathbf{z}_j^\dagger)}{\{n_1 f(\mathbf{z}_i^\dagger) + n_0 g(\mathbf{z}_i^\dagger)\} \{n_1 f(\mathbf{z}_j^\dagger) + n_0 g(\mathbf{z}_j^\dagger)\}} \\ &\quad + \frac{n_0^2 g(\mathbf{z}_i^\dagger) g(\mathbf{z}_j^\dagger)}{\{n_1 f(\mathbf{z}_i^\dagger) + n_0 g(\mathbf{z}_i^\dagger)\} \{n_1 f(\mathbf{z}_j^\dagger) + n_0 g(\mathbf{z}_j^\dagger)\}} \\ &= \frac{n_1^2 f(\mathbf{z}_i^\dagger) f(\mathbf{z}_j^\dagger) + n_0^2 g(\mathbf{z}_i^\dagger) g(\mathbf{z}_j^\dagger)}{\{n_1 f(\mathbf{z}_i^\dagger) + n_0 g(\mathbf{z}_i^\dagger)\} \{n_1 f(\mathbf{z}_j^\dagger) + n_0 g(\mathbf{z}_j^\dagger)\}}, \end{aligned}$$

and thus

$$\mathbb{E}_{H_1}(T_{\text{DW}}^\dagger | \mathbb{Z}^\dagger) = \sum_{1 \leq i < j \leq N} n^{-1} \tilde{\gamma}_{ij}^\dagger \psi_{n_1, n_0}(\mathbf{z}_i^\dagger, \mathbf{z}_j^\dagger),$$

where $\tilde{\gamma}_{ij}^\dagger = \gamma_{ij}^\dagger + \gamma_{ji}^\dagger$. Again, following the same argument as in the proof of Theorem 2 of Henze and Penrose (1999), we conclude that

$$\mathbb{E}_{H_1}(T_{\text{DW}}^\dagger) = \mathbb{E}_{H_1} \left\{ \sum_{1 \leq i < j \leq n} n^{-1} \tilde{\gamma}_{ij}^\dagger \psi(\mathbf{z}_i^\dagger, \mathbf{z}_j^\dagger) \right\} + o(1), \quad (\text{S45})$$

where

$$\psi(\mathbf{z}_i^\dagger, \mathbf{z}_j^\dagger) = \frac{\kappa^2 f(\mathbf{z}_i^\dagger) f(\mathbf{z}_j^\dagger) + (1 - \kappa)^2 g(\mathbf{z}_i^\dagger) g(\mathbf{z}_j^\dagger)}{\{\kappa f(\mathbf{z}_i^\dagger) + (1 - \kappa) g(\mathbf{z}_i^\dagger)\} \{\kappa f(\mathbf{z}_j^\dagger) + (1 - \kappa) g(\mathbf{z}_j^\dagger)\}}.$$

We consider

$$\begin{aligned} \mathbb{E}_{H_1} \left\{ \sum_{1 \leq i < j \leq n} n^{-1} \tilde{\gamma}_{ij}^\dagger \psi(\mathbf{z}_i^\dagger, \mathbf{z}_j^\dagger) \right\} &= \mathbb{E}_{H_1} \left\{ n^{-1} \sum_{i=1}^n \sum_{j \neq i} \gamma_{ij}^\dagger \psi(\mathbf{z}_i^\dagger, \mathbf{z}_j^\dagger) \right\} \\ &= \mathbb{E}_{H_1} \left\{ \sum_{j=2}^n \gamma_{1j}^\dagger \psi(\mathbf{z}_1^\dagger, \mathbf{z}_j^\dagger) \right\} \\ &= \mathbb{E}_{H_1} \left[\mathbb{E}_{H_1} \left\{ \sum_{j=2}^n \gamma_{1j}^\dagger \psi(\mathbf{z}_1^\dagger, \mathbf{z}_j^\dagger) \middle| \mathbf{z}_1^\dagger \right\} \right]. \end{aligned} \quad (\text{S46})$$

Analogous to deriving (11) of Proposition 1 of [Henze and Penrose \(1999\)](#), we have

$$\mathbb{E}_{H_1} \left\{ \sum_{j=2}^n \gamma_{1j}^\dagger \psi(\mathbf{z}_1^\dagger, \mathbf{z}_j^\dagger) \middle| \mathbf{z}_1^\dagger \right\} = \psi(\mathbf{z}_1^\dagger, \mathbf{z}_1^\dagger) \mathbb{E}_{H_1} \left\{ \sum_{j=2}^n \gamma_{1j}^\dagger \middle| \mathbf{z}_1^\dagger \right\} + o(1), \quad (\text{S47})$$

which equals $\psi(\mathbf{z}_1^\dagger, \mathbf{z}_1^\dagger) + o(1)$ due to $\sum_{j=2}^N \gamma_{1j}^\dagger = 1$ by Definition 2 of the Delaunay weight.

Combining (S45)–(S47), we have

$$\begin{aligned} \mathbb{E}_{H_1}(T_{\text{DW}}^\dagger) &\rightarrow \mathbb{E}_{H_1} \{ \psi(\mathbf{z}_1^\dagger, \mathbf{z}_1^\dagger) \} \\ &= \int_{\mathbf{z} \in \mathcal{M}} \psi(\mathbf{z}, \mathbf{z}) \{ \kappa f(\mathbf{z}) + (1 - \kappa) g(\mathbf{z}) \} d\mathbf{z} = \int_{\mathbf{z} \in \mathcal{M}} \frac{\kappa^2 f^2(\mathbf{z}) + (1 - \kappa)^2 g^2(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa) g(\mathbf{z})} d\mathbf{z}, \end{aligned}$$

as $n \rightarrow \infty$. Applying Theorem 2.3 of [Henze and Voigt \(1992\)](#), which proves the almost sure convergence of symmetric multi-sample statistics, we conclude (S42).

We next prove (S41) by showing that under $H_1 : F \neq G$, the right-hand side of (S42) is larger than $\mathbb{E}_{H_0}(T_{\text{DW}}|\mathbb{Z})$ and $\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})$ tends to 0 a.s. as $n \rightarrow \infty$.

By (S42), Theorem 1 that

$$\mathbb{E}_{H_0}(T_{\text{DW}}|\mathbb{Z}) = \frac{n_1(n_1 - 1) + n_0(n_0 - 1)}{n(n - 1)} \rightarrow \kappa^2 + (1 - \kappa)^2 \quad (\text{S48})$$

as $n \rightarrow \infty$, and

$$\begin{aligned}
1 &= (1 - \kappa) + \kappa \\
&= \int_{\mathbf{z} \in \mathcal{M}} (1 - \kappa) f(\mathbf{z}) d\mathbf{z} + \int_{\mathbf{z} \in \mathcal{M}} \kappa g(\mathbf{z}) d\mathbf{z} \\
&= \int_{\mathbf{z} \in \mathcal{M}} \frac{\kappa(1 - \kappa)f^2(\mathbf{z}) + (1 - \kappa)^2 f(\mathbf{z})g(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} + \int_{\mathbf{z} \in \mathcal{M}} \frac{\kappa^2 f(\mathbf{z})g(\mathbf{z}) + \kappa(1 - \kappa)g^2(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} \\
&= \int_{\mathbf{z} \in \mathcal{M}} \frac{\kappa(1 - \kappa)f^2(\mathbf{z}) + \{\kappa^2 + (1 - \kappa)^2\}f(\mathbf{z})g(\mathbf{z}) + \kappa(1 - \kappa)g^2(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z},
\end{aligned}$$

we have under $H_1 : F \neq G$,

$$\begin{aligned}
&T_{\text{DW}} - E_{H_0}(T_{\text{DW}}|\mathbb{Z}) \\
&\rightarrow \left\{ \int_{\mathbf{z} \in \mathcal{M}} \frac{\kappa^2 f^2(\mathbf{z}) + (1 - \kappa)^2 g^2(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} \right\} - \{\kappa^2 + (1 - \kappa)^2\} \text{ a.s.} \\
&= \left\{ \int_{\mathbf{z} \in \mathcal{M}} \frac{\kappa^2 f^2(\mathbf{z}) + 2\kappa(1 - \kappa)f(\mathbf{z})g(\mathbf{z}) + (1 - \kappa)^2 g^2(\mathbf{z}) - 2\kappa(1 - \kappa)f(\mathbf{z})g(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} \right\} \\
&\quad - \{\kappa^2 + 2\kappa(1 - \kappa) + (1 - \kappa)^2 - 2\kappa(1 - \kappa)\} \\
&= \left[\int_{\mathbf{z} \in \mathcal{M}} \frac{\{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})\}^2 - 2\kappa(1 - \kappa)f(\mathbf{z})g(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} \right] \\
&\quad - \{(\kappa + 1 - \kappa)^2 - 2\kappa(1 - \kappa)\} \\
&= \left[\int_{\mathbf{z} \in \mathcal{M}} \{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})\} d\mathbf{z} - \int_{\mathbf{z} \in \mathcal{M}} \frac{2\kappa(1 - \kappa)f(\mathbf{z})g(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} \right] - \{1 - 2\kappa(1 - \kappa)\} \\
&= \left\{ 1 - 2\kappa(1 - \kappa) \int_{\mathbf{z} \in \mathcal{M}} \frac{f(\mathbf{z})g(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} \right\} - \{1 - 2\kappa(1 - \kappa)\} \\
&= 2\kappa(1 - \kappa) \left\{ 1 - \int_{\mathbf{z} \in \mathcal{M}} \frac{f(\mathbf{z})g(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} \right\} \\
&= 2\kappa(1 - \kappa) \left[\int_{\mathbf{z} \in \mathcal{M}} \frac{\kappa(1 - \kappa)f^2(\mathbf{z}) + \{\kappa^2 + (1 - \kappa)^2 - 1\}f(\mathbf{z})g(\mathbf{z}) + \kappa(1 - \kappa)g^2(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} \right] \\
&= 2\kappa(1 - \kappa) \left\{ \int_{\mathbf{z} \in \mathcal{M}} \frac{\kappa(1 - \kappa)f^2(\mathbf{z}) - 2\kappa(1 - \kappa)f(\mathbf{z})g(\mathbf{z}) + \kappa(1 - \kappa)g^2(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} \right\} \\
&= 2\kappa^2(1 - \kappa)^2 \left\{ \int_{\mathbf{z} \in \mathcal{M}} \frac{f^2(\mathbf{z}) - 2f(\mathbf{z})g(\mathbf{z}) + g^2(\mathbf{z})}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} \right\}
\end{aligned}$$

$$= 2\kappa^2(1 - \kappa)^2 \left[\int_{\mathbf{z} \in \mathcal{M}} \frac{\{f(\mathbf{z}) - g(\mathbf{z})\}^2}{\kappa f(\mathbf{z}) + (1 - \kappa)g(\mathbf{z})} d\mathbf{z} \right] > 0 \quad (\text{S49})$$

as $n \rightarrow \infty$.

Because $\lim_{n \rightarrow \infty} n_1/n = \kappa$ and $\lim_{n \rightarrow \infty} n_0/n = 1 - \kappa$, recalling the expressions of $\mathcal{P}_1$ to $\mathcal{P}_3$ from (S3)–(S5), we have

$$\begin{aligned} \mathcal{P}_0 - \mathcal{P}_2^2 &= \frac{4n_1(n_1 - 1)n_0(n_0 - 1)}{n(n - 1)(n - 2)(n - 3)} - \frac{4n_1^2 n_0^2}{n^2(n - 1)^2} \\ &= \frac{4n_1 n_0 \{n(n - 1)(n_1 - 1)(n_0 - 1) - (n - 2)(n - 3)n_1 n_0\}}{n^2(n - 1)^2(n - 2)(n - 3)}, \\ &= \frac{4n_1 n_0 \{-n^2(n_1 + n_0) - nn_1 n_0 + n(n + n_1 + n_0 - 1) + 5nn_1 n_0 - 6n_1 n_0\}}{n^2(n - 1)^2(n - 2)(n - 3)}, \\ &= \frac{4n_1 n_0 \{-n^3 + 2n^2 - n + 4nn_1 n_0 - 6n_1 n_0\}}{n^2(n - 1)^2(n - 2)(n - 3)}, \\ &= O(n^{-1}), \\ \mathcal{P}_1 - \mathcal{P}_2^2 &= \frac{n_1 n_0}{n(n - 1)} - \frac{4n_1^2 n_0^2}{n^2(n - 1)^2} \leq (\mathcal{P}_2 - \mathcal{P}_2^2)/2 \leq 1/2, \\ \mathcal{P}_2 - \mathcal{P}_2^2 &= \frac{2n_1 n_0}{n(n - 1)} - \frac{4n_1^2 n_0^2}{n^2(n - 1)^2} \leq 1. \end{aligned}$$

By (S7) and

$$\sum_{1 \leq i < j < k < l \leq n} 2n^{-2}(\check{\gamma}_{ij}\check{\gamma}_{kl} + \check{\gamma}_{ik}\check{\gamma}_{jl} + \check{\gamma}_{il}\check{\gamma}_{jk}) \leq n^{-2} \left(\sum_{1 \leq i < j \leq n} \check{\gamma}_{ij} \right)^2 = 1,$$

we have

$$\begin{aligned} \text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z}) &= \sum_{1 \leq i < j \leq n} n^{-2} \check{\gamma}_{ij}^2 \times (\mathcal{P}_2 - \mathcal{P}_2^2) \\ &\quad + \sum_{1 \leq i < j < k \leq n} 2n^{-2}(\check{\gamma}_{ij}\check{\gamma}_{ik} + \check{\gamma}_{ij}\check{\gamma}_{jk} + \check{\gamma}_{ik}\check{\gamma}_{jk}) \times (\mathcal{P}_1 - \mathcal{P}_2^2) \\ &\quad + \sum_{1 \leq i < j < k < l \leq n} 2n^{-2}(\check{\gamma}_{ij}\check{\gamma}_{kl} + \check{\gamma}_{ik}\check{\gamma}_{jl} + \check{\gamma}_{il}\check{\gamma}_{jk}) \times (\mathcal{P}_0 - \mathcal{P}_2^2), \\ &\leq \sum_{1 \leq i < j \leq n} n^{-2} \check{\gamma}_{ij}^2 + \sum_{1 \leq i < j < k \leq n} n^{-2}(\check{\gamma}_{ij}\check{\gamma}_{ik} + \check{\gamma}_{ij}\check{\gamma}_{jk} + \check{\gamma}_{ik}\check{\gamma}_{jk}) + (\mathcal{P}_0 - \mathcal{P}_2^2) \end{aligned}$$

$$\begin{aligned}
&= \sum_{1 \leq i < j \leq n} n^{-2} \check{\gamma}_{ij}^2 + \sum_{1 \leq i < j < k \leq n} n^{-2} (\check{\gamma}_{ij} \check{\gamma}_{ik} + \check{\gamma}_{ij} \check{\gamma}_{jk} + \check{\gamma}_{ik} \check{\gamma}_{jk}) + O(n^{-1}) \\
&= \frac{1}{2n^2} \left\{ \sum_{1 \leq i < j \leq n} 2\check{\gamma}_{ij}^2 + \sum_{1 \leq i < j < k \leq n} 2(\check{\gamma}_{ij} \check{\gamma}_{ik} + \check{\gamma}_{ij} \check{\gamma}_{jk} + \check{\gamma}_{ik} \check{\gamma}_{jk}) \right\} + O(n^{-1}) \\
&= \frac{1}{2n^2} \sum_{i=1}^n \left(\sum_{j \neq i} \check{\gamma}_{ij} \right)^2 + O(n^{-1}). \tag{S50}
\end{aligned}$$

Since $\gamma_{ij} \in [0, 1]$ and $\check{\gamma}_{ij} = \gamma_{ij} + \gamma_{ji}$, we have $\check{\gamma}_{ij} \leq 2 \times I(\check{\gamma}_{ij} > 0)$ for all $1 \leq i < j \leq n$.

Thus, from (S50) and Lemma 3, we have

$$\begin{aligned}
\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z}) &\leq \frac{1}{2n^2} \sum_{i=1}^n \left(\sum_{j \neq i} \check{\gamma}_{ij} \right)^2 + O(n^{-1}) \\
&\leq \frac{1}{2n^2} \sum_{i=1}^n \left\{ \sum_{j \neq i} 2 \times I(\check{\gamma}_{ij} > 0) \right\}^2 + O(n^{-1}) \\
&= \frac{2}{n^2} \sum_{i=1}^n \left\{ \sum_{j \neq i} I(\check{\gamma}_{ij} > 0) \right\}^2 + O(n^{-1}) \\
&= \frac{2}{n^2} \sum_{i=1}^n d_i^2 + O(n^{-1}) \\
&= o(n^{(\nu-1)/2-1}) + O(n^{-1}) = o(n^{-3/8}), \tag{S51}
\end{aligned}$$

because $\nu < 9/4$. Thus by (S49) and (S51), (S41) holds under $H_1 : F \neq G$.

Since $T_{\text{DW}}^{(1)}, \dots, T_{\text{DW}}^{(B)}$ generated from the permutation procedure in Section 3.4 follow the null distribution of T_{DW} , according to Theorem 2, we have, for $i = 1, \dots, B$,

$$\frac{T_{\text{DW}}^{(i)} - \mathbb{E}_{H_0}(T_{\text{DW}}|\mathbb{Z})}{\sqrt{\text{Var}_{H_0}(T_{\text{DW}}|\mathbb{Z})}} \xrightarrow{D} N(0, 1) \tag{S52}$$

as $n \rightarrow \infty$. Using (S41), (S52) and the standard consistency result of the permutation test (Hoeffding, 1952), the proof is concluded.

S8 Additional simulation results

To provide empirical evidence of Theorem 2, we present in Figures S1–S2 QQ-plots of the standardized test statistic in (S39) over 1000 randomly sampled $\mathbb{Z}$ under the null. We consider two cases, $F = G = \text{MVN}(\mathbf{0}_d, \mathbf{I}_{d \times d})$ or $\text{Unif}\{[-1, 1]^d\}$, and it is observed that under both cases, the ECDF of the standardized test statistic as in (S39) is close to $N(0, 1)$ even for a small sample size $n_1 = n_0 = 20$ with $d = 20$.

References

- Abadie, A. and L’Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116(536):1817–1834.
- Bai, Z., Rao, C., and Zhao, L. (1988). Kernel estimators of density function of directional data. *Journal of Multivariate Analysis*, 27(1):24–39.
- Budninskiy, M., Yin, G., Feng, L., Tong, Y., and Desbrun, M. (2019). Parallel transport unfolding: A connection-based manifold learning approach. *SIAM Journal on Applied Algebra and Geometry*, 3(2):266–291.
- Chang, T. H., Watson, L. T., Lux, T. C. H., Butt, A. R., Cameron, K. W., and Hong, Y. (2020). Algorithm 1012: DELAUNAYSPARSE: Interpolation via a sparse subset of the Delaunay triangulation in medium to high dimensions. *ACM Transactions on Mathematical Software*, 46(4):38:1–38:20.
- Chen, H., Chen, X., and Su, Y. (2018). A weighted edge-count two-sample test for multivariate and object data. *Journal of the American Statistical Association*, 113(523):1146–1155.
- Facco, E., d’Errico, M., Rodriguez, A., and Laio, A. (2017). Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Scientific Reports*, 7(1):12140.
- Hall, P., Watson, G. S., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762.

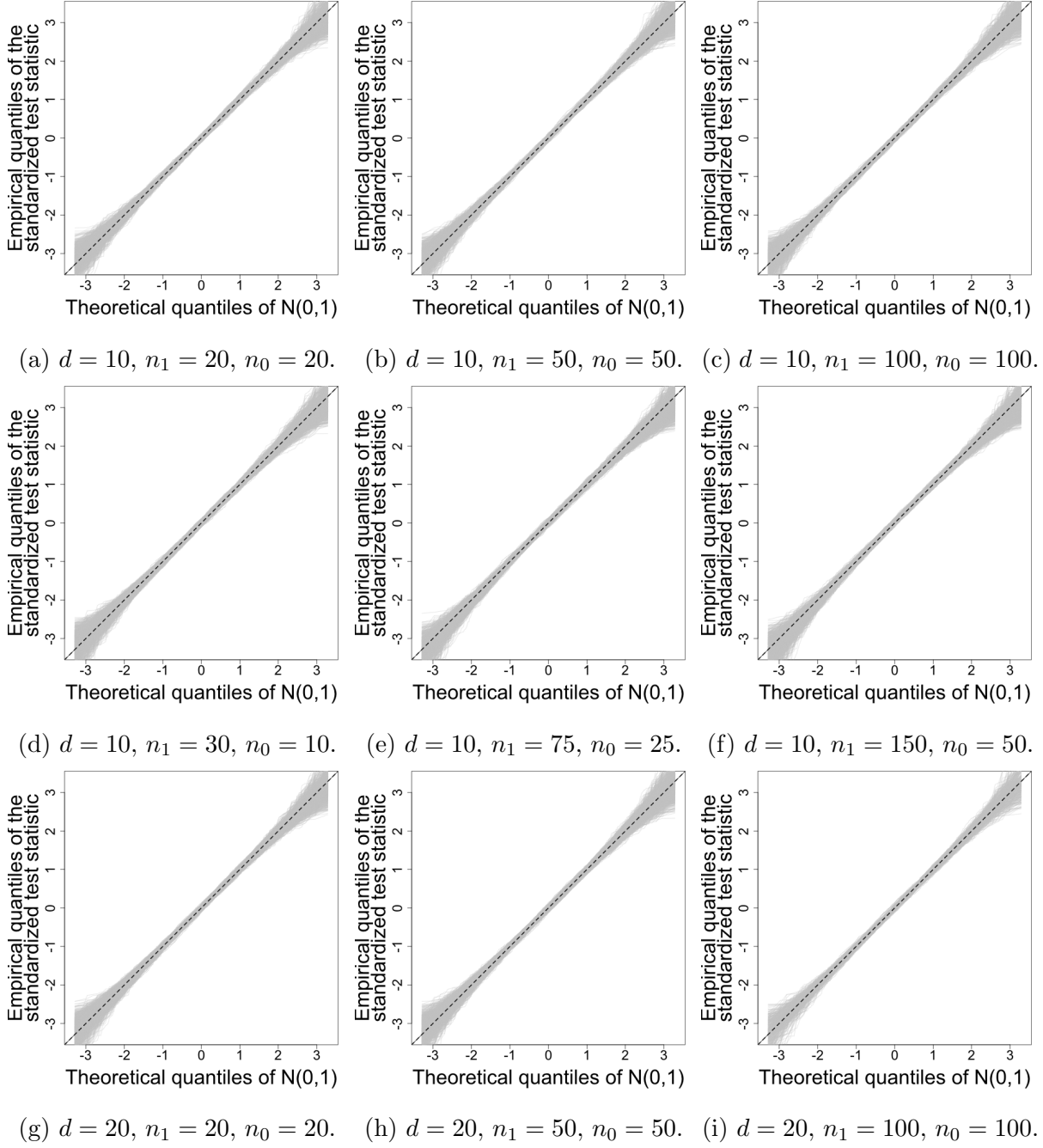


Figure S1: QQ-plots of the standardized test statistic in (S39) over 1000 randomly sampled $\mathbb{Z}$ under the null $H_0 : F = G = \text{MVN}(\mathbf{0}_d, \mathbf{I}_{d \times d})$ with different dimensions (d) and sample sizes (n_1 and n_0). Each gray solid line corresponds to the standardized test statistic under one randomly sampled $\mathbb{Z}$, while the dashed line is the reference line $y = x$.

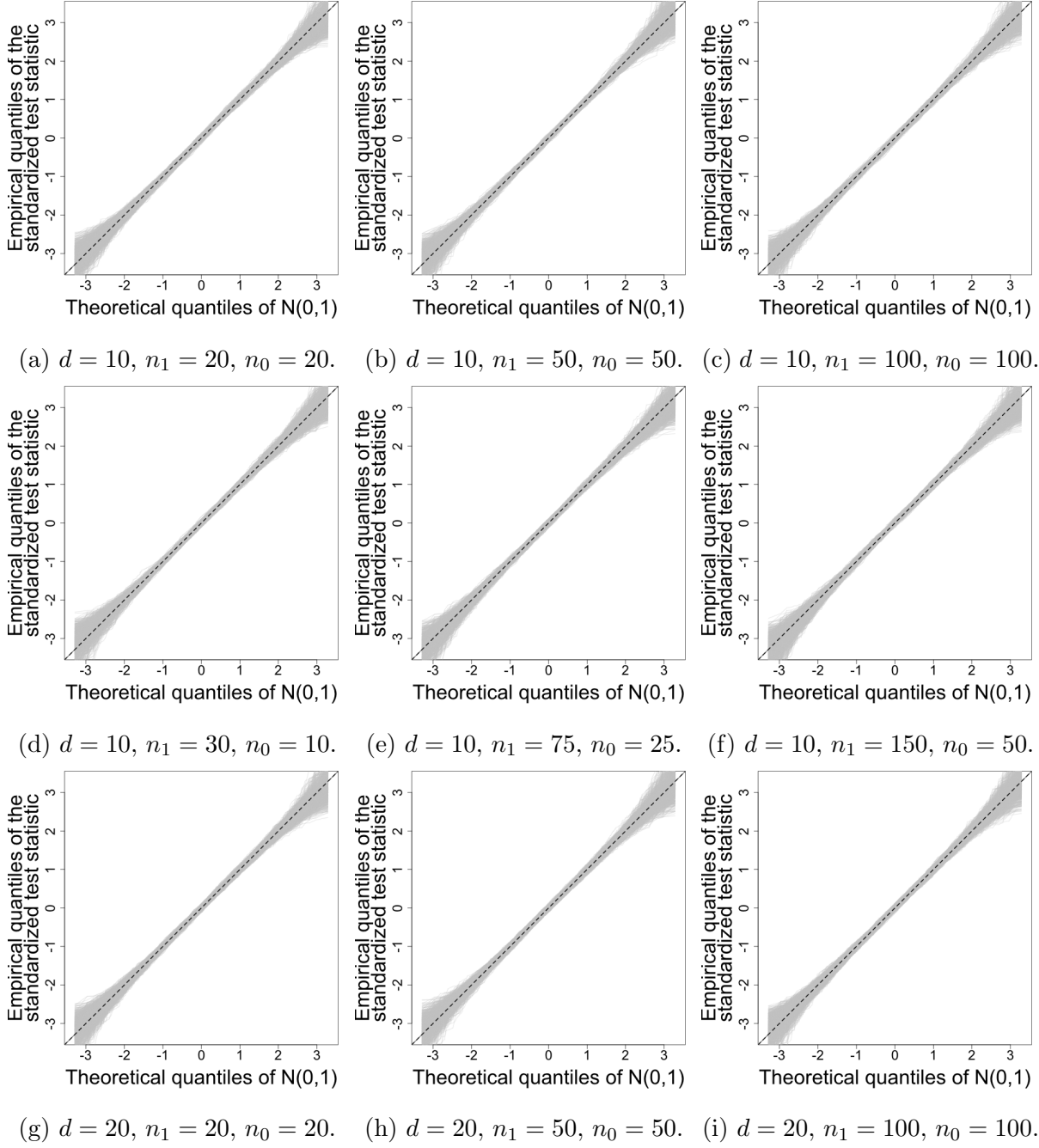


Figure S2: QQ-plots of the standardized test statistic in (S39) over 1000 randomly sampled $\mathbb{Z}$ under the null $H_0 : F = G = \text{Unif}\{[-1, 1]^d\}$ with different dimensions (d) and sample sizes (n_1 and n_0). Each gray solid line corresponds to the standardized test statistic under one randomly sampled $\mathbb{Z}$, while the dashed line is the reference line $y = x$.

- Henze, N. and Penrose, M. D. (1999). On the multivariate runs test. *The Annals of Statistics*, 27(1):290–298.
- Henze, N. and Voigt, B. (1992). Almost sure convergence of certain slowly changing symmetric one- and multi-sample statistics. *The Annals of Probability*, 20(2):1086–1098.
- Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *The Annals of Mathematical Statistics*, 23(2):169–192.
- Leibon, G. and Letscher, D. (2000). Delaunay triangulations and Voronoi diagrams for Riemannian manifolds. In *Proceedings of the sixteenth annual symposium on Computational geometry*, pages 341–349, New York, New York, USA. Association for Computing Machinery.
- Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323.