Main stages of the MST-DBSCAN algorithm

Stage 1: Identifying transmission clusters

An MST-DBSCAN transmission cluster (Kuo et al., 2018) is formed by a set of case points, including a core point and a minimum number of spatio-temporal neighbors (at least 3 in our study).
The definition of the spatio-temporal neighbors, given below, models the transmission relationship between cases by ensuring that the neighbors j of a case i appear spatially near but later so that j could be infected by i.

$$STNB_{i,t} = \{j \in D_t \mid dist(i,j) \leq EpsS \bigcap EpsT2 \leq \triangle T(i,j) \leq EpsT1 \bigcap i \neq j\} \tag{1}$$

where
D_t denotes the set of case points included at time t,
$dist(i,j)$ denotes the spatial distance from i to j,
$EpsS$ is the threshold value of the spatial adjacency (1000m in our study),
$EpsT1$ and $EpsT2$ are the threshold values of the possible shortest and longest transmission (respectively 1 and 7 days in our study)
and $\triangle T(i,j)$ denotes the difference between the appearance time of i and j

Stage 2: Classifying the diffusion type of transmission clusters at each epoch

The algorithm identifies the diffusion type (e.g., emerge, grow, merge, steady, move, split, reduce) of each transmission cluster by calculating the movement distance of the mean center and the proportion of the area change between two epochs, according to the formulas below:

$$\triangle MC_{i,t} = dist(MC_{i,t}, MC_{i,t-1}) \tag{2}$$

$$\triangle A_{i,t} = \frac{A_{i,t} - A_{i,t-1}}{A_{i,t-1}} \tag{3}$$

where $MC_{i,t}$ and $A_{i,t}$ are, respectively, the spatial mean center and the area of the transmission cluster i at time t.

For example, the diffusion type "Move" corresponds to a transmission cluster for which the area remains the same, but the center moves.

Stage 3: Delineating diffusion zones

To compare the diffusion dynamics across our study area, we first project the evolution patterns of transmission clusters into administrative units (i.e., postal codes in our study). Then, a diffusion similarity coefficient S is computed for any two areas according to the formula below:

$$S_{i,j} = \frac{E_{i,j}}{E_a} \tag{4}$$

where E_a denotes the total number of epochs of the epidemic and $E_{i,j}$ denotes the number of epochs in which units i and j have the same evolution pattern.

Finally, the Louvain method (Blondel et al., 2008) groups administrative units by finding a set that has high similarities within each group and low similarities between groups.

For more details, please refer to the publication of Kuo et al. (2018)

References

- Kuo F-Y, Wen T-H, Sabel CE. 2018. Characterizing Diffusion Dynamics of Disease Clustering: A Modified Space–Time DBSCAN (MST-DBSCAN) Algorithm. Ann Am Assoc Geogr 108:1168–1186.
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008:P10008.