frontiers

Supplementary Material for "Diffusion and synchronization dynamics reveal the multi-scale patterns of spatial segregation"

Aleix Bassolas *, **Sergio Gómez and Alex Arenas**

Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Spain

*Correspondence: aleix.bassolas@urv.cat

1 SUPPLEMENTARY RESULTS FOR DIFFUSION AND SYNCHRONIZATION DYNAMICS AND ECONOMIC SEGREGATION

We provide here supplementary results related to the study of income segregation in US cities. Figure S1 reports (A) the mean and (B) standard deviation of x_i^k in Boston, Cleveland, Detroit and Denver. Both of them reach minimum values between 8-10.

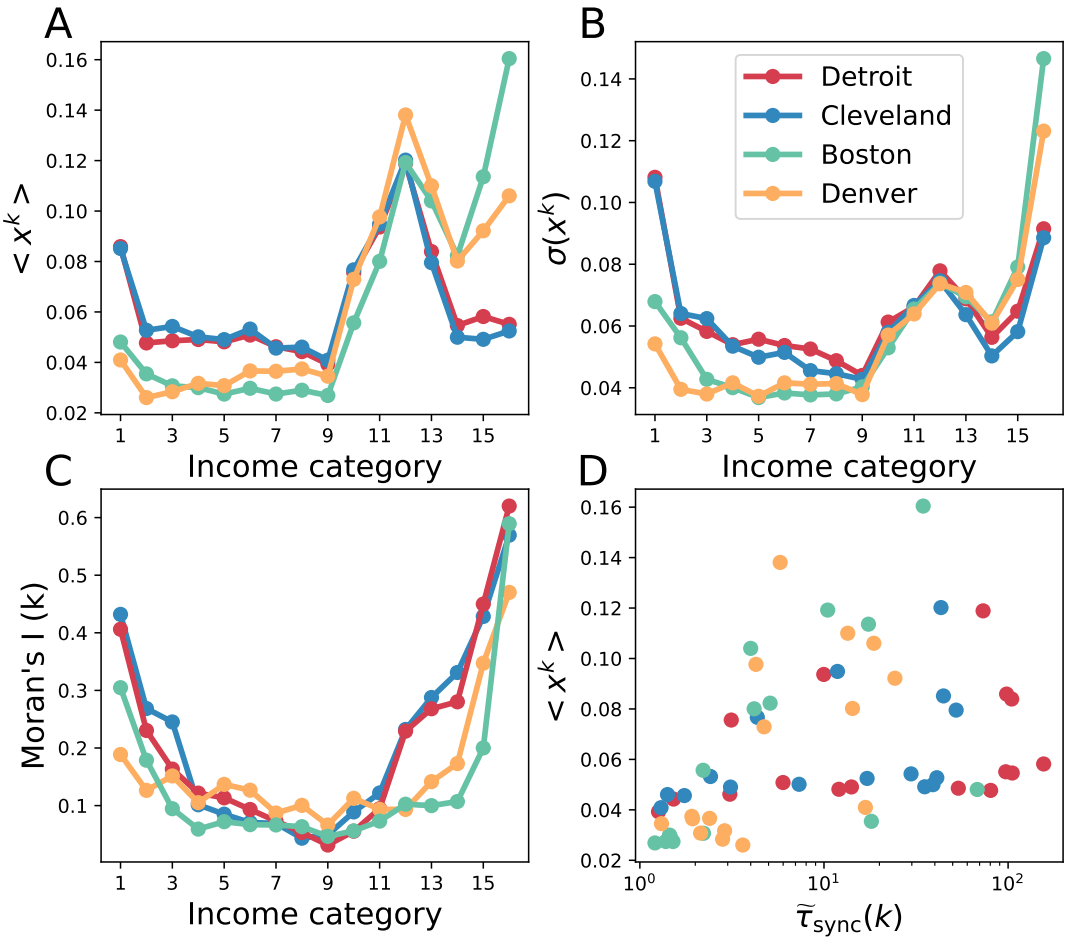

Figure S1. (A) Mean of x^k for each income category in Boston, Cleveland, Detroit and Denver. (B) Standard deviation $\sigma(x^k)$ for each income category in Boston, Cleveland, Detroit and Denver. (C) Moran's I for each income category in Boston, Cleveland, Detroit and Denver. (D) Scatter plot of the mean of x^k as a function of $\widetilde{\tau}_{\text{sync}}(k)$.

The fact that classes 8-10 appear to be the less segregated is also supported by the Moran's I as Fig. S1(C) shows. To further assess that the mean $< x^k >$ does not strongly determine the values of $\widetilde{\tau}_{\rm sync}(k)$, we plot both quantities in Fig. S1(D), where no strong pattern is observed. Categories with low $< x^k >$ display high variability in $\widetilde{\tau}_{\rm sync}(k)$ and vice-versa.

In Fig. S2 we provide the ranking of the selected US cities according to the value of $\widetilde{\tau}_{\rm sync}(k)$ for the lowest and highest income categories 1 and 16, respectively. As can be seen, there are significant variations in the ranking depending on which economic category is shown; for example, Cleveland is close to the top for category 1 but far apart for 16, and the other way around for Seattle.

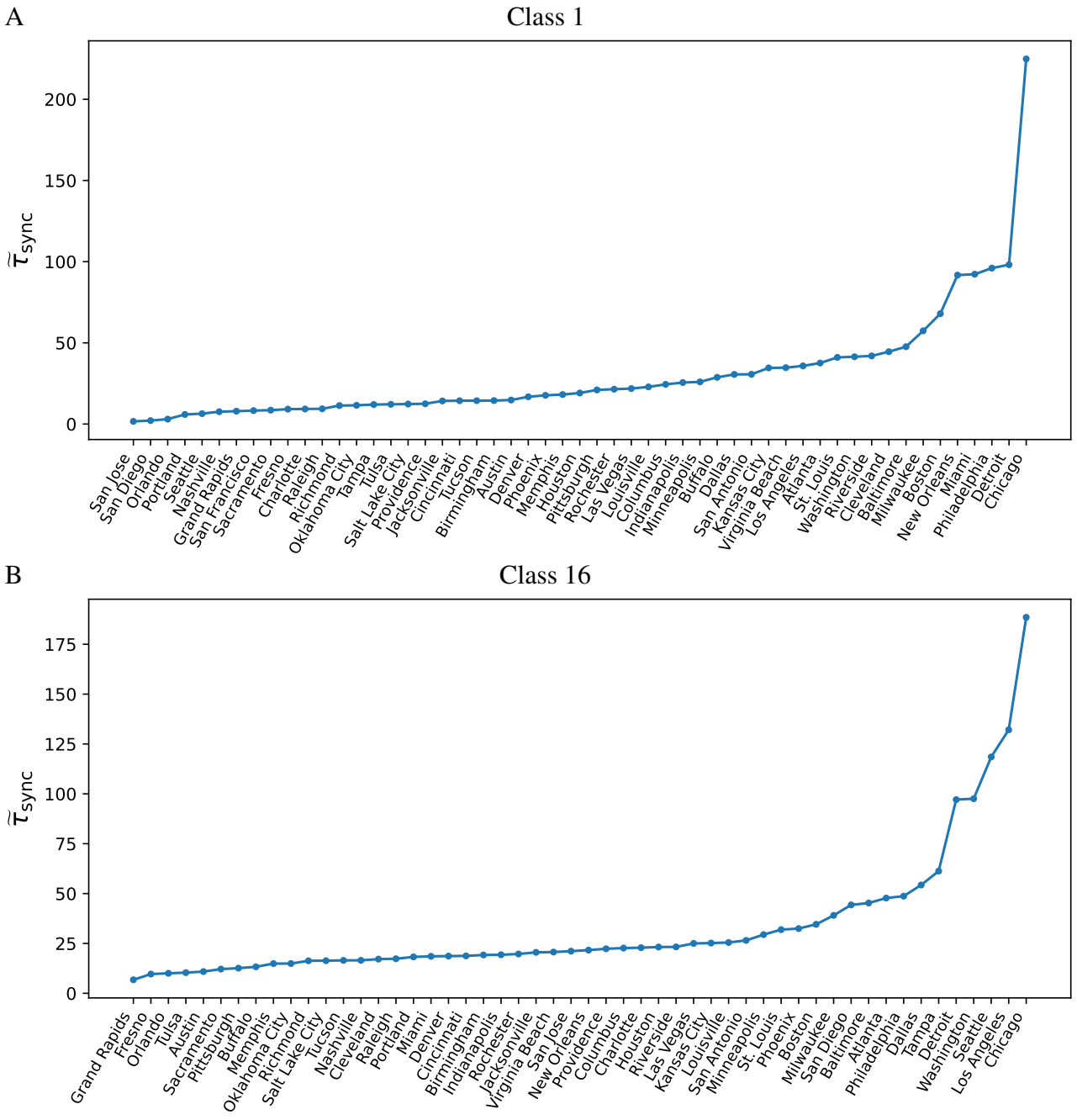

Figure S2. Ranking of the selected US cities according to the value of $\widetilde{\tau}_{\rm sync}(k)$, for income class 1 (A) and income class 16 (B).

2 COMPARISON WITH OTHER SEGREGATION MEASURES

In this section we assess how the normalized synchronization time $\widetilde{\tau}^k_{\rm sync}$ relates to other segregation measures. In particular we focus on the widely used Moran's I [1], which focuses on local correlations, and one obtained from class mean first passage times (CMFPT) developed in [2, 3], which captures long-range spatial correlations.

For each city and category k the Moran's I can be written as

$$I^k = \frac{\frac{1}{W}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i^k - \bar{x}^k)(x_j^k - \bar{x}^k)}{\frac{1}{n}\sum_{i=1}^{n}(x_i^k - \bar{x}^k)^2}, \tag{S1}$$

where x_i^k is the fraction of population in i that belongs to category k, $\bar{x}^k$ is its mean across all spatial units, the weights w_{ij} correspond in our case to the spatial adjacency matrix a_{ij}, and $W = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$ is the total weight.

As an index to assess the long-range correlations in the spatial distribution of the income categories, we will use the class mean first passage times between classes. In this methodology [2, 3], random walkers start from each of the spatial units in a system and move through the spatial adjacency graph until they have visited the 16 classes at least once. For this, each location is assigned to a class with probability proportional to its corresponding fraction of population. By averaging the number of steps that a walker needs to reach class j across all the units that belong to category i and for multiple realizations, we can obtain the class mean first passage times τ_{ij}, which encapsulate the average number of steps needed to reach a unit of category j when a walker departs from a unit of category i. After normalizing by a null-model in which colors are uniformly reshuffled at random to compensate for uneven class abundances, we finally obtain the normalized class mean first passage times $\widetilde{\tau}_{ij}$. The quantity $\widetilde{\tau}_{ij}$ provides thus information on how much time you need to reach category j when a walker departs from a unit of category i as compared to the null-model, values below 1 mean that two categories are closer than in the null-model and vice-versa for values above 1. To summarize the segregation of category k in a city we will use the CMFPT index, i.e., the $\text{med}(\widetilde{\tau})_k$ given by the median value of τ_{jk} $\forall j$.

For each city included in our analysis, we measure the Pearson correlation coefficient r_p between each of the additional segregation quantities and $\widetilde{\tau}^k_{\rm sync}$ for all the 16 categories k. More specifically, for each city r_p is calculated over a set of 16 points. The distribution of r_p across cities is shown in Fig. S3 for the Moran's I (A) and $\text{med}(\widetilde{\tau})_k$ (B), where a skewness towards high values is clearly observed. Most of the cities display correlations above 0.8 with the Moran's I and 0.7 with the CMFPT index. Additionally, we also show in Fig. S4 the significance of the correlations observed in each of the cities, which are also below 0.001 in most of the cases.

In the main text we discuss the potential of our methodology to assess the multiscale patterns of segregation in front of traditional first-neighbor approaches. In Fig. S5 we further investigate this fact by plotting for Boston, Cleveland, Denver and Detroit the local normalized synchronization times, the local Moran's $I_i^{\mathrm{loc}}(k)$, and the raw ratio of population of category k in each of the census tracts.

Although the segregation hotspots detected by our methodology and the local Moran's I seem similar, the patterns detected are significantly different. Whereas $I_i^{\mathrm{loc}}(k)$ captures strong differences between

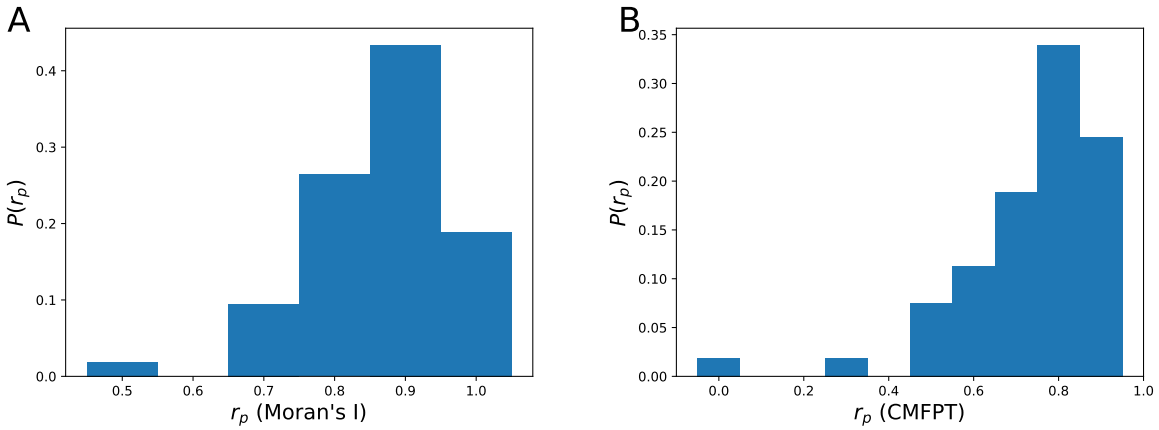

Figure S3. Correlation between $\widetilde{\tau}^k_{\mathrm{sync}}$ and the additional segregation indicators. For each city in our study, we calculate the Pearson correlation coefficient r_p between $\widetilde{\tau}^k_{\mathrm{sync}}$ and the additional segregation metrics over the 16 income categories. The correlation coefficient for a city is thus obtained from a set of 16 points, one per category. (A) Distribution of r_p between Moran's I and $\widetilde{\tau}^k_{\mathrm{sync}}$ across cities. (B) Distribution of r_p between the segregation calculated through normalized CMFPT $\mathrm{med}(\widetilde{\tau})_k$ and $\widetilde{\tau}^k_{\mathrm{sync}}$ across cities.

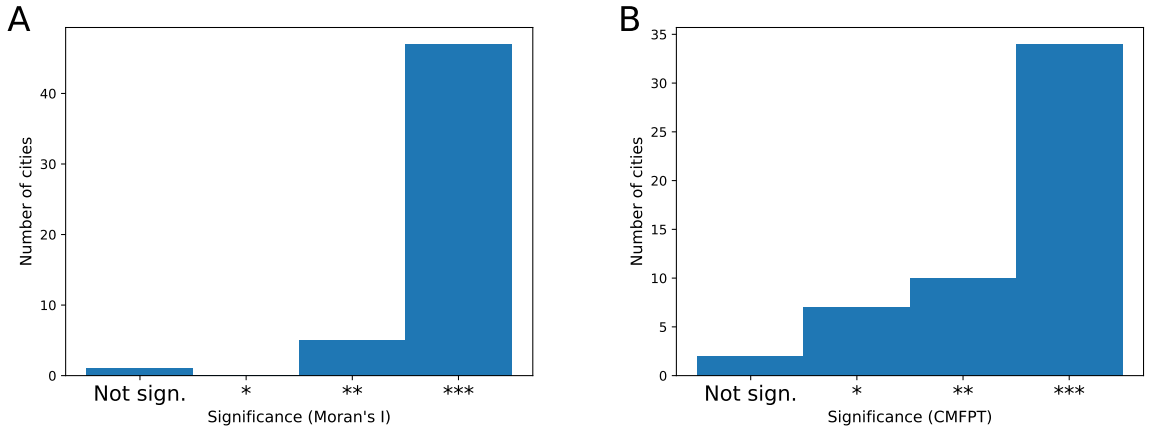

Figure S4. Significance of the Pearson correlation coefficients between $\widetilde{\tau}^k_{\mathrm{sync}}$ and the other segregation indicators. For each of the additional indices, we display the significance of the correlations across cities. (A) Significance of correlations between Moran's I and $\widetilde{\tau}^k_{\mathrm{sync}}$. (B) Significance of correlations between the segregation calculated through normalized class mean first passage times $\mathrm{med}(\widetilde{\tau})_i$ and $\widetilde{\tau}^k_{\mathrm{sync}}$. The correlation coefficient and significance for each city is obtained by comparing the segregation values for the 16 income categories. The significance values are depicted as * for p-value < 0.05, ** for p-value < 0.01, and *** for p-value < 0.001.

neighboors, $\widetilde{\tau}_i^{\mathrm{loc}}(k)$ highlights isolated regions even if the differences with their first-neighboors is low; most likely, this is because they are far apart from regions displaying ratios of population closer to the city average and require more time to reach the global synchronized state. In fact, the areas highlighted by synchronization dynamics have a larger scale and allow us to identify common mesoscale patterns of segregation across cities: a downtown that displays high values, a ring around it with low values, and finally the suburbs with high values again. By focusing on Detroit, we can see that not only the poorer downtown appears highlighted but also the suburbs due to their very low ratio of population of category 1. Similar patterns can also be observed in Cleveland and Denver.

The segregation index developed in the main text is calculated as the median of $\widetilde{\tau}_{\mathrm{sync}}^k$ which confers an equal weight to each of the income categories, disregarding the amount of population in each category. However, we can also construct a weighted index $\widetilde{\widetilde{\tau}}_{\mathrm{sync}}$ that can be built as

$$\widetilde{\widetilde{\tau}}_{\mathrm{sync}} = \frac{\sum_k P_k \widetilde{\tau}_{\mathrm{sync}}(k)}{\sum_k P_k}, \tag{S2}$$

where P_k is the total number of citizens that belong to category k in a given city. The ranking of cities according to the value of $\widetilde{\widetilde{\tau}}_{\mathrm{sync}}$ (Fig. S6) displays only slight changes with, for example, Philadelphia and Los Angeles closer to the top of the ranking. We test the relation between both indices in Fig. S7, where a clear relationship between both quantities is revealed.

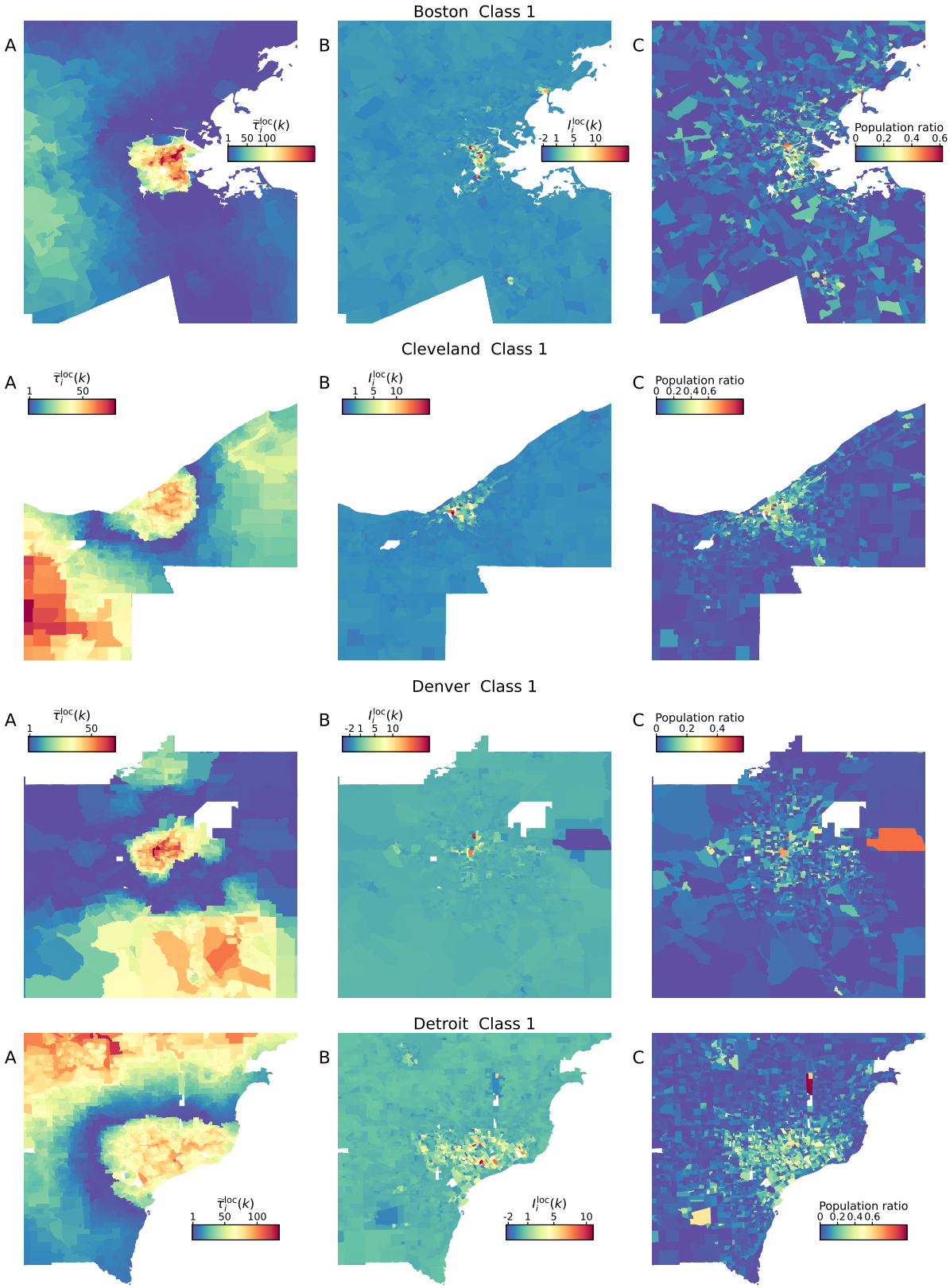

Figure S5. Comparison of local segregation indicators in Boston, Cleveland, Denver and Detroit. (A) Normalized synchronization time, (B) Local Moran correlation, and (C) proportion of citizens for each census tract and income class 1 (most deprived).

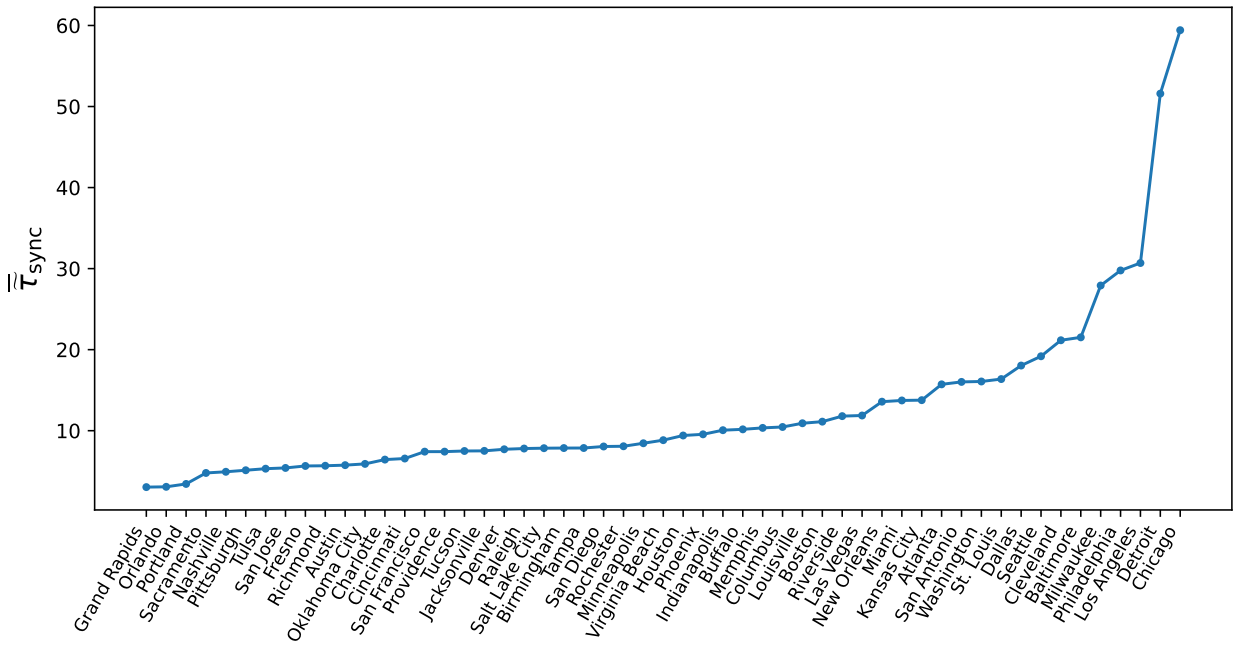

Figure S6. Segregation in US cities according to an index calculated through a weighted average. Ranking of cities according to the weighted index of segregation $\widetilde{\overline{\tau}}_{\text{sync}}$.

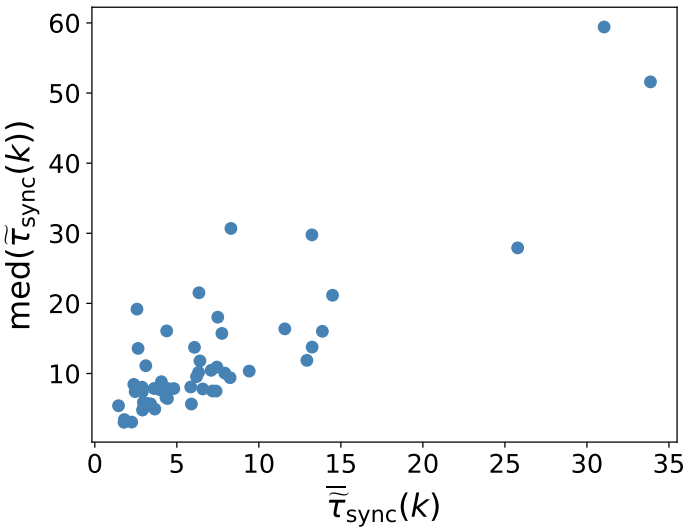

Figure S7. Comparison between segregation indicators obtained through synchronization dynamics. Comparison between the weighted index of segregation $\widetilde{\overline{\tau}}_{\text{sync}}$ and the index $\text{med}(\widetilde{\tau}_{\text{diff}}(k))$ used in the main text.

3 BEYOND ECONOMIC SEGREGATION: PARIS AROUND THE CLOCK

Besides only economic segregation, our methodology can be used to assess the spatial heterogeneity of any other quantity, and to exemplify it, we assess in this section the segregation of the population in Paris according to a wide set of socioeconomic indicators. The data compiles the fraction of population per district within a certain category at each hour of the day in French cities; in this work, we focus on Paris [4, 5, 6]. The list of indicators and categories analyzed can be found in Table S1.

Indicator	Categories				
Activity type	At home	At work	Studying	Shopping	Leisure
Age	16-24	25-34	35-64	65 and more	
Educational level	Low	Middle-low	Middle-high	High	
Socioprofessional status	Inactive	Low	Middle-low	Middle-high	High
Last travel mode	Pub. trans.	Private motor	Soft mobility		
Occupational status	Active	Student	Unemployed	Retired	Inactive
Sex	Male	Female			

Table S1. Socio-economic indicators and activity types analyzed for Paris.

For each indicator or category, we have a certain distribution of population per spatial unit and hour of the day, thus we can compute how the quantity $\widetilde{\tau}_{\rm sync}(k)$ varies during the day, as we show in Fig. S8(A,B) for the five activity types, and the five socio-professional status; the patterns of synchronization through time turn out to be very distinct. For example, the level of synchronization remains basically constant throughout the day for low, middle and high socio-professional status, while it increases (decreases) between 8am and 8pm for inactive (high) socio-professional status. If we focus instead on the ranking of $\widetilde{\tau}_{\rm sync}(k)$ at 10am and 10pm, see Fig. S8(C), the lower occupational and socio-professional status seem to be the most segregated indicators as they are on top of the ranking at both times of the day. Other categories that should be uniformly distributed across the city, such as sex, are very close to 1, thus indicating no segregation.

The hourly patterns of each metric allow for the grouping of indicators behaving similarly as we did for US cities. As before, we focus more on the time-series profile rather than the specific values taken by bon $\widetilde{\tau}^h_{\rm sync}(k)$, thus analyzing the normalized $P(\widetilde{\tau}^h_{\rm sync}(k))$ for each hour of the day h. The k-means clustering reveals four distinct clusters (see Fig. S9) which correspond to: those increasing during workings, those decreasing, those remaining almost constant, and those with a more characteristic behavior with a peak during midday and at the end of the day, roughly around the lunch and dinner times.

Finally, we assess the local segregation of districts by measuring their local normalized synchronization time. In particular, we show an example in Fig. S10 for the population performing leisure activities and those with inactive socio-professional status. In agreement with the temporal pattern shown in Fig. S8, the segregation is much higher at 10pm compared to 10am, especially concentrated in the centre of the city; a not so surprising result given that most of the leisure activities are concentrated in that part of the city. In the case of the population with inactive socio-professional status, the hotspots seem to be concentrated in the northern part of the city, a region known for suffering a thriving inequality.

REFERENCES

[1]Moran PA. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)* **10** (1948) 243–251.

[2]Bassolas A, Nicosia V. First-passage times to quantify and compare structural correlations and heterogeneity in complex systems. *Communications Physics* **4** (2021) 1–14.

[3]Bassolas A, Sousa S, Nicosia V. Diffusion segregation and the disproportionate incidence of covid-19 in african american communities. *Journal of the Royal Society Interface* **18** (2021) 20200961.

[4]Lecomte C, Vallée J, Le Roux G, Commenges H. Le mobiliscope, un outil de géovisualisation des rythmes quotidiens des métropoles. *Mappemonde. Revue trimestrielle sur l'image géographique et les formes du territoire* (2018).

[5][Dataset] Vallée J, Douet A, Villard E, Lecomte C, Le Roux G. Mobiliscope (2021). doi:10.5281/zenodo.4670766.

[6]Vallée J, Lenormand M. Intersectional approach of everyday geography. *arXiv* (2021) arXiv:2106.15492.

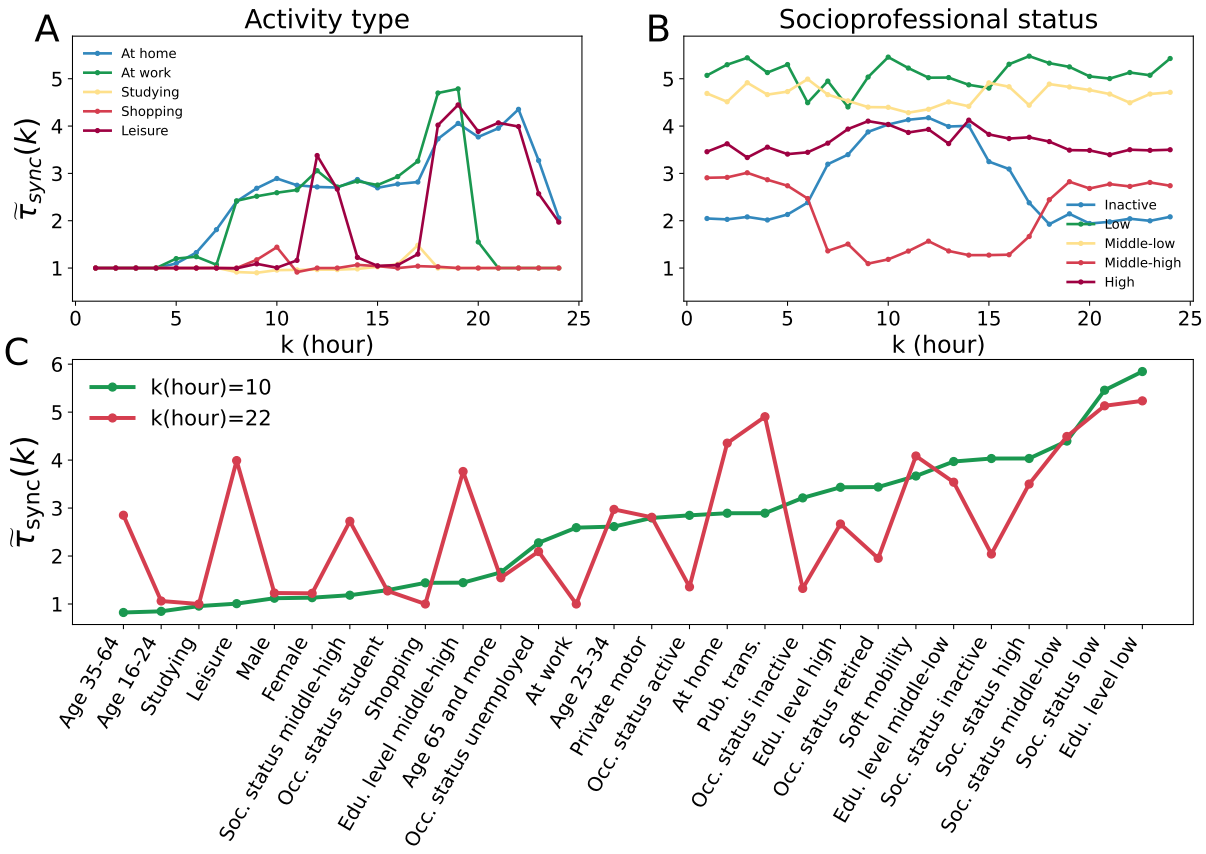

Figure S8. Synchronization around the clock in Paris. (A) Normalized synchronization time for the distribution of population performing each of the five types of activities. (B) Synchronization time for the distribution of population of each socio-professional status. (C) Change of synchronization times for all of the indicators at 10am (green) and 10pm (red).

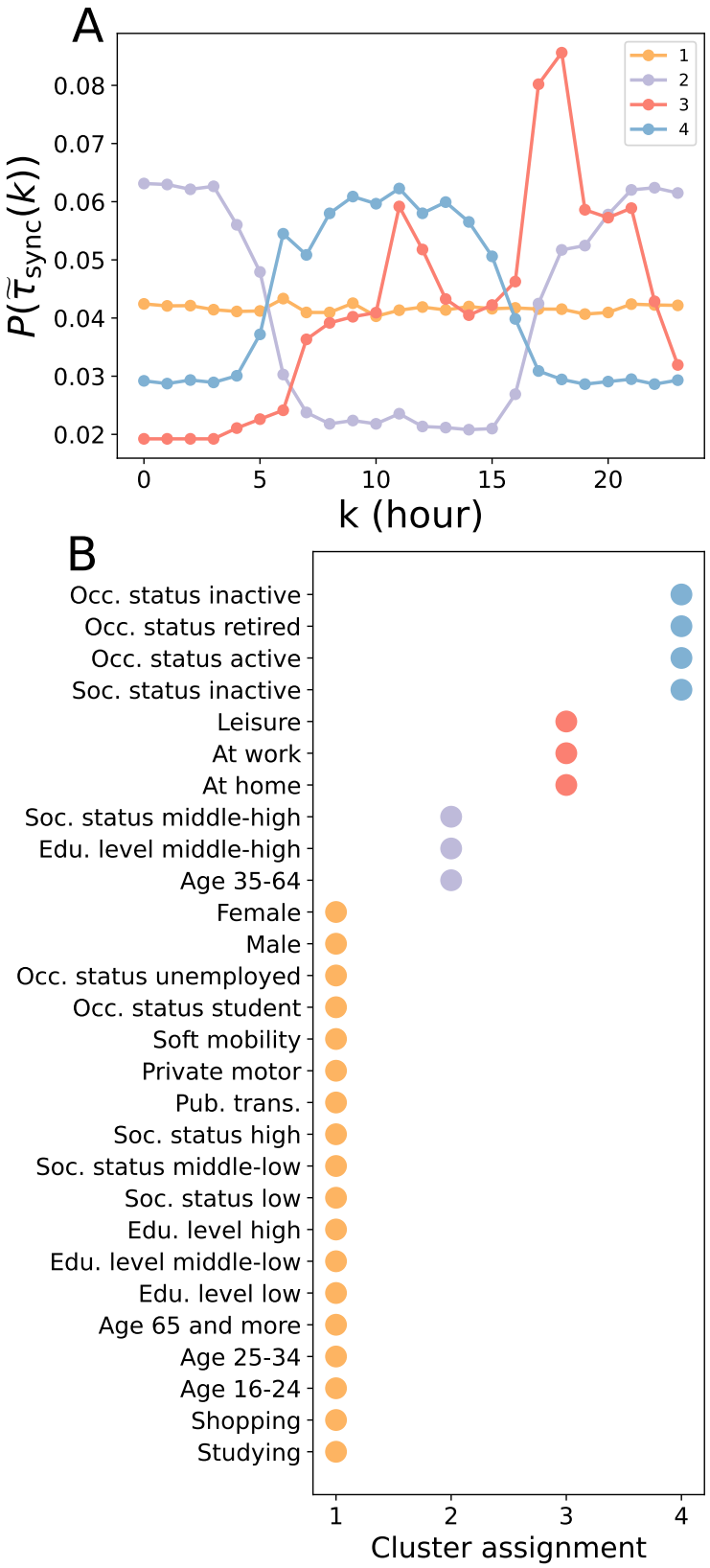

Figure S9. Clustering analysis of segregation around the clock in Paris. (A) Pattern of synchronization times P for each of the four main groups detected with the K-Means algorithm. (B) Cluster assignment for each of the indicators analyzed.

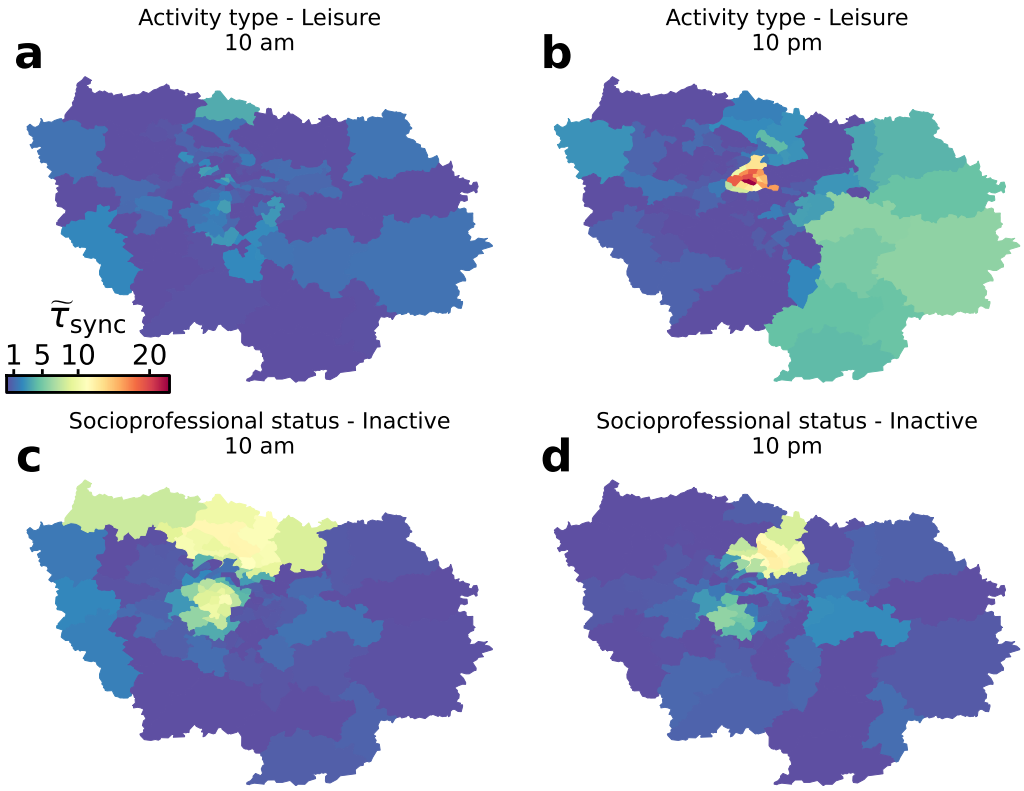

Figure S10. Local synchronization around the clock in Paris. (A, B) Normalized synchronization time for each Paris district for the population performing leisure activities at 10am and 10pm. (C, D) Normalized synchronization time for each Paris district for the population with inactive socio-professional status at 10am and 10pm. For visualization purposes the color range is common to all four maps.