Assessing the spatial scale of segregation in the Netherlands.

Lucas Spierenburg *Corresponding author*
Dr. Oded Cats
Dr. Sander van Cranenburgh

1 Introduction

Spatial segregation, here understood as the uneven distribution of social groups in space, is a persisting problem in many cities in the world (Veneri et al., 2021). It can occur along one or several social dimensions, such as income, religion, or migration background. This situation is prejudicial for society, as segregation can result in exacerbating inequality between groups; in terms of education achievements, well-being, or health condition, among other aspects of people's life (Tselios et al., 2015).

The spatial scale at which such segregation unfolds matters: in a city with large segregated areas, individuals from different groups are distant from each other, and thus less likely to encounter. This impedes interaction between groups, which is found to further contribute to inequality (Tóth et al., 2021). Several studies have proposed methods to determine the spatial scale of segregation, and the factors influencing it (Veneri et al., 2021). For instance, Petrović et al. (2018) do that by assessing the variation of a scale-dependent segregation indicator. Their indicator measures the social diversity in each neighborhood's local environment, defined by a varying spatial extent around them, called scale. By computing the segregation indicator for a wide range of scales, and studying its maxima, they identify scales of interest. However, the indicator is aggregated at the city level, and does not convey the size, nor the number of segregated areas in the city.

Hitherto, studies have not drawn the size distribution of segregated areas in cities, while it could help understanding how segregation unfolds. This study proposes a direct approach to measure the size of segregated areas, by delineating their spatial extent. Using the proposed approach, we are able to make the following substantive contributions: we determine the size distribution of segregated areas per city, and the extent to which spatial segregation can be characterized by a representative geographical scale in the Netherlands. The size distribution per city could then be exploited further by relating it to geographic, demographic, and urban characteristics of cities to determine the leverage of municipalities in reducing the spatial extent of segregation.

In this study, we focus on the spatial segregation of people with a non-western migration background in the Netherlands. In this country, the segregation between Dutch natives and people with a non-western migration background tend to impede interactions between the two groups (Tselios et al., 2015).

Data for this study on the population mix is obtained from Netherlands Statistics (CBS), at a 6-digit postcode resolution level (around $100 \times 100\ \mathrm{m}^2$ in densely populated areas). In this data, an individual is considered as having a migration background if at least one of their parents was born abroad. This study also extracts the buildings footprint and the street layout from the OpenStreetMap database, to define the postcodes' centroids and to compute the walking time between them.

2 Method

2.1 Measuring exposure

Exposure to a given group is understood here as the potential to encounter an individual from this group. It is computed for a spatial unit using the potential presence in the zone of individuals from each group. Since we do not have data on the actual presence of people from the different social groups in the zones, we model it using the visitation law determined by Schläpfer et al. (2021).

2.1.1 Travel impedance

The shortest walking distance from any postcode to any other postcode is computed using the street network. The walking time is computed from the walking distance, using a walking speed of 4.5 km/h. Then, for a given destination zone, we determine the origin zones located within an acceptable walking distance from it. The inhabitants living in a zone j able to visit the destination zone i are weighted based on the walking time t_{ij} and the travel impedance function $w(t_{ij})$ described by equation 1. This impedance is derived from the work of Schläpfer et al. (2021). A constant factor is set to 3600 s^2 in $w(t_{ij})$, so that $w(t = 60s)$ is 1. This constant does not affect the subsequent calculation of exposure. Finally, we set a minimum walking time to 1 minute.

$$w(t_{ij}) = \begin{cases} 1 & \text{if } 0 \leq t_{ij}[s] < 60 \\ \frac{3600}{t_{ij}^2} & \text{if } 60 \leq t_{ij} < 1200 \\ 0 & \text{if } t_{ij} \geq 1200 \end{cases} \tag{1}$$

2.1.2 Exposure indicator

Exposure to group k in zone i, here defined as the potential $E_i(k)$ to encounter someone from group k in zone i is computed using formula 3. We first calculate the number of individuals from group k living in all neighborhoods j able to reach zone i — $n_j(k)$ is the population from group k living in j —, weighted by the impedance function $w(t_{ij})$ (see equation 2). This value reflects the potential for group k to be present in zone i, coined the *presence* in this work. We compute it for all groups. The exposure to group k in zone i, $E_i(k)$ is the ratio between the presence of group k, and the presence of all groups (equation 3).

$$presence_i(k) = \sum_j w(t_{ij}) \cdot n_j(k) \tag{2}$$

$$E_i(k) = \frac{\sum_j w(t_{ij}) \cdot n_j(k)}{\sum_{k'} \sum_j w(t_{ij}) \cdot n_j(k')} \tag{3}$$

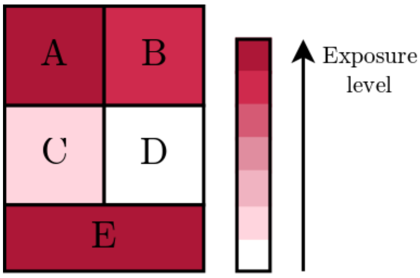
2.2 Detection of spatially segregated areas

To detect segregated areas and determine their geographical demarcation we use agglomerative clustering. This clustering technique groups spatial units together into spatially continuous areas, in which exposure is by and large homogeneous.

2.2.1 Agglomerative clustering

This subsection describes the clustering analysis. In the initialization phase, all zones are considered as independent clusters. Then, we merge iteratively the most similar contiguous clusters, in terms of exposure. In this work, we choose the Ward distance to measure dissimilarity, shown in equation 4, minimizing the within-cluster variance. Figure 1 illustrates the agglomerative process in a city composed of 5 zones. In this work, we ban merging operations resulting in a non-continuous area (zone A and E are not adjacent) using a connectivity matrix. The dendrogram summarizes the successive merges of clusters, as well as the distances between merged clusters.

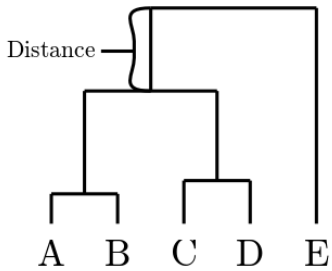
$$d(i \cup j, k) = \sqrt{\frac{(n_i + n_k)d(i,k) + (n_j + n_k)d(j,k) - n_k d(i,j)}{n_i + n_j + n_k}} \tag{4}$$



Map representation of exposure in a city composed of 5 zones.

$$\begin{array}{c|c} & \begin{matrix} A & B & C & D & E \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \end{array}$$

Connectivity matrix corresponding to the city's topology.



Dendrogram obtained when aggregating zones together.

Figure 1: An example of a dendrogram (right) obtained after applying agglomerative clustering on a city composed of 5 zones with different exposure levels (left), using the connectivity matrix corresponding to the city's topology (middle).

The agglomerative clustering stops when all possible merging operations would result in the aggregation of two dissimilar clusters: when the dissimilarity between clusters to be merged exceeds a certain threshold. We tested various threshold levels, and set its value to the ratio 2.5 multiplied by the average exposure in the city.

3 Results

3.1 Exposure and detection of segregated areas

We compute exposure to individuals with a non-western migration background in all 6-digits postcodes of all Dutch municipalities. As an illustration, we display the exposure in the city of Leiden, as well as the segregated areas detected by our method in figure 2. In this map, exposure is expressed in relative terms, in comparison to a situation where groups would be evenly distributed across the city. This means that if the relative exposure is greater than 100% in a zone, then the potential to encounter someone with a non-western migration background is larger than the city's average.

Figure 2: Exposure to individuals with a non-western migration background in Leiden, compared to the benchmark scenario: all groups are evenly distributed across the city. The black contours delineate the segregated areas detected.

The results in figure 2 illustrate that our clustering analysis detects well the areas in which the exposure to individuals with non-western migration background is above the city's average. Similar figures are produced for all municipalities in the data set. The next step consists of computing the surface of these areas and investigating how their size is distributed.

3.2 Size distribution of segregated areas

When applied to all cities in the data set, the clustering analysis detects around 3,000 segregated areas in the Netherlands. Figure 3 shows how their size is distributed. While the vast majority of segregated areas are smaller than 3 km^2 (around 95%), a significant number of areas are considerably larger. This pattern indicates that it is impossible to set a scale that would represent well segregation in the Netherlands.

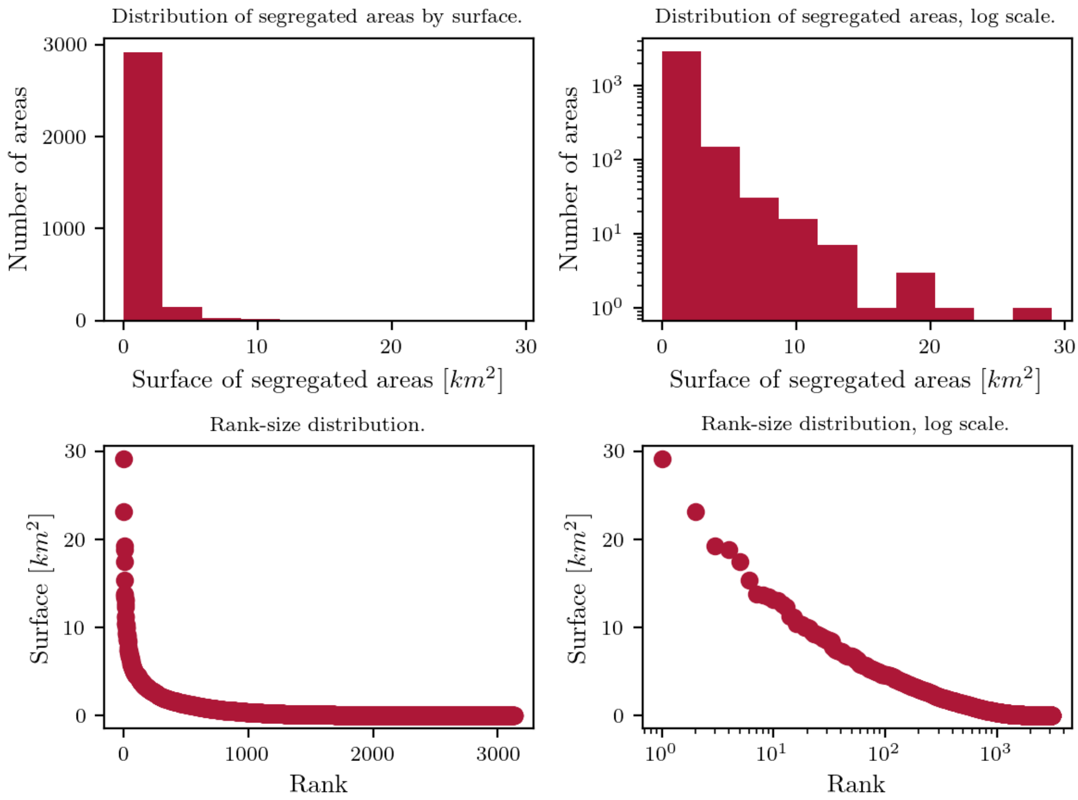


Figure 3: Size distribution of segregated areas in the Netherlands. The top histograms show the number of segregated areas per size group, in linear and logarithmic scale. The bottom two graphs are rank-size distribution: areas are ordered by size and the area's surface is plotted against the area's rank.

4 Conclusion and outlook

This study proposes a novel data-driven approach to delineate the geographical demarcation of segregated areas. We applied the proposed method to assess segregation of individuals with a non-western migration background in all Dutch municipalities, and drawn the size distribution of segregated areas in these cities. This study demonstrates that there is no characteristic scale for spatial segregation in the Netherlands.

The results will be explored further to determine whether the size distribution varies at different scales (e.g. city vs country), or whether some cities host segregated areas of a certain scale (some cities have many a large number of small segregated areas, while other have few large ones). After investigating the variation in size of segregated areas, we will relate the differences between cities to geographic, demographic and urban planning characteristics.

References

Petrović, A., van Ham, M., and Manley, D. (2018). Multiscale measures of population: Within- and between-city variation in exposure to the sociospatial context. *Annals of the American Association of Geographers*, 108(4):1057–1074.

Schläpfer, M., Dong, L., O'Keeffe, K., Santi, P., Szell, M., Salat, H., Anklesaria, S., Vazifeh, M., Ratti, C., and West, G. B. (2021). The universal visitation law of human mobility. *Nature*, 593:522–527.

Tselios, V., Noback, I., van Dijk, J., and McCann, P. (2015). Integration of immigrants, bridging social capital, ethnicity, and locality. *Journal of Regional Science*, 55(3):416–441.

Tóth, G., Wachs, J., Clemente, R. D., Ákos Jakobi, Ságvári, B., Kertész, J., and Lengyel, B. (2021). Inequality is rising where social network segregation interacts with urban topology. *Nature Communication*, 12.

Veneri, P., Comandon, A., Àngel Garcia-López, M., and Daams, M. N. (2021). What do divided cities have in common? an international comparison of income segregation. *Journal of Regional Science*, 61:162–188.