RESIDENTIAL SEGREGATION IN MESSINA (SOUTHERN ITALY): AN INFORMATION THEORY BASED STUDY

Francesca Bitonti, Angelo Mazza, Daniela Ghio, Massimo Mucciardi

Abstract. Residential segregation stands out as one of the most noticeable and potentially concerning consequences of urbanization. Adopting the framework proposed by the Information Theory, the study investigates residential segregation patterns in the Italian municipality of Messina that has recently experienced deteriorating urban conditions. We rely on anonymized individual data sourced from the Population Register to examine the major immigrant groups residing in Messina in 2016 and 2022, Sri Lankans, Filipinos, Romanians, and Moroccans. The analysis computes the Shannon's entropy index and Kullback-Leibler (KL) divergence, aiming at: 1. drawing comparisons in the residential segregation patterns among immigrant populations; 2. appraising changes in residential patterns between 2016 and 2022; 3. assessing to what extent ethnic concentration depends on the adoption of different territorial scales to classify metropolitan areas. Results reveal nuanced patterns of residential segregation among the selected migrant populations, with Filipinos and Moroccans remaining the most segregated groups, both in 2016 and 2022. However, two common dynamics are affecting all immigrant groups: a. the presence of micro-scale segregation; b. the increase of segregation degrees over time. Furthermore, when comparing the distribution of immigrant groups with native populations, concentration levels, detected by the Shannon's entropy index, have not always implied significant KL divergence. These results suggest complex interactions between migrant and the local populations, challenging simplistic assumptions about segregation. Accounting for the multi-scalar dimensionality of segregation, this study contributes to a deeper understanding of residential dynamics and provides insights for fostering social cohesion in diverse spatial urban settings.

1. Introduction

Residential segregation, as complex multidimensional phenomenon (Massey and Denton, 1988), has been extensively studied through the lens of the Information Theory. In this framework, concepts, such as information and entropy, have been revived (Theil and Finizza, 1971). Entropy appraises the degree of randomness in a system or the informational content of a message (Coulter, 1989; Cover and Thomas, 2006; Shannon, 1948; Theil, 1967). Introduced to the social sciences by Theil (Theil, 1967; 1972), entropy became a valuable tool for assessing population diversity and

income inequality (Reardon and Firebaugh, 2002; White, 1986). Essentially, entropy can be interpreted as a measure of uncertainty, representing the amount of information needed to describe a probability distribution. Social researchers apply entropy to gauge residential segregation, as it quantifies the homogeneity of neighbourhoods or regions based on demographic and social characteristics, like ethnicity, education level, or income. In residential segregation, high entropy indicates an equal probability distribution of population groups across space, while lower entropy signifies a higher probability for one group, reflecting reduced uncertainty about the distribution (Cover and Thomas, 2006).

Previous studies on residential segregation, particularly ethnic segregation, have emphasized the importance of setting appropriate spatial scales and implications of this methodological choice in its measurement (Arcaya *et al.*, 2018). Distinctive racial residential patterns, and consequently a wide variation in segregation, become evident when changing the range of spatial scales (Reardon *et al.*, 2008). This is due to the decision-making processes regarding where to live within a city. Excluding public housing allocations, residential decisions are individual-group-varying and involve: 1) housing rental prices, 2) access to work, school, and leisure (for single and potential family members), influenced by the presence and efficiency of local services, 3) proximity to significant others like family or co-ethnics. Consequently, some groups might be highly concentrated in certain parts of a metropolitan area but scattered within those parts, while others might form tight, exclusive clusters spread across different housing market segments (Jones *et al.*, 2015). For these reasons, different levels of residential segregation can be detected at various spatial scales.

Utilizing anonymized individual-level data coming from the Population Register, this paper investigates the residential segregation patterns of migrants within the metropolitan area of Messina, Italy. The study focusses on the four largest immigrant groups — Sri Lankans, Filipinos, Romanians, and Moroccans — in 2016 and 2022. The primary inquiries center around 1) Carrying out a multi-scalar analysis of the segregation levels exhibited by the selected immigrant groups; 2) Measuring ethnic residential segregation patterns by immigrant group; 3) Leveraging entropy and divergence when comparing the spatial distribution of immigrants with native populations over time

The reimaging part of the paper includes four sections. The next section analyzes data and the methodology implemented, while the third section portrays the geographical area of Messina and depicts descriptive migrant population patterns. Then, the last two sections are dedicated to the presentation of results and the discussion of the main findings and related implications, drawing the conclusions of the work.

2. Data and Methods

Characteristics of immigrants' groups come from the Population Register, accounting for individuals recorded as of June 30th, 2016 and November 30, 2022. This dataset includes all individuals (Italians and immigrants) residing in Messina, along with their children or nephews who were born abroad (in their respective countries of origin) and obtained solely the citizenship of parents at birth. Their residential addresses have been geocoded by querying the Google Maps Geocoding API exploiting the R "ggmap" library (Kahle and Wickham, 2013). The statistical analysis was conducted with the R software (R Core Team, 2023).

In this study, the residential segregation of the selected foreign groups is examined through the concept of entropy, a measure often utilized in Information Theory. Entropy represents the amount of information required to describe a (in our case spatial) probability distribution. When two outcomes (e.g., two ethnic groups) are equally probable, the uncertainty about the outcome is high, resulting in high entropy. Conversely, if one outcome is more likely than the other, there is less uncertainty and, consequently, lower entropy. As a consequence, the higher the entropy, the lower the segregation[1]. In this context, Shannon's entropy metric is calculated (Shannon, 1948). For a categorical variable X with I possible outcomes (or groups), Shannon's entropy is defined as:

$$H(X) = \sum_{i=1}^{I} p(x_i) log \frac{1}{p(x_i)} \quad (1)$$

where $p(x_i)$ for $i = 1, 2, \dots, I$ is the probability of occurrence of group i in a given area. The relative entropy, measuring the discrepancy of the immigrant groups' probability distributions from the local population's one, is also evaluated. This involved computing the Kullback-Leibler (KL) divergence (Kullback, 1987) to assess the spatial distribution differences between Italians and each of the four immigrant groups individually, as follows:

$$D(p|q) = \sum_{m=1}^{M} p_m log \frac{p_m}{q_m} \quad (2)$$

where the q distribution defines the reference (Italians) against which the p distribution of the immigrant group is compared in a given area m. In general, the KL divergence is able to better control the contextual factors influencing residential

[1] In residential segregation studies, entropy is inversely related to the concept of diversity, intended as ethnic or group mixing: higher diversity corresponds to lower segregation.

choices in general (e.g., building distribution, industrial areas etc.) because it compares the distributions themselves.

3. Context and descriptive results

Over the past fifty years, Messina has been struggling with challenges due to deteriorating urban conditions, a rise in youth emigration, and a shift in population towards neighbouring villages, leading to a decline and spatial redistribution of its population (Scrofani, 2018).

Table 1 – *Demographic characteristics of the four main immigrant groups residing in the metropolitan areas of Messina, 2016 and 2022.*

Characteristic	Sri Lanka		Philippines		Romania		Morocco	
	2016	**2022**	**2016**	**2022**	**2016**	**2022**	**2016**	**2022**
Age								
0-25	30.7 %	28.3 %	29.7 %	25.8 %	21.7 %	18.8 %	28.8 %	27.1 %
26-65	66.7 %	64.7 %	66.2 %	64.4 %	76.7 %	76.5 %	66.6 %	65.1 %
65+	2.6 %	6.9 %	4.1 %	9.8 %	1.6 %	4.7 %	4.6 %	7.8 %
Sex								
Female	46.8 %	49.0 %	52.8 %	53.0 %	68.0 %	67.3 %	36.6 %	38.7 %
Male	53.2 %	51.0 %	47.2 %	47.0 %	32.0 %	32.7 %	63.4 %	61.3 %
N° individuals	4199	4030	2555	2226	1661	1668	1218	1155
% of total immigrant population	33.2	30.3	20.2	16.8	13.1	12.6	9.6	8.7
N° households	1742	1823	902	917	1015	1122	573	636
% of total number of immigrants' households	27.6	24.1	14.3	12.1	16.1	14.8	9.1	8.4
Mean n° of individuals in households	2.4	2.2	2.8	2.4	1.6	1.5	2.1	1.8

Source: authors' elaborations on Population Register data.
Notes: 2016 observations refer to 30.06.2016: 2022 observations refer to 30.11.2022

From the 1980s onwards, the city has received the inflows of immigrants mostly coming from Sri Lanka, the Philippines, Romania and Morocco. These first four groups accounted for the 68.35% of the total foreign population in 2022 (against 76.2% in 2016, suggesting an increase in the ethnic diversification of the population – table 1). Their demographic characteristics show important similarities and

differences. Beside a general shrinkage in the immigrant groups' size, also the mean household size has decreased over time, with Sri Lankans and Filipinos accounting for the highest mean of 2.2 and 2.4 individuals per household as for 2022, respectively. The population pyramids (figure 1) illustrate that overall, the middle and adult age classes are the most numerous. As regard the gender composition, Sri Lankans and Filipinos are balanced with a bottleneck for the young adult age classes, whereas a gender characterization can be observed for Romanians (with a female dominance) and Moroccans (with a male prevalence) (see both figure 1 and table 1).

The four immigrant groups are settled in the urban core of the municipality of Messina and along the seaside, following the overall resident population density (figure 2). Nevertheless, Sri Lankans and Romanians seem to be more dispersed than the other groups, being in more peripheral northern and in the western inner areas.

4. Results: residential segregation patterns

The computation of the H index and the KL divergence at the district[2] level (mean area: 35.33 km^2) yielded different results.

According to the H index Sri Lankans (table 2) and Filipinos appear more mixed with the Italians and, in general, the H values for all the groups do not vary much between the two years.

The KL divergence, on the contrary, highlights a general increase of segregation through time for all the groups with Sri Lankans and Romanians being the least segregated with respect to Italians.

Considering the KL values together with demographics, it is possible to draw residential profile by immigrant group. Specifically, in 2022:

- Filipinos: highly segregated in «middle-size household» groups (mean household size: 2.2)
- Moroccans: segregated in «small household» groups (mean household size: 1.8)
- Sri Lankans: segregated in «middle-size household» groups (mean household size: 2.1)
- Romanians: poorly segregated in «small household» groups (mean household size: 1.5)

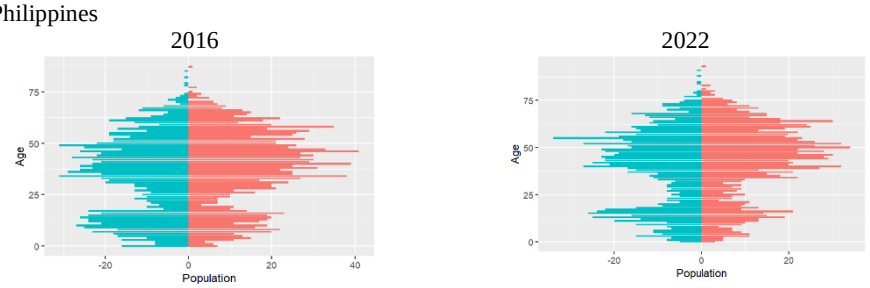
Overall, with the increase of the household size the KL divergence, and hence the segregation, unsurprisingly increases too.

[2] The city is divided into six municipal districts, or "*circoscrizioni*" (I to VI), which function as subdivisions of the urban area, similarly to the municipal arrondissements of Paris or the London boroughs.

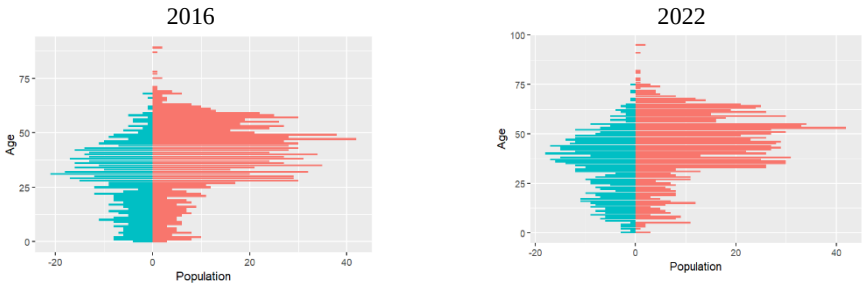
Figure 1 – *Population structure for the main four foreign groups in the municipality of Messina, 2016 and 2022.*

a) Sri Lankans

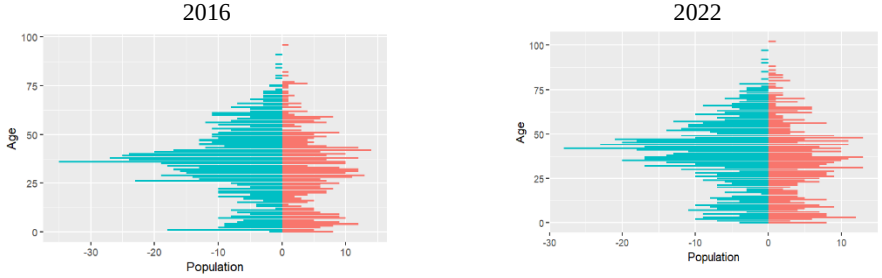
b) Philippines



c) Romanians

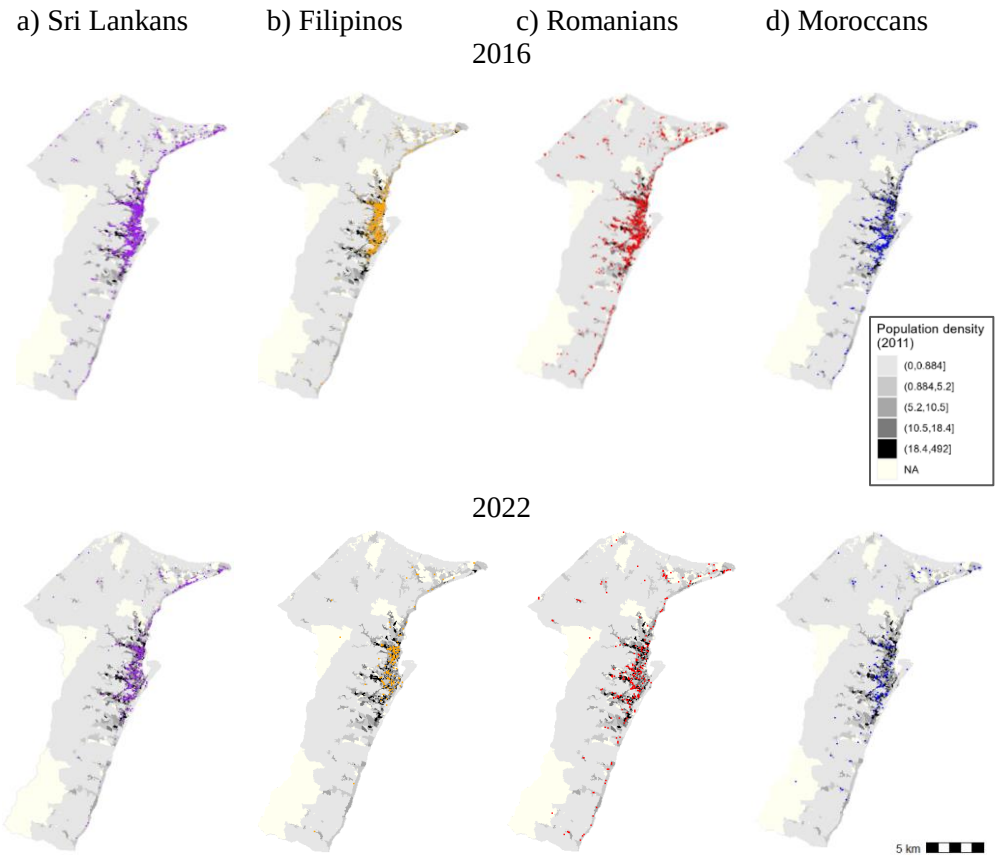


d) Moroccans



Note: females in pink; males in light blue.

Figure 2 – *Spatial distribution of the main four foreign groups in the municipality of Messina in 2016 and 2022.*



Note: the points referring to the individual residential addresses have been jittered for privacy concerns.

Table 2 – *Shannon's entropy index and KL divergence values for the four selected immigrant groups against Italians in Messina in 2016 and 2022 at the district level.*

Foreign group	Shannon's entropy H(X)		KL divergence	
	2016	2022	2016	2022
Sri Lankans	0.129	0.146	0.070	0.132
Filipinos	0.087	0.072	0.207	0.309
Romanians	0.061	0.072	0.050	0.073
Moroccans	0.047	0.048	0.169	0.164

As already pointed out by a large body of work (Leckie et al., 2012; Louf, 2016; Östh et al., 2015), segregation is not a scale-invariant phenomenon. Instead, high variation in segregation level can be detected at different scales of measurement depending on several demographic, socioeconomic but also urban factors. For this reason, the KL divergence was measured at different scales by superimposing to the urban area of Messina several regular grids composed by quadrats of different side's lengths (i.e.: 100, 200, 500, 1000 meters). As reported in table 3, except the general decrease in divergence with the time, for both years, the higher the scale, the lower the segregation for all of the groups considered. This should imply that if at small scales some sort of divergence is detected, there is lower level of concentration for these groups when lager scales are adopted.

Table 3 – *Multilevel dynamics of the KL divergence against Italians.*

Foreign group	2016				2022			
	Quadrat side's length (meters)							
	100	200	500	1000	100	200	500	1000
Sri Lankans	0.922	0.752	0.532	0.404	1.522	1.125	0.721	0.490
Filipinos	1.357	1.161	0.892	0.760	2.450	2.039	1.448	0.978
Romanians	0.692	0.514	0.332	0.217	2.100	1.613	0.910	0.499
Moroccans	1.420	1.150	0.743	0.524	2.218	1.834	1.200	0.734

5. Conclusions

The results showed that the concentration assessed by the Shannon index do not always simply a divergence in KL terms. The KL metrics highlighted heterogeneity in the levels of divergence when comparing the spatial distribution of selected immigrant groups with Italians. This mirrors potential differences in integration dynamics, not just in terms of sharing space, but also different life domains.

Yet, KL divergence results suggest disparities in segregation levels across foreign groups that interact with and are exposed to the same environments as the local population. This could imply heterogeneous levels of integration between the different foreign groups and the local population, not just in terms of sharing the same neighbourhoods, but also similar workplaces and social contexts. In particular, the analysis detected a general increase of segregation for all the groups investigated with the passing of the time. Despite this, the consideration of different scales of analysis ruled out the presence of large neighbourhoods of segregation. Nonetheless, the different divergence levels characterizing the groups, especially at the smallest scales, are evidence that group-specific policies should be implemented to guarantee the integration of all the groups.

The present work represents the initial effort made to map and understand the evolution of residential allocation patterns of the main immigrant groups in Messina. The influence of additional individual, group-specific and contextual socioeconomic factors on residential choices will be assessed in future steps of the analysis to unveil the causes, dynamics and possible socioeconomic implications of residential segregation in Messina. Some of the research questions that will be investigated concern the potential differences in the spatial distribution of the first and second generations of immigrants, the individual residence movements across time, the socioeconomic factors influencing residential choices, as job opportunities and spatial heterogeneities in housing cost (as partially investigated in Bitonti *et al.* (2023)), as well as the differences in settlement models based on ethnic-specific migratory behaviours: e.g., family migration, family reunion and "solo-migration".

Acknowledgements

This study has been partially supported by the PRIN-PNRR 2022 research project "Foreign population and territory: integration processes, demographic imbalances, challenges and opportunities for the social and economic sustainability of the different local contexts (For.Pop.Ter)" [P2022WNLM7], funded by European Union—Next Generation EU, component M4C2, Investment 1.1. It has also been partially supported by the Italian Ministerial grant PRIN 2022 "Depopulation Risk: Experiences, Mobility and Subjective Well-being (DREAMS)" [2022RNKSEL] CUP E53D23010380006. It has also been partially supported by the PNRR - Missione 4 "Istruzione e Ricerca" - Componente 2 "Dalla ricerca all'impresa" - Investimento 1.3 - *Partenariato Esteso "GRINS – Growing Resilient, INclusive and Sustainable",* cod. PE0000018, CUP E63C22002120006, Spoke 8 - "*Social sustainability",* finanziato dall'Unione Europea – NextGenerationEU.

References

ARCAYA M.C., SCHWARTZ G., SUBRAMANIAN S. 2018. A multi-level modeling approach to understanding residential segregation in the United States. *Environment and Planning B: Urban Analytics and City Science*, Vol. 45, No. 6, pp. 1090–1105.

BITONTI F., MAZZA A., MUCCIARDI M., SCROFANI L. 2023. Urban transformations and the spatial distribution of foreign immigrants in Messina. *Models for Data Analysis. Selected papers of 49th Meeting of Italian Statistical Society*, pp.53-67.

COULTER P.B. 1989. *Measuring Inequality: A Methodological Handbook*. Boulder: Westview Press.

COVER T.M., THOMAS J.A. 2006. *Elements of Information Theory*. Hoboken, N.J.: Wiley-Interscience.

JONES K., JOHNSTON R., MANLEY D., DEWI O., CHARLTON C. 2015. Ethnic Residential Segregation: A Multilevel, Multigroup, Multiscale Approach Exemplified by London in 2011. *Demography*, Vol. 52, pp. 1995–2019.

KAHLE D., WICKHAM H. 2013. ggmap: Spatial Visualization with ggplot2. *The R Journal*, *5*(1), 144–161.

KULLBACK S. 1987. Letters to the Editor. *The American Statistician*, Vol. 41, pp. 338–341.

LECKIE G., PILLINGER R., JONES K., GOLDSTEIN H. 2012. Multilevel modeling of social segregation. *Journal of Educational and Behavioral Statistics*, Vol. 37, No. 1, pp. 3–30.

LOUF R. 2016. Patterns of residential segregation. *PLoS ONE*, No. e0157476.

MASSEY D.S., DENTON N.A. 1988. The dimensions of residential segregation. *Social Forces*, Vol. 67, pp. 281–315.

ÖSTH J., CLARK W.A.V. MALMBERG B. 2015. Measuring the scale of segregation using k-nearest neighbor aggregates. *Geographical Analysis*,Vol. 47, No. 1, pp. 34–49.

R CORE TEAM. 2023. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

REARDON S.F., FIREBAUGH G. 2002. Measures of Multigroup Segregation. *Sociological Methodology*, Vol. 32, pp. 33–67.

REARDON S.F., MATTHEWS S.A., O'SULLIVAN D., LEE B.A., FIREBAUGH G., FARRELL C.R., BISCHOFF K. 2008. The Geographic Scale of Metropolitan Racial Segregation. *Demography*, Vol. 45, No. 3, pp. 489–514.

SCROFANI L. 2018. Le aree urbane nei processi di periferizzazione e di sviluppo del Mezzogiorno. In *Sussidiarietà e… giovani al Sud. Rapporto sulla Sussidiarietà 2017/2018*, pp. 167–198. Fondazione Sussidiarietà.

SHANNON C.E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*, Vol. 27, No. 3, pp. 379–423.

THEIL H. 1967. *Economics and Information Theory*. Amsterdam: North- Holland Publishing Company.

THEIL H. 1972. *Statistical Decomposition Analysis.* In THEIL H. (Ed.). Amsterdam: North- Holland Publishing Company.

THEIL, H., FINIZZA, A.J. 1971. A note on the measurement of racial integration of schools by means of informational concepts. *The Journal of Mathematical Sociology*, Vol. 1, No. 2, pp. 187–193.

WHITE M.J. 1986. Segregation and Diversity Measures in Population Distribution. *Population Index*, Vol. 52, No. 2, pp. 198–221.

Francesca BITONTI, University of Catania, francesca.bitonti@unict.it
Angelo MAZZA, University of Catania, a.mazza@unict.it
Daniela GHIO, University of Catania, daniela.ghio@unict.it
Massimo MUCCIARDI, University of Messina, massimo.mucciardi@unime.it