Supplements

Appendix A

Derivation of mathematical formulas

When a row-standardized spatial weights matrix is applied, G_i^* for a given area i can be simplified as

$$G_i^* = \frac{\sum_{j=1}^{N} w_{ij}x_j - \bar{x}\sum_{j=1}^{N} w_{ij}}{S\sqrt{\frac{N\sum_{j=1}^{N} w_{ij}^2 - \left(\sum_{j=1}^{N} w_{ij}\right)^2}{N-1}}} = \frac{\frac{1}{n_i}\left(\sum_{k=1}^{n_i} x_k - \sum_{k=1}^{n_i} \bar{x}\right)}{S\sqrt{\frac{N\sum_{k=1}^{n_i}(1/n_i)^2 - \left(\sum_{k=1}^{n_i} 1/n_i\right)^2}{N-1}}} = \frac{\frac{1}{n_i}\sum_{k=1}^{n_i}\frac{x_k-\bar{x}}{S}}{\sqrt{\frac{N/n_i-1}{N-1}}} = \frac{\frac{1}{n_i}\sum_{k=1}^{n_i} z_{x_k}}{\sqrt{\frac{N/n_i-1}{N-1}}}$$

$$= \sqrt{\frac{N-1}{N/n_i-1}}\,\tilde{z}_{x_i} \quad (1)$$

where x_j is the attribute value of area j, $\bar{x}$ is the mean of attribute values, w_{ij} is the spatial weight between area i and j, S is the standard deviation of attribute values, n_i is the number of neighbors (including i itself) of i, z_{x_k} is the z-score value of x_k and $\tilde{z}_{x_i}$ is the average z-score value of neighboring areas of i.

L statistic (Lee 2001) can be expanded as

$$L = \sqrt{SSS_X}\cdot\sqrt{SSS_Y}\cdot r_{\tilde{X},\tilde{Y}} = \sqrt{\frac{\sum_i(\tilde{x}_i-\bar{x})^2}{\sum_i(x_i-\bar{x})^2}}\cdot\sqrt{\frac{\sum_i(\tilde{y}_i-\bar{y})^2}{\sum_i(y_i-\bar{y})^2}}\cdot\frac{\sum_i(\tilde{x}_i-\bar{\tilde{x}})(\tilde{y}_i-\bar{\tilde{y}})}{\sqrt{\sum_i(\tilde{x}_i-\bar{\tilde{x}})^2}\sqrt{\sum_i(\tilde{y}_i-\bar{\tilde{y}})^2}}$$

$$= \frac{N}{\sum_i\left(\sum_j w_{ij}\right)^2}\frac{\sum_i\left[\left(\sum_j w_{ij}x_j - \sum_j w_{ij}\bar{x}\right)\left(\sum_j w_{ij}y_j - \sum_j w_{ij}\bar{y}\right)\right]}{\sqrt{\sum_i(x_i-\bar{x})^2}\sqrt{\sum_i(y_i-\bar{y})^2}} = \frac{\sum_i\left[\left(\frac{\sum_j w_{ij}x_j - \sum_j w_{ij}\bar{x}}{S_X}\right)\left(\frac{\sum_j w_{ij}y_j - \sum_j w_{ij}\bar{y}}{S_Y}\right)\right]}{\sum_i\left(\sum_j w_{ij}\right)^2} \quad (2)$$

where SSS_X and SSS_Y are the spatial smoothing scalars (SSS) of variable X and Y, respectively. SSS_X and SSS_Y can be used to represent the univariate spatial association of each variable (SSS will be larger if a variable is spatially clustered). $r_{\tilde{X},\tilde{Y}}$ is the Pearson's correlation coefficient between the spatial lags of two variables. $\tilde{x}_i$ ($\sum_j w_{ij}x_j$) and $\tilde{y}_i$($\sum_j w_{ij}y_j$) are the spatial lags of variable X and Y of area I, respectively. S_X ($\sqrt{\frac{\sqrt{\sum_i(x_i-\bar{x})^2}}{N}}$) and S_Y ($\sqrt{\frac{\sqrt{\sum_i(y_i-\bar{y})^2}}{N}}$) are the standard deviations of variable

X and Y, respectively. $\bar{x}$ and $\bar{y}$ are the means of variable X and Y, respectively. $\bar{\bar{x}}$ and $\bar{\bar{y}}$ are the means of the spatial lags of variable X and Y, respectively.

Form Eq. (2), we can find that a local bivariate spatial association L_i^* can be regarded as the relative contribution each area makes to L. Therefore, a local bivariate spatial association measure L_i^* can be easily derived from the L statistic as:

$$L_i^* = \frac{\left(\frac{\sum_j w_{ij}x_j - \sum_j w_{ij}\bar{x}}{s_X}\right)\left(\frac{\sum_j w_{ij}y_j - \sum_j w_{ij}\bar{y}}{s_Y}\right)}{\left(\sum_j w_{ij}\right)^2} = \frac{\left(\sum_j w_{ij}x_j - \sum_j w_{ij}\bar{x}\right)\left(\sum_j w_{ij}y_j - \sum_j w_{ij}\bar{y}\right)}{\left(\sum_j w_{ij}\right)^2 s_X s_Y} \tag{3}$$

When a row-standardized spatial weights matrix is applied, L_i^* can be simplified as:

$$L_i^* = \frac{\left[\frac{1}{n_i}\sum_{k=1}^{n_i}(x_k - \bar{x})\right]\left[\frac{1}{n_i}\sum_{k=1}^{n_i}(y_k - \bar{y})\right]}{S_X S_Y} = \left[\frac{1}{n_i}\sum_{k=1}^{n_i}\left(\frac{x_k - \bar{x}}{S_X}\right)\right]\left[\frac{1}{n_i}\sum_{k=1}^{n_i}\left(\frac{y_k - \bar{y}}{S_Y}\right)\right]$$

$$= \left(\frac{1}{n_i}\sum_{k=1}^{n_i} z_{x_k}\right)\left(\frac{1}{n_i}\sum_{k=1}^{n_i} z_{y_k}\right) = \tilde{z}_{x_i}\tilde{z}_{y_i} \tag{4}$$

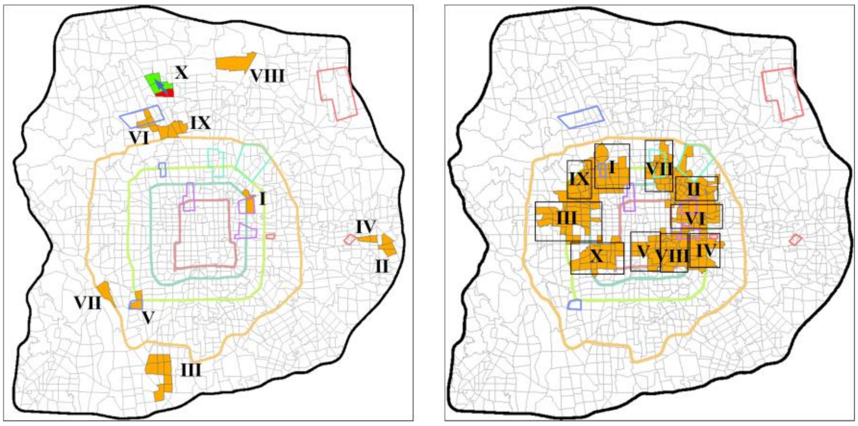
Where $\tilde{z}_{x_i}$ and $\tilde{z}_{y_i}$ are the local average z-scores for variable X and Y, respectively.

Appendix B

Bivariate flow clusters identified from Beijing smart card transaction and taxi trajectory data during afternoon rush hours

$H_I - H_{II}$ clusters. In Figure B1(a), $H_I - H_{II}$ clusters identified by BiFlowAMOEBA demonstrate that traffic demands revealed by taxi and public transit were extremely high in these regions. The clusters were discovered around commercial areas, residential areas, technology hubs and transportation hubs. Cluster I was located around Sanlitun. Cluster Ⅱ and IV were located around Tongzhou West Railway Station. Cluster V and VI were around Fengtai Science and Technology Park and Zhongguancun Software Park, respectively. Other clusters were located around residential areas. For example, Cluster VIII was located around Tiantongyuan, the largest housing community in Asia.

The $H_I - H_{II}$ clusters identified by BiFlowLISA are shown in Figure B1(b). AntScan_flow does not find $H_I - H_{II}$ clusters. The means of A^I and A^{II} for clusters identified by BiFlowAMOEBA are 8.78 and 1100.70. The means of A^I and A^{II} for clusters identified by BiFlowLISA are 2.37 and 248.74. The quality of the clustering results obtained by BiFlowAMOEBA is higher than that obtained by BiFlowLISA. We found that BiFlowLISA identify some clusters that were not identified by BiFlowAMOEBA. For these clusters, the maximum means of A^I and A^{II} are 2.95 and 348.82. For BiFlowAMOEBA, the means of A^I and A^{II} for the cluster with the lowest absolute value of BG_i^* are 3.5 and 1090. It indicates the $H_I - H_{II}$ clusters identified by BiFlowAMOEBA are more anomalous than that identified by BiFlowLISA.



(a) (b)

Figure B1. $H_I - H_{II}$ clusters identified during afternoon rush hours: (a) clusters identified by BiFlowAMOEBA; (b) clusters identified by BiFlowLISA.

$\boldsymbol{H_I - L_{II}}$ **clusters.** The clusters identified by BiFlowAMOEBA are shown in Figure B2(a), Cluster II, V, VI and IX mainly represent the traffic interactions between different regions and Beijing Capital International Airport. In these regions, taxi was a more popular travel tool than public transit for coming and leaving the airport during afternoon rush hours. Cluster I, Cluster III, Cluster IV, Cluster VII and Cluster VIII were the other five clusters discovered between the Fifth and Sixth Ring

Road. This may indicate that people residing in these areas prefer to choose taxi for commuting during afternoon rush hours.

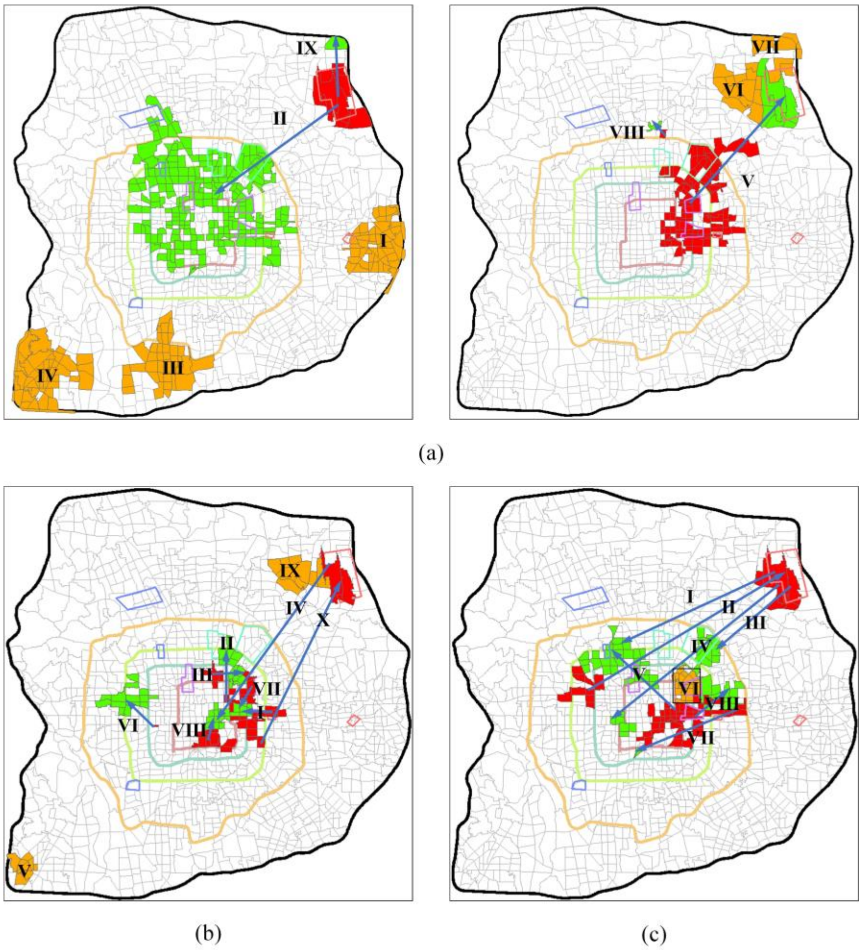


Figure B2. $H_I - L_{II}$ clusters identified during afternoon rush hours: (a) clusters identified by BiFlowAMOEBA; (b) clusters identified by BiFlowLISA; (c) clusters identified by AntScan_flow.

The $H_I - L_{II}$ clusters identified by BiFlowLISA and AntScan_flow are shown in Figure B2(b) and (c). The means of A^I and A^{II} for clusters identified by BiFlowAMOEBA, BiFlowLISA, AntScan_flow are 4.34 and 4.26, 1.53 and 6.22, and 1.00 and 10.46, respectively. The quality of the clustering results obtained by BiFlowAMOEBA is higher than that identified by BiFlowLISA and AntScan_flow. BiFlowLISA and AntScan_flow found some clusters which were not identified by BiFlowAMOEBA, e.g., Cluster I-III and Cluster VI-VIII in Figure B2 (b) and Cluster V-VIII in Figure B2 (c). For these clusters, the maximum mean of A^I is 2.23 and the minimum mean of A^{II} is 3.69. For BiFlowAMOEBA, the means of A^I and A^{II} for the cluster with the lowest absolute value of BG_i^* are 3.21 and 3.43. We can find that the $H_I - L_{II}$ clusters identified by BiFlowAMOEBA are more anomalous than that identified by BiFlowLISA and AntScan_flow.

$L_I - H_{II}$ clusters. In Figure B3(a), $L_I - H_{II}$ clusters identified by BiFlowAMOEBA mainly represent traffic interactions from commercial area and transportation hub to residential areas around Tongzhou West Railway Station. Cluster I, Ⅳ, Ⅷ and Ⅹ were from Sihui Coach Station to residential areas. Cluster II, Ⅵ and Ⅸ were from Beijing Central Business District to residential areas. Cluster Ⅶ was around Tongzhou West Railway Station. This may indicate that public transit was a more popular travel tool than taxi for people residing around the Tongzhou West Railway Station during afternoon rush hours. Cluster III and V represent the urban commuting between residential areas and technology hubs.

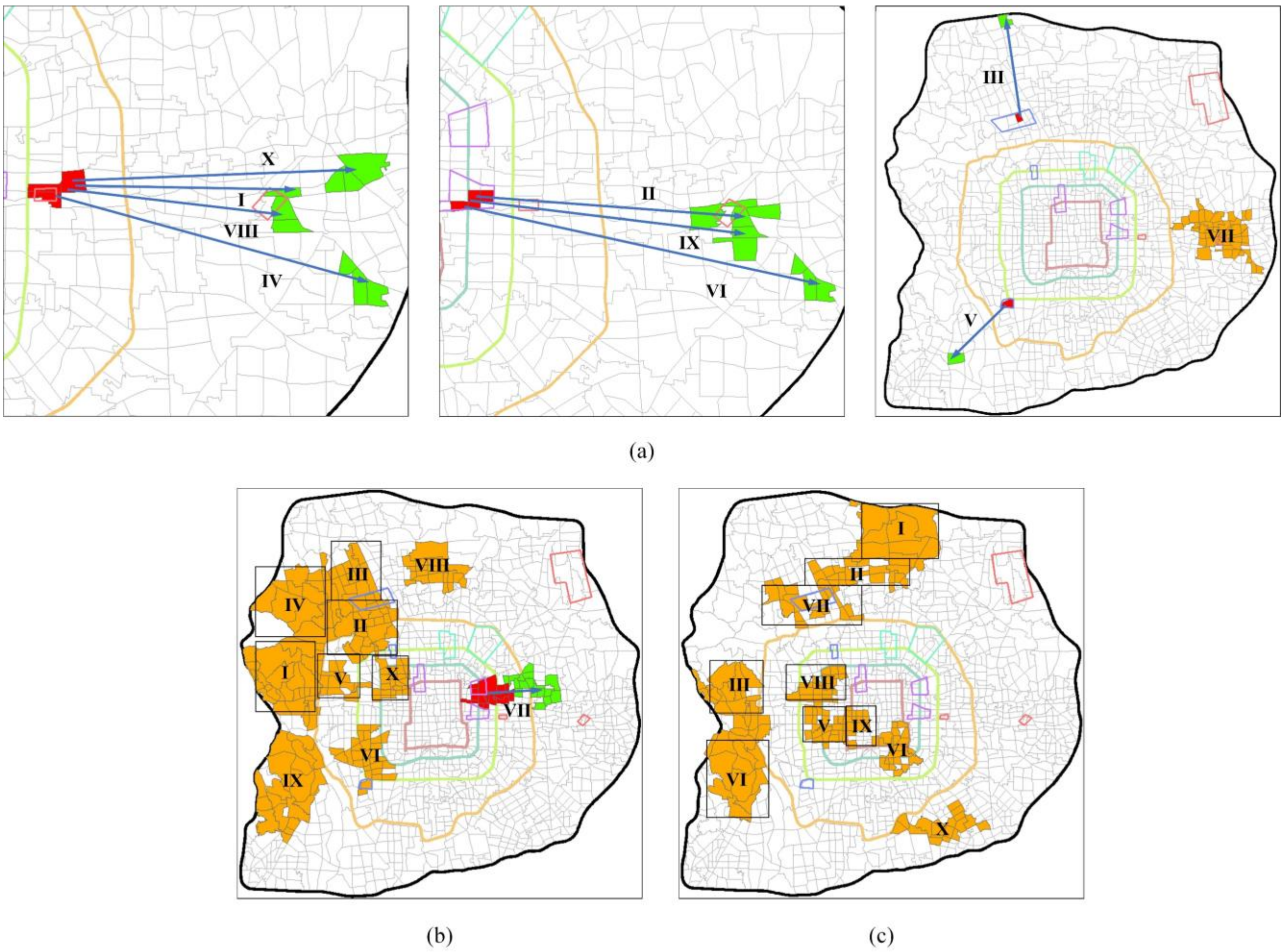


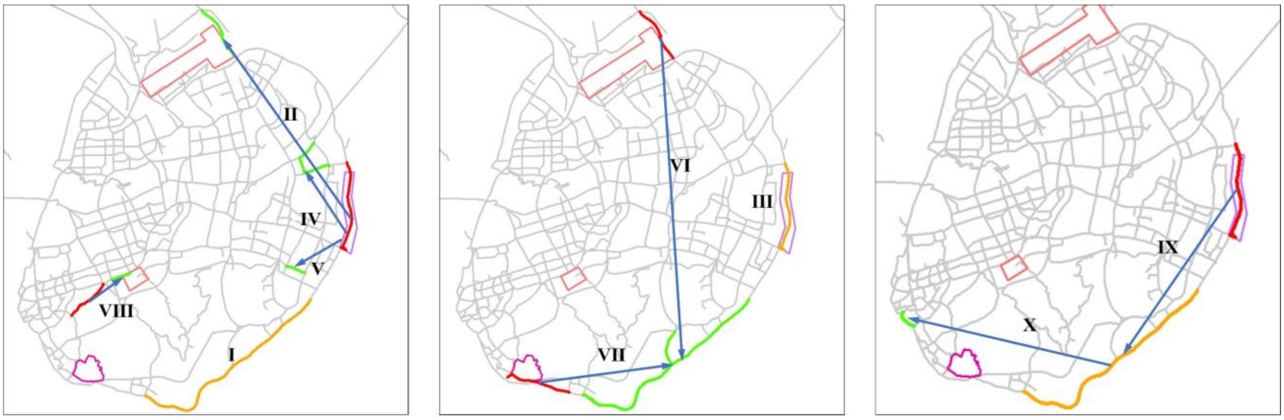
Figure B3. $L_I - H_{II}$ clusters identified during afternoon rush hours: (a) clusters identified by BiFlowAMOEBA; (b) clusters identified by BiFlowLISA; (c) clusters identified by AntScan_flow.

The $L_I - H_{II}$ clusters identified by BiFlowLISA and AntScan_flow are shown in Figure B3(b) and (c). The means of A^I and A^{II} for clusters identified by BiFlowAMOEBA, BiFlowLISA, and AntScan_flow are 0 and 922.95, 0 and 166.83, and 0.05 and 418.34, respectively. The quality of the clustering results obtained by BiFlowAMOEBA is higher than that identified by BiFlowLISA and AntScan_flow. For the clusters that were not identified by BiFlowAMOEBA, the minimum mean of A^I is 0 and the maximum mean of A^{II} is 449.35. For BiFlowAMOEBA, the means of A^I and A^{II} for the cluster with the lowest absolute value of BG_i^* are 0 and 669.5. Therefore, we can infer that the $L_I - H_{II}$ clusters identified by BiFlowAMOEBA are more anomalous than that identified by the other two methods.

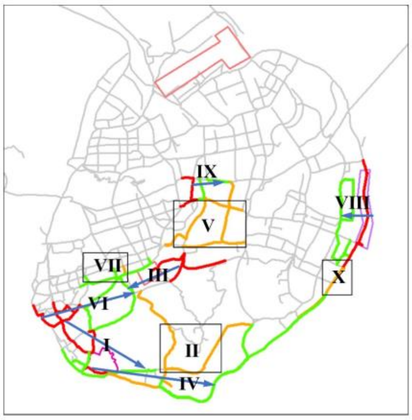
Bivariate flow clusters identified Xiamen taxi and ride-hailing data during afternoon rush hours

$\boldsymbol{H_I - H_{II}}$ clusters. In Figure B4(a), $H_I - H_{II}$ clusters identified by BiFlowAMOEBA were mainly discovered between commercial areas and transportation hubs or residential areas**.** These clusters indicate that taxi and ride-hailing are popular for traveling around or between some regions during afternoon rush hours. Cluster I was located around a residential area near Huandao South Road. Cluster Ⅱ was located from Guanyin Mountain to Xiamen Gaoqi International Airport, Cluster III was located around Guanyin Mountain. Cluster Ⅳ, Ⅴ and Ⅸ were located from Guanyin Mountain to residential areas**.** Cluster Ⅵ was found from Xiamen Gaoqi International Airport to a residential area near Huandao South Road and Cluster Ⅹ was located from the residential area to a commercial area (i.e., Lujiang)**.** Cluster Ⅷ was located from a commercial area to Xiamen Railway Station. Transport demands were also high between universities and residential areas during afternoon rush hours. For example, Cluster VII showed the flows from Xiamen University to the residential area near Huandao South Road.

The $H_I - H_{II}$ clusters identified by BiFlowLISA are shown in Figure B4(b). AntScan_flow does not find $H_I - H_{II}$ clusters. The means of A^I and A^{II} for clusters identified by BiFlowAMOEBA are 6.14 and 20.07. The means of A^I and A^{II} for clusters identified by BiFlowLISA are 3.11and 2.26. Therefore, the quality of the clustering results obtained by BiFlowAMOEBA is higher than that obtained by BiFlowLISA. In Figure B4(b), for Cluster I, Cluster III, Cluster V-VII, and Cluster IX that were not discovered by BiFlowAMOEBA, the maximum means of A^I and A^{II} are 3.33 and 3.21. For BiFlowAMOEBA, the means of A^I and A^{II} for the cluster with the lowest absolute value of BG_i^* are 5 and 11. It indicates that the $H_I - H_{II}$ clusters identified by BiFlowAMOEBA are more anomalous than that identified by BiFlowLISA.



(a)



(b)

Figure B4. $H_I - H_{II}$ clusters identified during afternoon rush hours: (a) clusters identified by BiFlowAMOEBA; (b) clusters identified by BiFlowLISA.

$H_I - L_{II}$ clusters. In Figure B5(a), $H_I - L_{II}$ clusters identified by BiFlowAMOEBA mainly represent the traffic interactions between different transportation hubs or between commercial areas and transportation hubs. Cluster I-Ⅳ were located from Xiamen North Railway station to Xiamen Gaoqi International Airport or vice versa, Cluster Ⅴ and Ⅸ were located around Xiamen Gaoqi International Airport, Cluster Ⅵ was located from a commercial area near Huandao East Road to Xiamen North railway station. These clusters may indicate that taxi is more popular than ride-hailing for traveling between these areas during afternoon rush hours. Cluster Ⅶ, Ⅷ and Ⅹ were located from commercial areas to residential areas. This may indicate that people in these residential areas prefer to choose taxi for commuting during afternoon rush hours.

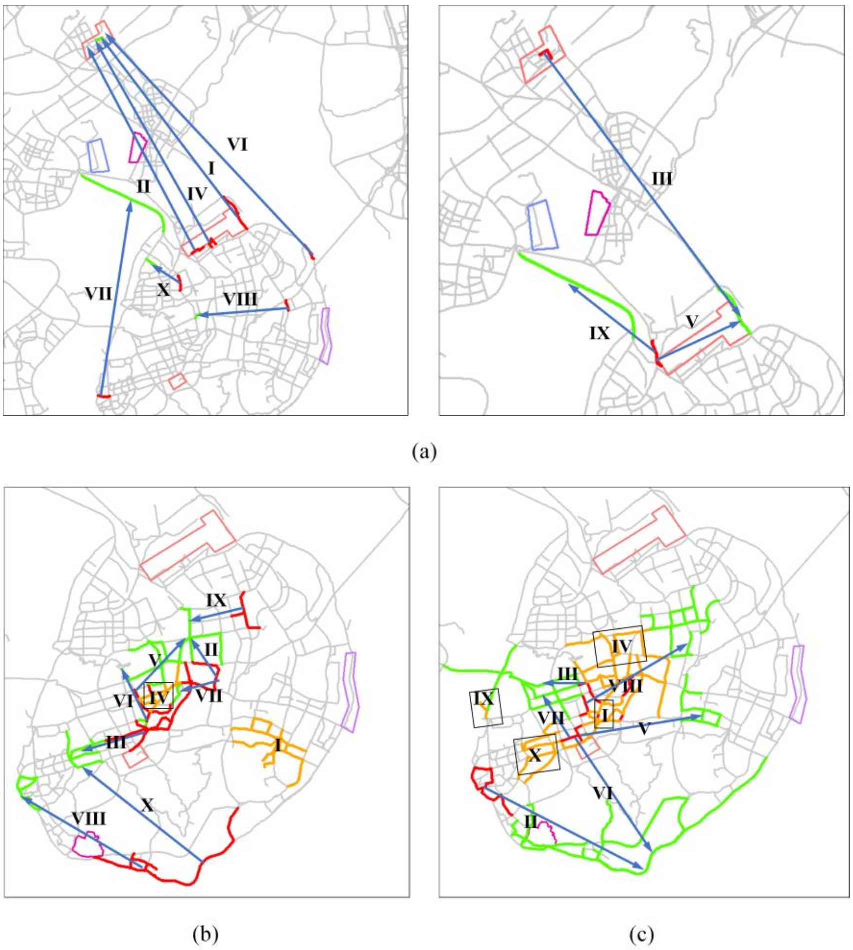


Figure B5. $H_I - L_{II}$ clusters identified during afternoon rush hours: (a) clusters identified by BiFlowAMOEBA; (b) clusters identified by BiFlowLISA; (c) clusters identified by AntScan_flow.

The $H_I - L_{II}$ clusters identified by BiFlowLISA and AntScan_flow are shown in Figure B5(b) and (c). The means of A^I and A^{II} for clusters identified by BiFlowAMOEBA, BiFlowLISA, and

AntScan_flow are 12.10 and 0, 1.80 and 0.18, and 1.38 and 0.34, respectively. The quality of the clustering results obtained by BiFlowAMOEBA is higher than that identified by BiFlowLISA and AntScan_flow.

$L_I - H_{II}$ clusters. In Figure B6(a), $L_I - H_{II}$ clusters identified by BiFlowAMOEBA mainly represent the urban commuting behavior between commercial areas and residential areas. Cluster III and Ⅵ were located from commercial areas to residential areas in Tongan District, Cluster Ⅴ-Ⅶ and Ⅸ-Ⅹ were located from commercial areas to residential areas around Jimei University, Cluster II was located from a commercial area to a residential area around Xiamen North Railway Station. These clusters may indicate that people residing in these residential areas prefer ride-hailing for commuting during afternoon rush hours. Cluster I and Ⅷ represent the traffic interactions between commercial areas and Xiamen North Railway Station.

The $L_I - H_{II}$ clusters identified by BiFlowLISA and AntScan_flow are shown in Figure B6(b) and (c). The means of A^I and A^{II} for clusters identified by BiFlowAMOEBA, BiFlowLISA, and AntScan_flow are 0 and 21.41, 0 and 2.41, and 0.05 and 2.88, respectively. We can infer that the clusters identified by BiFlowAMOEBA are more anomalous than that identified by BiFlowLISA and AntScan_flow.

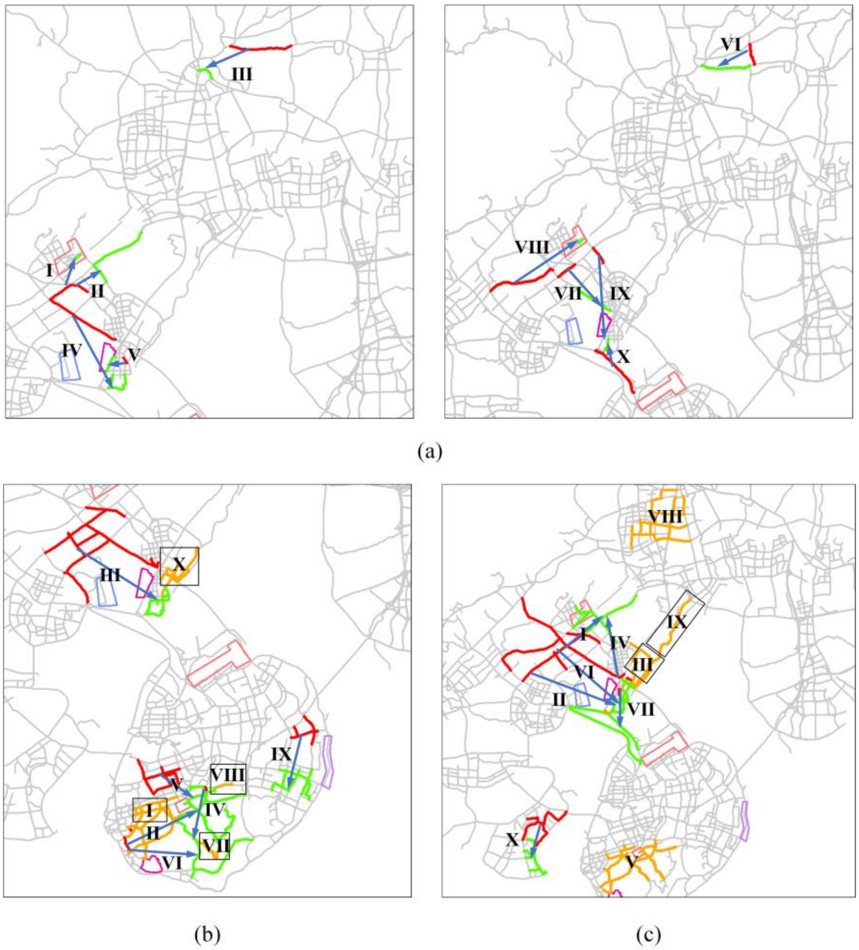


Figure B6. $L_I - H_{II}$ clusters identified during afternoon rush hours: (a) clusters identified by BiFlowAMOEBA; (b) clusters identified by BiFlowLISA; (c) clusters identified by AntScan_flow.

Reference

Lee, S.I., 2001. Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. Journal of geographical systems, 3 (4), 369–385.