Supplementary Information for "Quantifying long-term evolution of intra-urban spatial interactions"

Lijun Sun,[1,2] Jian Gang Jin,[3] Kay W. Axhausen,[1,4] Der-Horng Lee,[2] and Manuel Cebrian[5]

[1]*Future Cities Laboratory, Singapore-ETH Centre for Global Environmental Sustainability (SEC), 138602, Singapore*

[2]*Department of Civil & Environmental Engineering, National University of Singapore, 117576, Singapore*

[3]*School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China*

[4]*Institute for Transport Planning and Systems (IVT), Swiss Federal Institute of Technology, Zürich, 8093, Switzerland*

[5]*National Information and Communications Technology Australia, University of Melbourne, Victoria 3010, Australia*

SI TEXT

Data

Trip records were collected from Singapore's smart-card-based fare collection system, covering more than 96% of public transit trips. The system collects data for both bus and metro (rail based, including Mass Rapid Transit (MRT) and Light Rapid Transit (LRT)) modes.

Fields and their contents are provided in Tab. S1.

TABLE S1. Fields and contents of trip record dataset

Field	Description
Trip ID	A unique number for each transit trip
Card ID	A unique coded number for each smart card (anonymized)
Passenger Type	The attribute of cardholder (Adult, Senior citizen and Child)
Travel Model	Bus, MRT
Service Number	Bus route service number (e.g. 96); NULL for Metro (MRT&LRT)
Direction	Direction of the bus route (0 and 1); NULL for Metro (MRT&LRT)
Bus Registration No.	A unique registration number for each vehicle (e.g. '0999'); NULL for MRT&LRT
Boarding Stop/Station ID	A unique number for boarding bus stop / MRT&LRT station (e.g. 40009/Clementi)
Alighting Stop/Station ID	A unique number for alighting bus stop / MRT&LRT station (e.g. 40009/Clementi)
Ride Date	Date of a trip (e.g. '2011-04-11')
Ride Start Time	Start (tapping-in) time of a trip (e.g. 08:00:00)
Ride End Time	End (tapping-out) time of a trip (e.g. 08:00:00)
Ride Distance	Distance of the trip (e.g. 12.0 km)

Defining spatial zones

The average distance (spacing) between two successive bus stops in Singapore is about $300 \sim 400m$. We choose the resolution of spatial zones as $500m \times 500m$. One one hand, it enables us to identify the heterogeneity among different locations and distinguish important differences among adjacent zones. On the other hand, considering the average spacing of transit systems, the value also prevents us from creating too many zones, which might be null without generating/attracting

any transit flows.

Defining transit journeys

The smart card data set is trip based. We group sequential trips into one journey is their inter-trip interval (i.e. waiting time or transfer time) are less than 30 minutes.

Grouping transit journeys on time

In grouping transit journeys according to time, we set time rouping interval $\Delta t = 1\,hour$. For certain time t, the corresponding transit journey set is defined as:

$$J(t) = \{j \,|\, t_s(j) \leq t + \frac{\Delta t}{2} \wedge t_e(j) > t - \frac{\Delta t}{2}\}, \tag{S1}$$

where j represents each transit journey, $t_s(j)$ and $t_e(j)$ are the start time and end time of journey j, respectively.

Determining states of evolution

Given that all passenger transit journeys are time-stamped, for a certain time t we are able to extract all the journeys using the previous grouping strategy (by setting t and Δt) and use it to characterize the community structure. However, to better distinguish/characterize the evolution properties of each zone, ideally we would like to observe mutability values to span the whole support $[0,1]$ with high variation. Therefore, determining the length of time interval to define temporal states becomes a crucial problem. In principle, both sharp and minor transitions should be captured by the definition of subsequent states. If a short interval is used, in general we would observe a slower transition between subsequent states, so that we will get a small mutability value for all zones, preventing us from observing the consistent spatial distribution and influencing the calculation of $\Delta\phi$. Using a large interval (such as 4 hours) could help us to obtain sharp transitions; however, in this case the limited number of observations could result in that the mutability values are distributed around some unique values, instead of a continuous distribution. This could also limit our ability to find the effective region when using $\Delta\phi$ as a proxy.

Therefore, in this paper, we take the hourly intervals as the default temporal states and characterize the community structures for further analysis.

SI FIGURES

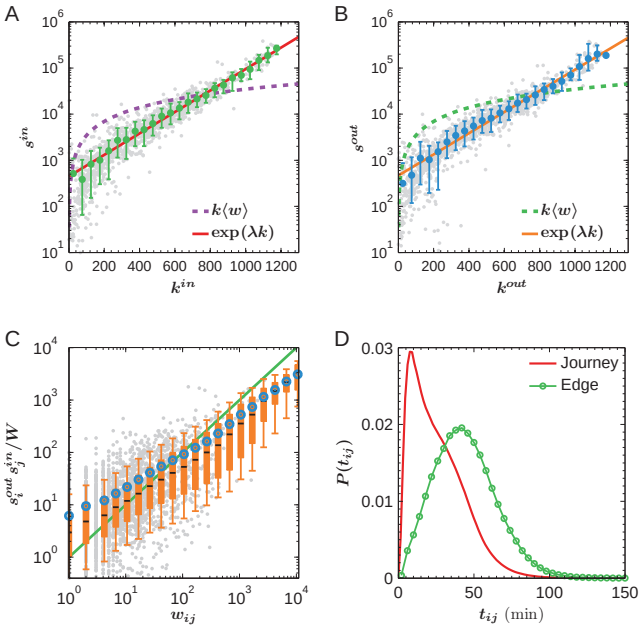

FIG. S1. Statistical properties of interaction intensity – w_{ij} and zone strength – s^{in} and s^{out}. (A) The dependence of s^{in} on k_{in}. Gray scatter plot shows zone in-strength s^{in} and corresponding in-degree k^{in}. The green circles correspond to mean value $\langle s^{in}\rangle$ of zones with their k^{in} falling in corresponding bin and the error bars indicate standard deviation of s^{in}. The dashed curve shows a function of $s^{in} = k^{in}\langle w\rangle$, implying the situation where all trips are equally distributed. The solid line shows an exponential fit $s^{in} \sim \exp\left(\lambda k^{in}\right)$ with $\lambda = 6.11 \pm 0.15$. The curves represent least square fit for $y = \log_{10}(s)$ and $x = k$. The coefficient of determination are 0.82 and 0.74, respectively. Clearly, the exponential increasing provides a better fit. (B) Same plot as in panel (A) but for out-strength s^{out}, which is well captured by $s^{out} \sim \exp\left(\lambda k^{out}\right)$ with $\lambda = 6.14 \pm 0.18$. (C) Compare the observed interaction intensity w_{ij} with a null model with $w'_{ij} = s_i^{out}s_j^{in}/W$ (shown as the green line). Box plot shows 9^{th} and 91^{st} percentiles in corresponding bins. The blue circles correspond to mean estimated intensities $\left\langle w'_{ij}\right\rangle$. As a guide, the black dashed line plots $s_i^{out}s_j^{in}/W \sim w_{ij}^{0.7}$. (D) Probability density function $P(t_{ij})$ of median travel time t_{ij} on link (i,j) (green circled curve) and $P(t)$ of individual travel time across all journeys (red curve).

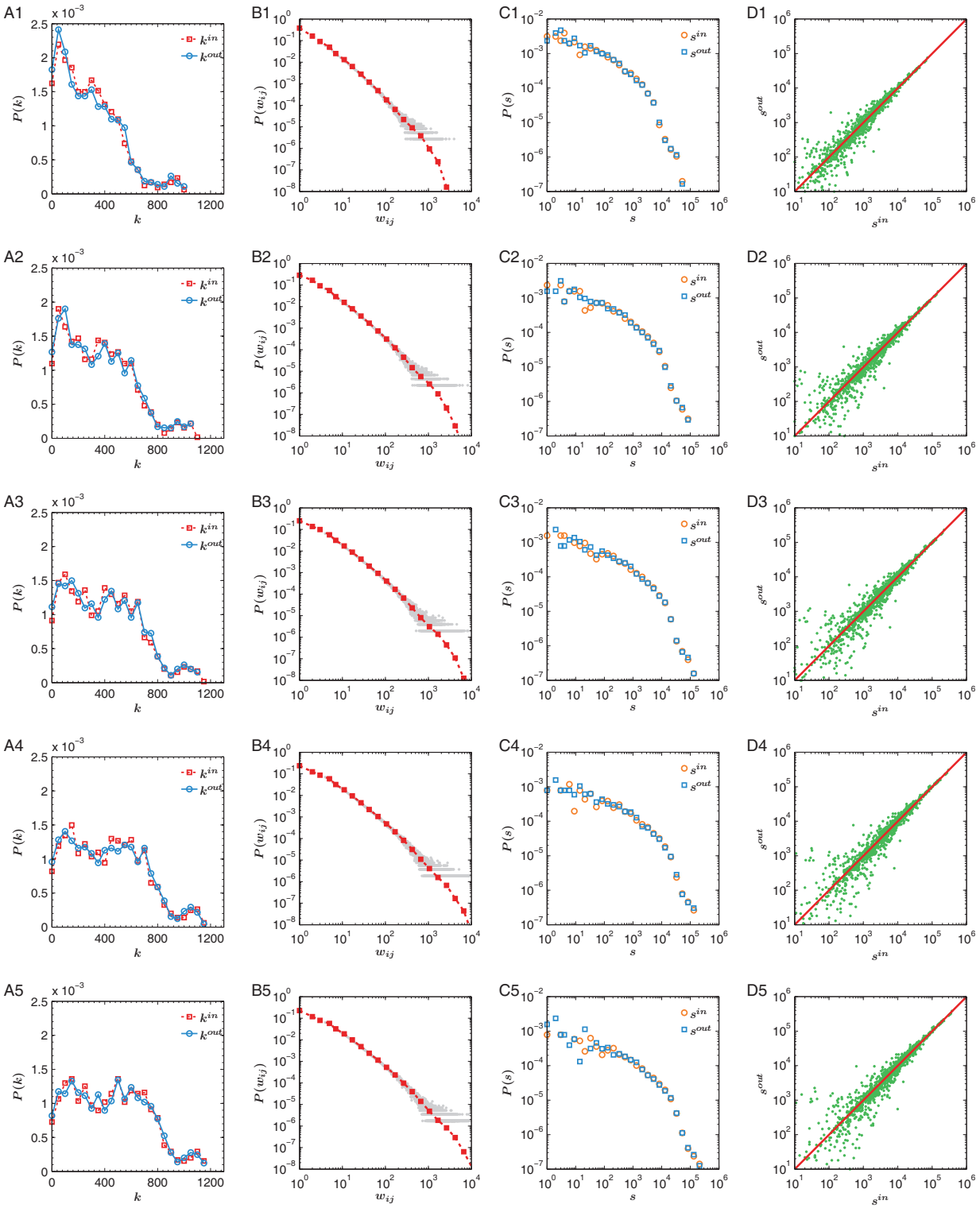

FIG. S2. Aggregated statistical properties (2012) for (A) in/out-degree, (B) interaction, (C) in/out-strength and (D) symmetrical plot of s^{in} and s^{out}. X1-X5 show the results aggregated over from 1 day to 5 days.

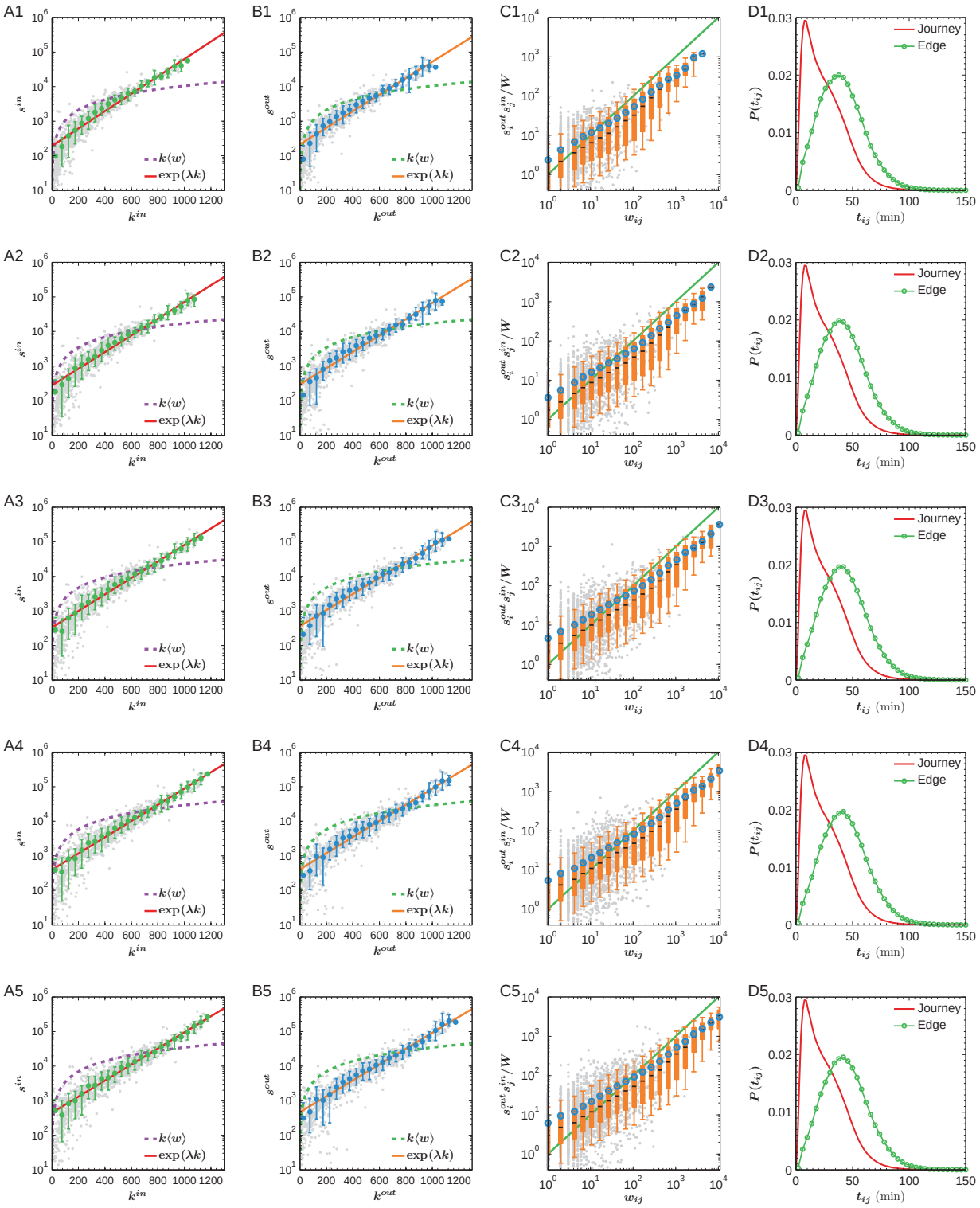

FIG. S3. Aggregated statistical properties (2012) for (A) the dependence of s^{in} on k_{in}, (B) the dependence of s^{out} on k_{out}, (C) observed interaction intensity w_{ij} with a null model with $w'_{ij} = s_i^{out}s_j^{in}/W$ and (D) $P(t_{ij})$ of median travel time t_{ij} on link (i,j) (green circled curve) and $P(t)$ of individual travel time across all journeys (red curve). X1-X5 show the results aggregated over from 1 day to 5 days.

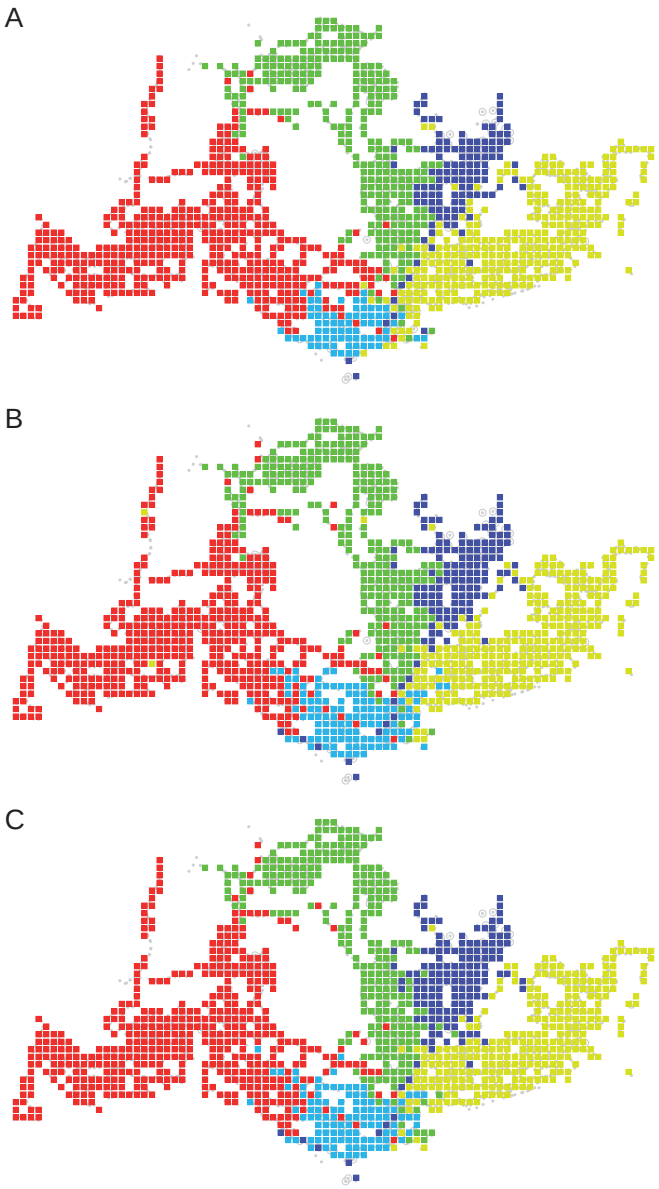

FIG. S4. The community structures of aggregated networks on weekdays in (A) 2011 ($Q_{2011} = 0.3142$), (B) 2012 ($Q_{2012} = 0.3163$) and (C) 2013 ($Q_{2013} = 0.3168$).

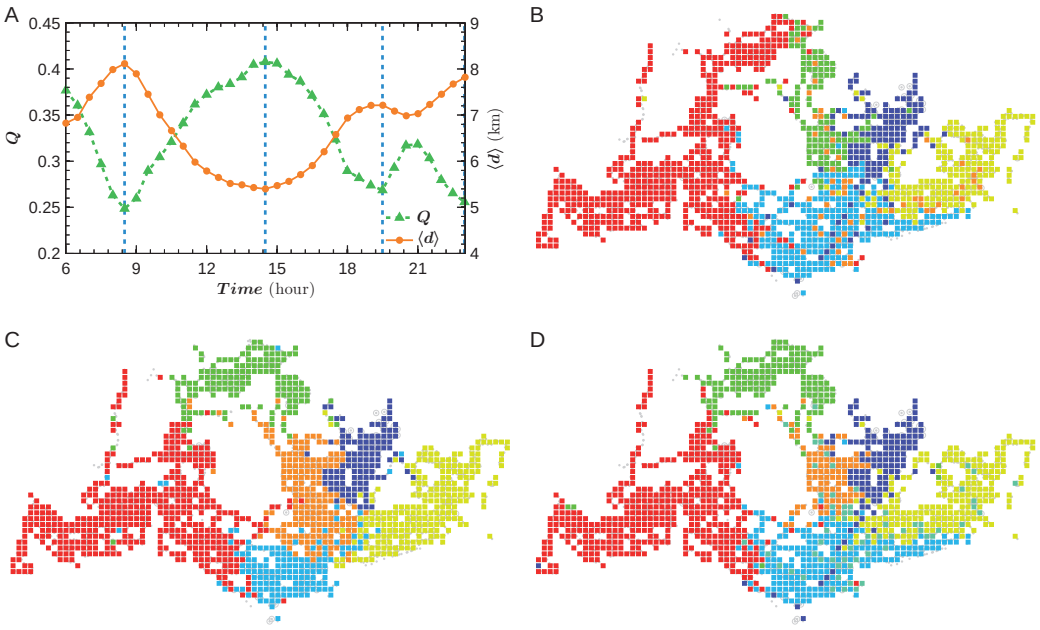

FIG. S5. (A) Comparison of Q_{12} and $\langle d \rangle_{12}$. Two local minimums of modularity are observed around 8:30am and 7:30pm, corresponding to daily transportation peaks. A maximum value is found around 2:30pm. (B) (D) Temporal community structure enabled by urban spatial interactions. Although spatial information is not used in identifying communities, the results still exhibit strong spatial relation/patterns. Panel (B) shows the community structure at 8:30am with $Q = 0.25$ and $\langle d \rangle = 8.11km$. Panel (C) displays community structure at 2:30pm, with a high modularity $Q = 0.41$ and $\langle d \rangle = 5.40km$. In between the morning and evening peaks, intra-urban movements are mainly composed by short local trips, resulting in a high modularity and well-ordered community structure spatially. In panel (D), we show the structure at 7:30pm with $Q = 0.27$ and $\langle d \rangle = 7.22km$.

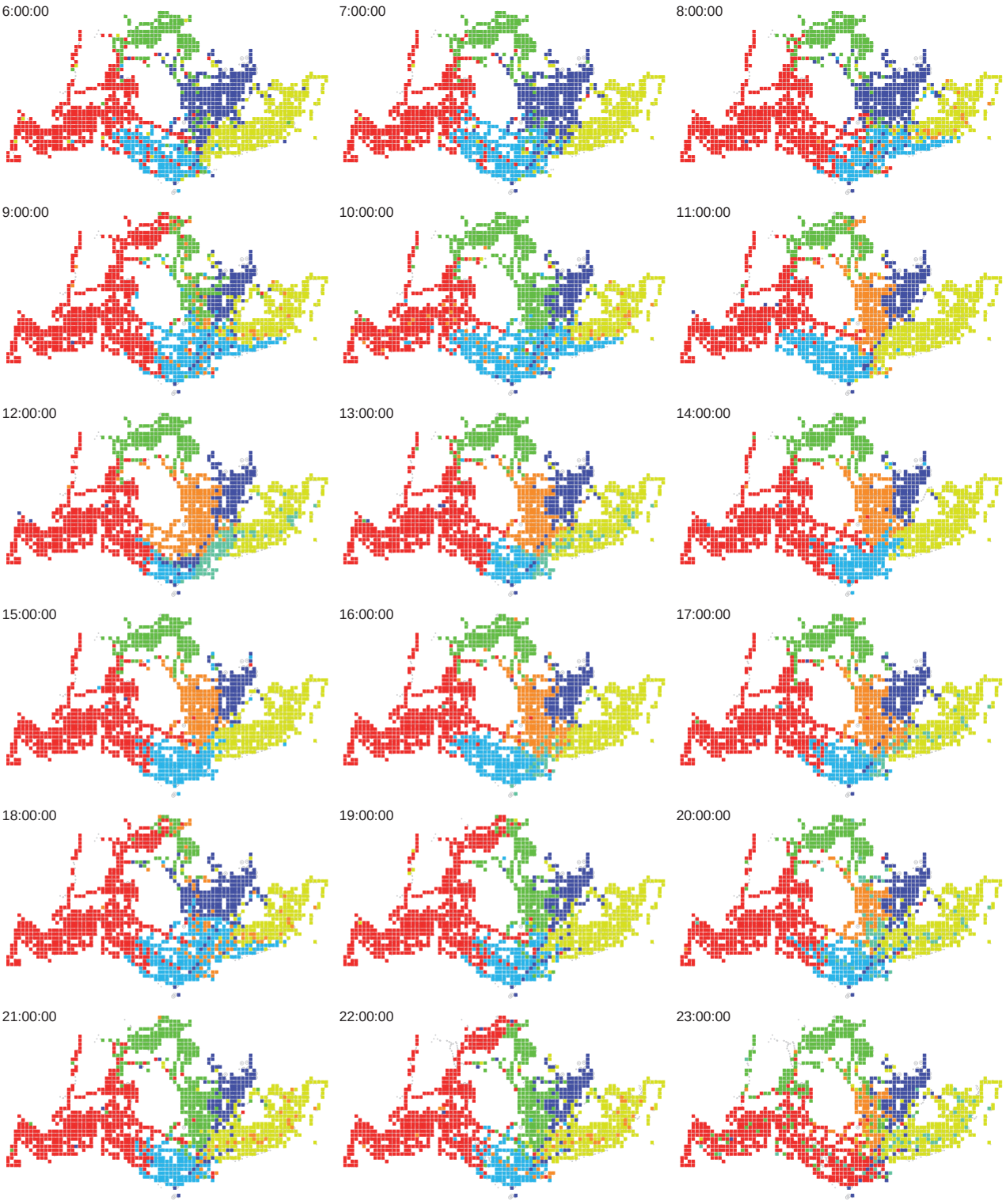

FIG. S6. Temporal community structure in year 2011

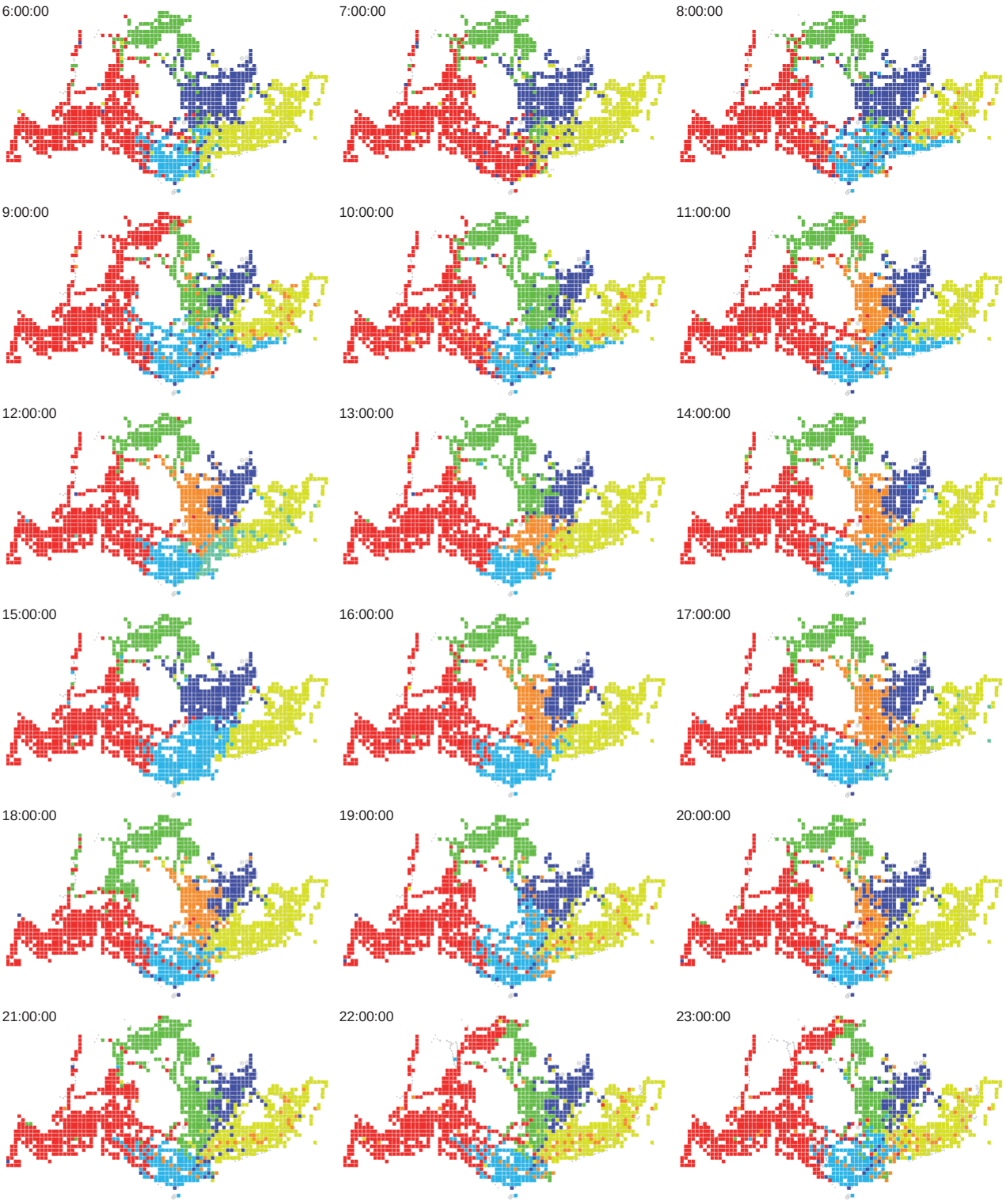

FIG. S7. Temporal community structure in year 2012

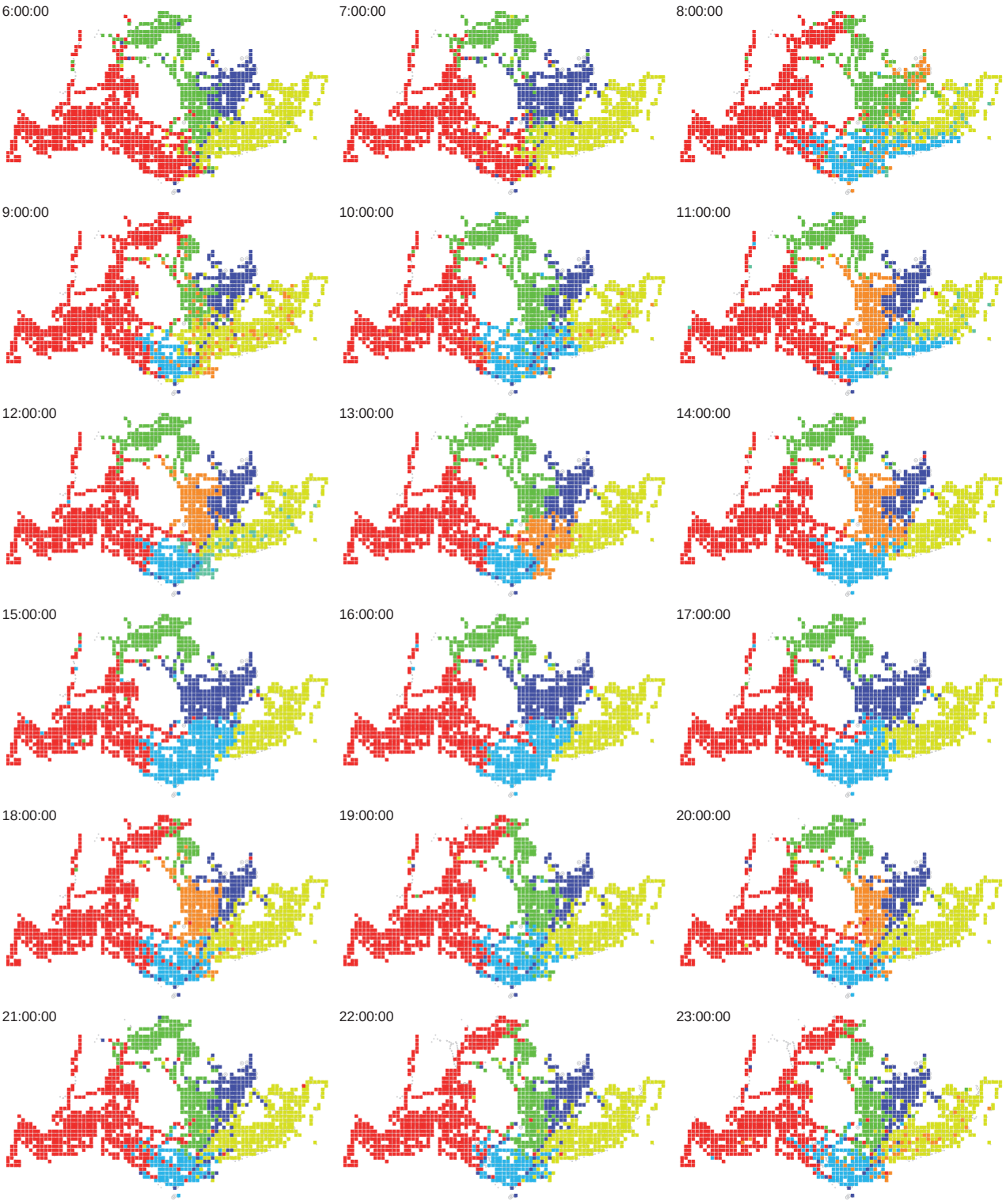

FIG. S8. Temporal community structure in year 2013