Algorithm 1: Connected Components of KS-CCDs

Algorithm 1: Given a dataset $\mathbf{X}$, this algorithm tests if there are multiple components or not. If it turns out that more than one components are detected, further investigation is need to decide whether these components are clusters or outliers;
Here, δ_0 is an initial value (big enough) for the density parameter δ in KS statistics, Δ is a small value that δ changes each time. N represents the number of datasets to simulate.

Input: δ_0, Δ, N and a dataset $\mathbf{X}$;
Output: Connected Components of $\mathbf{X}$;

Initialization:
1 $\mathbf{H_0}$: Assume $\mathbf{X}$ is generated from an underlying uni-modal distribution F with support S and there are no outliers.
2 $i \leftarrow 1$;
3 $M \leftarrow$ the size of $\mathbf{X}$;
4 $\delta_{seq} \leftarrow$ NULL ; `/* Store` δ `for each simulated dataset */`
5 **while** $i < N$ **do**
6     $\delta \leftarrow \delta_0$;
7   Simulate a dataset $\mathbf{X_0}$ of size M from the distribution F within S;
8     $D_{\mathbf{X_0}} = (\mathcal{V}_{\mathbf{X_0}}, \mathcal{A}_{\mathbf{X_0}}) \leftarrow$ the KS-CCD of $\mathbf{X_0}$ (with density δ);
9     $G_{\mathbf{X_0}} = (\mathcal{V}_{\mathbf{X_0}}, \mathcal{E}_{\mathbf{X_0}}) \leftarrow$ the intersection graph of D;
10   **while** $G_{\mathbf{X_0}}$ *is connected* **do**
11         $\delta \leftarrow \delta - \Delta$;
12   Repeat line 8 and 9 to update $D_{\mathbf{X_0}}$ and $G_{\mathbf{X_0}}$;
    end
13     $\delta_{seq} \leftarrow \delta_{seq} \cup \delta$;
end
14 The α quantile of δ_{seq} is denoted δ_α;
15 $D_{\mathbf{X}} = (\mathcal{V}_{\mathbf{X}}, \mathcal{A}_{\mathbf{X}}) \leftarrow$ the KS-CCD of $\mathbf{X}$ (with density δ_α);
16 $G_{\mathbf{X}} = (\mathcal{V}_{\mathbf{X}}, \mathcal{E}_{\mathbf{X}}) \leftarrow$ the intersection graph of $D_{\mathbf{X}}$;

17 **if** $G_{\mathbf{X}}$ *is connected* **then**
18   Do not reject $\mathbf{H_0}$ and return $\mathbf{X}$ as a single component;
else
19   Reject H_0 at $1 - \alpha$ level and return the connected components of $G_{\mathbf{X_0}}$;
end

Note:

1) In line 8, $\mathcal{V}_{\mathbf{X_0}} := \mathbf{X_0}$.
2) In line 9, for any $u, v \in \mathbf{X_0}$, the edge $uv \in \mathcal{E}_{\mathbf{X_0}}$ if and only if B_u and B_v cover v and u respectively.