



Delft University of Technology

Document Version

Final published version

Citation (APA)

Spierenburg, L. J. (2026). *Advances in the analysis of residential segregation and urban riots*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:50568802-0f7d-4d4c-8534-da0c812260f6>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

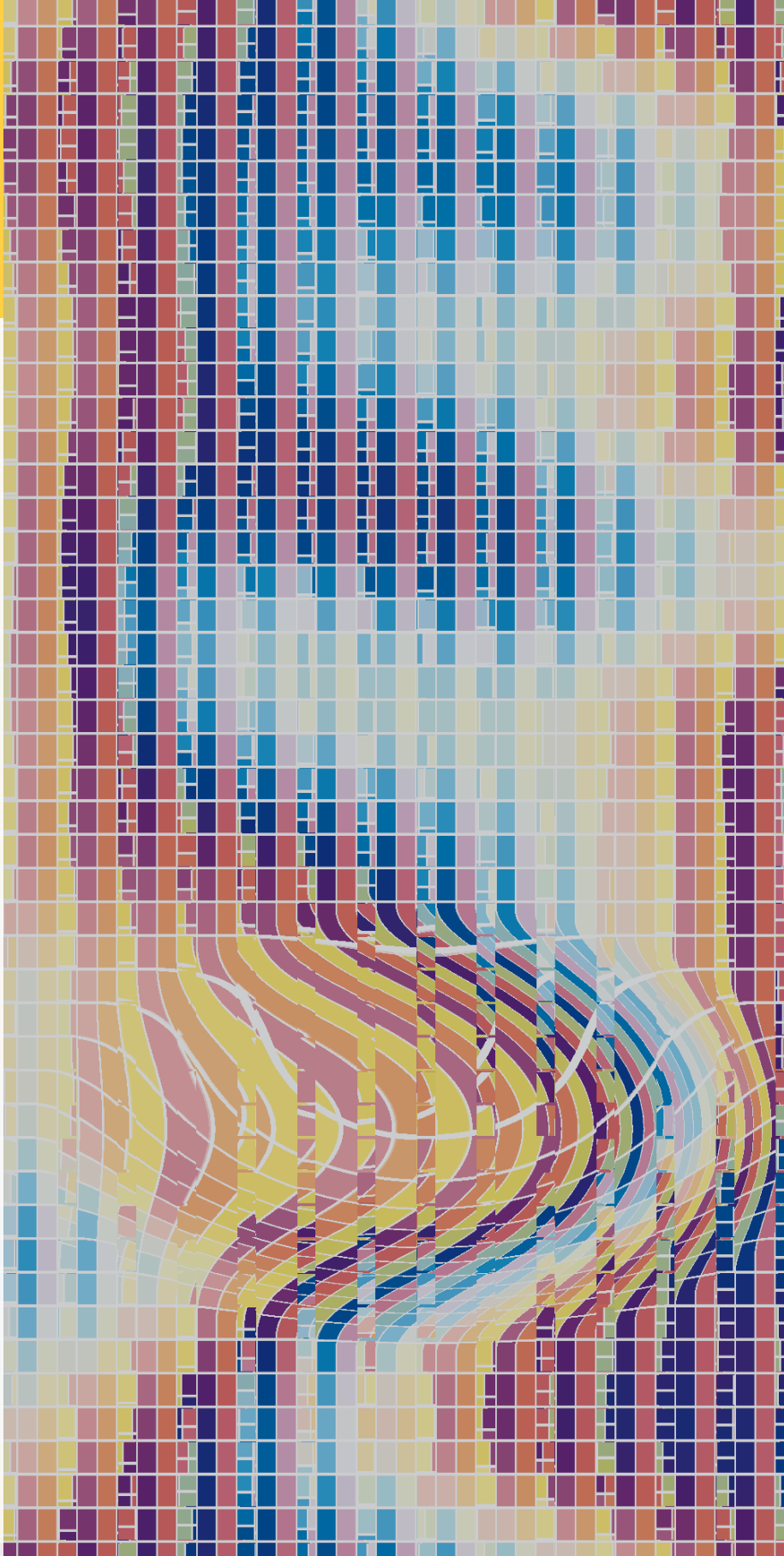
Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.



ADVANCES IN THE ANALYSIS OF RESIDENTIAL SEGREGATION AND URBAN RIOTS

by Lucas Spierenburg

Advances in the analysis of residential segregation and urban riots

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof. dr. ir. H. Bijl

Chair of the Board for Doctorates
to be defended publicly on Wednesday, 22 April 2026, 10:00

by

Lucas Johannes SPIERENBURG

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus

Dr. ir. S. van Cranenburgh

Prof. dr. O. Cats

chairperson

Delft University of Technology, promotor

Delft University of Technology, promotor

Independent members:

Prof. dr. M. van Ham

Prof. dr. M. Olteanu

Dr. L. M. Alessandretti

Dr. C. M. Lima Azevedo

Prof. dr. M. E. Warnier

Delft University of Technology

Paris Dauphine University, France

Technical University of Denmark, Denmark

Technical University of Denmark, Denmark

Delft University of Technology, *reserve member*



AI Initiative



The research in this thesis is supported by the TU Delft AI Labs programme.

Front & Back covers: The cover artwork is derived from code originally written by Naoto Hieda, published in the *Hydra Book* ("Textures" chapter), available at <https://hydra.ojack.xyz/docs/docs/learning/guides/textures> (accessed 2026/03/02). The original code is used under CC BY 4.0; modifications were made by Julija Cesnulaityte and Lucas Spierenburg.

TRAIL Thesis Series no. T2026/9, The Netherlands Research School TRAIL

TRAIL

P.O. BOX 5017

2600 GA Delft

The Netherlands

E-mail: info@rsTRAIL.nl

ISBN: 978-90-5584-384-8

Copyright © 2026 by Lucas Spierenburg

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission of the author.

An electronic version of this dissertation is available at <http://repository.tudelft.nl/>.

Contents

Summary	vii
Samenvatting (Summary in Dutch)	xiii
1 Introduction	1
1.1 Background	1
1.1.1 Social tensions in cities.	1
1.1.2 Residential segregation	2
1.1.3 Riots	4
1.2 Emerging fine-resolution data sources for urban analysis	6
1.2.1 Fine-resolution demographic data for mapping segregation . . .	6
1.2.2 Social media data for mapping rioting events	7
1.3 Problem statement	7
1.3.1 Limited spatial understanding of residential segregation.	8
1.3.2 Limited spatial analysis of urban riots.	8
1.4 Research objectives.	9
1.5 Contributions.	10
1.5.1 Methodological contributions.	10
1.5.2 Empirical contributions.	10
2 Characterizing residential segregation in cities	13
2.1 Introduction	13
2.2 Method.	15
2.2.1 Indicators	16
2.2.2 Delineation of spatially segregated regions	19
2.2.3 Potential exposure	21
2.3 Case study and datasets.	23
2.3.1 Case study	23
2.3.2 Datasets	23
2.3.3 Reflection on the modifiable areal unit problem	24
2.3.4 Code and data produced	25

2.4	Results	25
2.4.1	Characterization of segregation patterns in the Netherlands . . .	26
2.4.2	Association with municipality characteristics	28
2.5	Conclusion	31
3	Evolution of residential segregation in the Netherlands, 2015-2020	33
3.1	Introduction	33
3.2	Literature review and theoretical background	35
3.2.1	Segregation indicators	35
3.2.2	Comparative analysis	36
3.2.3	Data-driven approaches.	36
3.2.4	Theoretical framework	37
3.3	Method.	38
3.3.1	Spatial moving average.	39
3.3.2	Normalization	39
3.3.3	Statistical transformation	40
3.3.4	Regionalization.	40
3.3.5	Labeling regions	40
3.3.6	Residential segregation indicators.	41
3.4	Data and case study	41
3.4.1	Dataset.	41
3.4.2	Situation in 2015	42
3.5	Results	44
3.5.1	Descriptive statistics	44
3.5.2	Geographical trends	45
3.5.3	Convergence of segregation patterns	45
3.5.4	Relation between share of population with a non-western migration background and change in segregation patterns	47
3.6	Discussion and conclusion	48
3.6.1	Interpretation of the patterns observed	49
3.6.2	Limitation and further research	49
4	Misalignment between residential segregation and urban barriers	51
4.1	Introduction	51
4.2	Results	55
4.2.1	Absence of a uniform alignment between residential segregation and urban fragmentation across European cities.	55

4.2.2	Greater association between residential segregation and urban fragmentation in Germany and the Netherlands than in other countries	58
4.3	Discussion	60
4.3.1	Limitations.	62
4.3.2	Recommendations for further research	63
4.4	Methods	63
4.4.1	Data sources and preparation	64
4.4.2	Quantifying alignment between social frontiers and urban barriers.	65
4.4.3	Post-hoc statistical testing	68
5	Studying riots through the lens of social media	71
5.1	Introduction	72
5.2	Case study and datasets.	73
5.2.1	Case study	73
5.2.2	Social media data.	74
5.2.3	Press data	75
5.3	Method.	75
5.3.1	Computer vision model.	76
5.3.2	Spatiotemporal analysis	78
5.4	Results	80
5.4.1	Validation of the computer vision model	80
5.4.2	Spatiotemporal analysis of riots.	84
5.5	Discussion and conclusion	87
6	Conclusion	91
6.1	Main conclusions.	91
6.1.1	What types of spatial structure does residential segregation exhibit across urban contexts? (chapter 2).	91
6.1.2	How does the spatial structure of residential segregation patterns evolve over time and relate to demographic change at the municipal scale? (chapter 3)	92
6.1.3	To what extent the spatial structure of residential segregation follows urban barriers? (chapter 4)	93
6.1.4	How to identify riot-related content from visual social media data, and what does this reveal about the spatiotemporal nature of riots?(chapter 5)	93

6.2	Limitations and directions for future work	95
6.2.1	From "what patterns exist" to "how patterns emerge"	95
6.2.2	Understanding the mechanisms underlying segregation-barrier alignment	96
6.2.3	Modeling spatiotemporal dependencies in riot propagation . . .	96
6.2.4	Toward an integrated spatial analysis of segregation and riots . .	97
6.3	Implications for policymakers	98
Bibliography		101
About the Author		121
TRAIL Thesis Series publications		125

Summary

Cities are spatial concentrations of population that facilitate connections between individuals. By bringing people together, they create environments where social groups with different backgrounds and resources — whether defined by education, ethnicity, class, or other characteristics — interact and exchange. When these differences become sources of conflict, they generate social tensions which contribute to phenomena like residential segregation or riots. Residential segregation represents a latent, chronic expression, where social divisions are embedded within spatial patterns that evolve over years across broad urban areas. Riots constitute explicit, acute manifestations, emerging as concentrated episodes of violent unrest in specific locations and compressed time frames. These spatiotemporal phenomena have substantial consequences: they reinforce inequalities across education, health, and employment and undermine social cohesion between groups and trust in institutions. Such patterns represent failures in urban governance and planning to facilitate interaction and integration across diverse populations, placing them high on municipal agendas. Such attention underscores the need for analysis of how residential segregation and riots unfold spatially.

An analytical gap persists in understanding the spatial structure of these phenomena - the specific geographic patterns, configurations, and spatial relationships that define how segregation and riots actually unfold across urban landscapes. Despite the availability of fine-resolution demographic data and geolocated social media sources, researchers lack both the methods to leverage these datasets and the empirical applications needed to advance spatial understanding of social tensions. The overarching objective is therefore to reveal the spatial structure of social tensions in cities by developing analytical methods capable of capturing these phenomena at the scales where they actually manifest. This objective is addressed through two research streams: examining how residential segregation patterns organize across urban landscapes and how riots unfold spatiotemporally. The first stream comprises chapters 2 to 4 that investigate how residential segregation patterns organize spatially, developing and applying spatial analysis methods for this purpose. These chapters progress from method development with initial empirical findings in the Dutch context, to temporal analysis examining segregation evolution across Dutch municipalities, to large-scale cross-European examination of spatial alignment between segregation patterns and urban barriers such as railways, roads, and rivers. The second stream comprises chapter 5 that captures the fine spatiotemporal dynamics of riots as they unfold across urban neighborhoods, using computer vision techniques to analyze visual social media data from the French Nahel Merzouk riots (July 2023) as a case study.

Chapter 2 investigates the spatial structure of residential segregation and describes the extent to which it varies across urban contexts. To this end, I implement a regionalization approach that delineates demographically homogeneous regions and develop a framework for characterizing the spatial structure of residential segregation through three dimensions: **intensity** — the degree of overrepresentation of a group in segregated regions compared to the city average —, **separation** — the proportion of the group living in these segregated regions —, and **scale** — the size of these regions in terms of population. The approach is applied to Dutch 6-digit postcode data (spatial units of approximately 50 inhabitants) to study segregation of the population with non-Western migration background across all Dutch municipalities. The analysis reveals substantial diversity in segregation patterns across Dutch municipalities. Six of eight theoretically possible segregation types across Dutch municipalities, with municipalities spanning nearly the full conceptual space rather than expressing a single typical pattern. Figure 1 illustrates this three-dimensional characterization, displaying how Dutch municipalities vary across intensity, separation, and scale. When investigating the relation of these indicators to other city variables, both intensity and separation positively associate with municipality size, social mix, and income inequality, suggesting these associations stem from a common underlying phenomenon rather than independent effects. The absolute scale of segregation correlates linearly with municipality population: larger cities exhibit proportionally larger segregated regions rather than more numerous regions, indicating that segregation patterns in large municipalities are scaled-up versions of the ones in smaller municipalities.

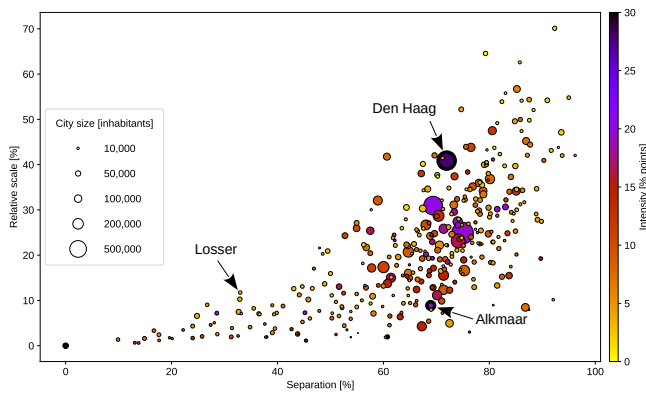


Figure 1 Relation between segregation indicators in Dutch municipalities. The separation and the relative scale are shown on the x and y axes, respectively, while the color of a point depicts the segregation intensity. The size of points is proportional to the municipality population.

Chapter 3 extends the analysis from chapter 2 to track segregation patterns over time. It investigates the temporal evolution of segregation patterns across Dutch municipalities between 2015 and 2020 and their relationship to demographic change. The analysis reveals a convergence of segregation patterns across Dutch municipalities over this

period. Figure 2 displays the spatial distribution of residential mix and the three segregation dimensions in 2015, providing baseline context for understanding these temporal dynamics. Municipalities with initially low intensity and separation experienced increases in these indicators, while those with historically high intensity and separation — particularly the major Randstad cities — experienced decreases, resulting in reduced variance across the country. The increase in the share of population with non-Western migration background positively correlates with increases in separation and scale but not with changes in intensity, suggesting that the increase in population with non-Western migration background leads to spatial expansion of segregated regions or emergence of new ones rather than increased homogeneity within existing ones.

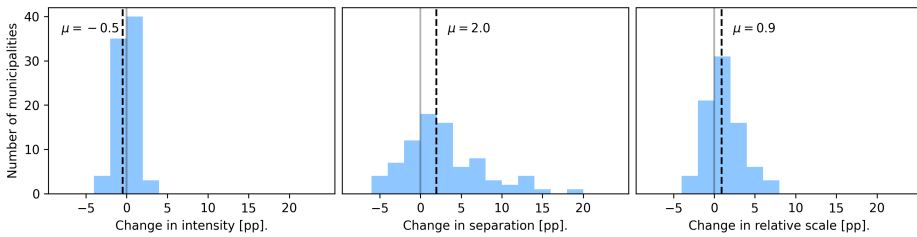


Figure 2 Statistical distribution of the change in intensity, separation, and relative scale in percentage points (pp).

Chapter 4 investigates whether residential segregation aligns with urban fragmentation — the partition of urban space by barriers, such as railways, motorways, and waterways — in contexts without historical segregation policies. While such alignment has been documented in US and South African cities with historical segregation policies, theory suggests it could also emerge naturally from residential preferences. Testing this hypothesis across European contexts addresses an empirical gap with implications for both urban theory and policy. Methodologically, the chapter develops a Monte Carlo approach to test the statistical significance of spatial alignment between demographic regions and urban barriers, focusing on the segregation of population with non-EU migrants. For each of 520 cities across eight Western European countries (Germany, France, Ireland, Italy, the Netherlands, Portugal, Spain, and the UK), the method generates 200 synthetic fragmentation patterns matching the actual size distribution of fragments. I measure the alignment between demographic regions and urban barriers with a cluster similarity metric and assess its significance by comparing its value to the alignment of demographic regions to the 200 synthetic partitions. The findings reveal no systematic alignment between residential segregation and urban fragmentation across European cities — contrary to expectations based on US and South African contexts. Many cities actually exhibit less alignment than expected by chance. Decentralization proves a significant predictor of alignment, while segregation intensity does not. Important regional variations emerge: the Netherlands show significantly stronger alignment than other countries, even after controlling for urban decentralization. These results suggest urban barriers do not generally function as social frontiers in European contexts without

explicit segregation policies, though country-specific urban development patterns may influence this relationship. Figure 3 displays the statistical and geographic distribution of quantile ranks, illustrating both the overall lack of systematic alignment and the concentration of cities exhibiting significant alignment in Germany and the Netherlands.

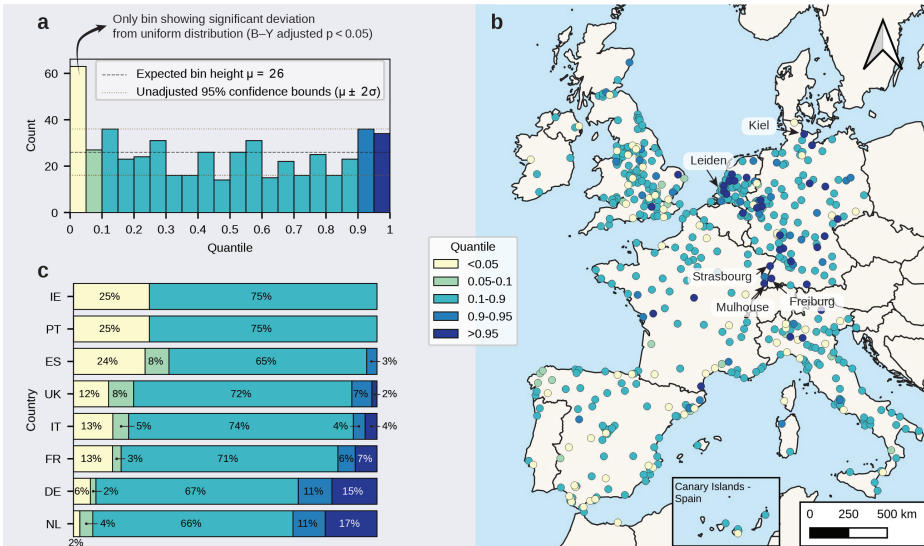


Figure 3 Statistical and geographical patterns of alignment between residential segregation and urban fragmentation across European cities. (a) Distribution of quantile ranks across all cities in the dataset. (b) Geographic distribution of quantile ranks across cities in the dataset. (c) Distribution of quantile ranks across countries.

Chapter 5 shifts focus to mapping riots — the second manifestation of social tensions examined in this thesis. Unlike residential segregation’s latent, chronic patterns, riots represent explicit, acute events that erupt into the open, requiring fine-grain spatiotemporal data to understand their dynamics. Capturing this level of detail has traditionally proven difficult, but social media platforms — where users document events in real-time with precise geolocation — offer new potential to trace how riots unfold. In this chapter, I investigate the spatiotemporal distribution of riots using the 2023 Nahel Merzouk riots as a case study, leveraging social media data from Snapchat — 107,000 posts collected across Paris, Marseille, Lyon, Toulouse, Lille, Bordeaux, and Grenoble. I identify riot-related content from visual posts, by finetuning a computer vision model on an annotated subset of the data, reaching 77% precision and 68% recall on held-out data. Applying the model to the full dataset, results show strong spatiotemporal correspondence with external press data at the municipality-day level. The fine spatiotemporal resolution enables unprecedented mapping of riot evolution — Figure 4 illustrates the hour-by-hour spatial dynamics across the Paris agglomeration. Then, leveraging this resolution of the data, I group posts into discrete events and reveal that riots manifest as highly localized,

short-lived bursts. Specifically, I show that 80% of rioting events last less than 50 minutes and remain confined to areas under 0.3 km².

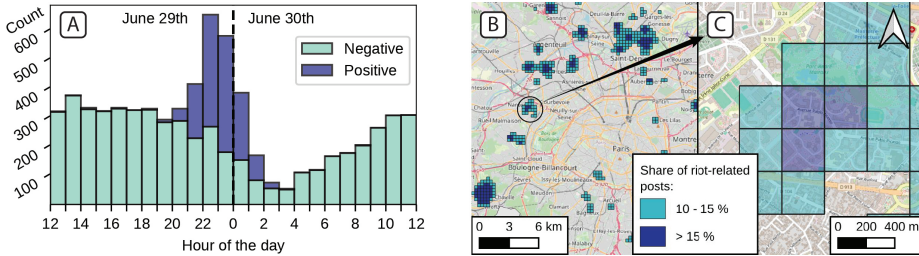


Figure 4 A: Temporal distribution of riot-related and non riot-related posts in the Paris agglomeration between June 29th and June 30th. B: Spatial distribution of riot-related posts in the Paris agglomeration between June 29th and June 30th. C: Zoom on a neighborhood in Rueil-Malmaison.

These findings point to several promising directions for future research. First, the analysis could move from describing 'what patterns exist' to explaining 'how patterns emerge' by examining the housing mobility processes — individual moves into, out of, and within neighborhoods — that generate segregation. Second, future research could investigate why segregation aligns more strongly with urban barriers in the Netherlands and Germany than in other European countries, helping to identify whether specific urban planning practices drive this pattern. Third, while the analysis characterizes rioting events as bursty and localized, examining spatiotemporal dependencies between events would reveal whether riots spread through contagion-like processes or remain independent, informing containment strategies. The dissertation yields several implications for policymakers. First, tracking segregation trends enables targeted intervention where patterns are emerging. The Dutch case illustrates this: between 2015 and 2020, suburban Randstad towns and rural municipalities experienced increases in separation and scale, converging toward levels in major cities. This suggests prioritizing secondary towns and rural areas through mixed-tenure developments that integrate social and private housing, as addressing segregation at earlier stages may be more feasible than reversing entrenched patterns. Second, since segregation does not systematically align with urban barriers in European contexts, infrastructure modifications such as putting railways or motorways underground should not be justified based on expected segregation-reduction effects. Third, this work underscores the importance of fine-resolution demographic data for analyzing urban social phenomena. National statistical agencies and EU institutions (particularly Eurostat) should continue making fine-resolution demographic data openly accessible while employing anonymization techniques to protect privacy, and establish EU-wide frameworks to enable comparative research across countries. Finally, despite its research potential for mapping riots, social media analysis presents a slippery slope if applied to participant identification or predictive policing. While this thesis demonstrates the value of social media for understanding riot dynamics, researchers and practitioners

should remain mindful of potential misuse for surveillance or preemptive intervention. Ultimately, understanding the spatial structure of social tensions — whether chronic patterns like segregation or acute events like riots — is essential for developing cities that facilitate integration rather than division.

Samenvatting

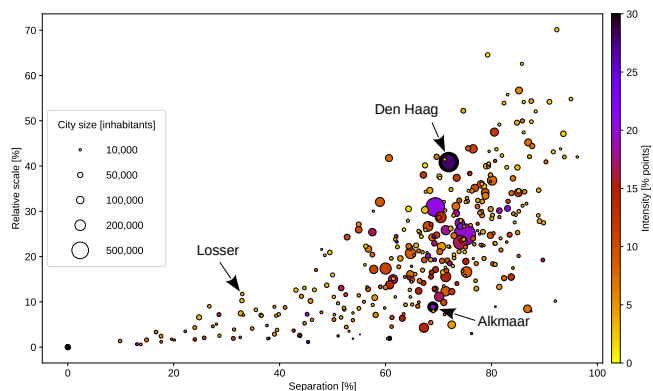
Steden zijn ruimtelijke concentraties van de bevolking die verbindingen tussen individuen faciliteren. Door mensen samen te brengen, ontstaan er omgevingen waarin sociale groepen met verschillende achtergronden en middelen — bijvoorbeeld opleiding, etniciteit of klasse — met elkaar in contact komen en ervaringen uitwisselen. Wanneer deze verschillen ontaarden in conflicten, ontstaan er sociale spanningen die bijdragen aan fenomenen zoals woonsegregatie of onlusten. Woonsegregatie is een latente, chronische uiting hiervan, waarbij sociale scheidslijnen verankers raken in ruimtelijke patronen die zich gedurende meerdere jaren en over meerdere stedelijke gebieden ontwikkelen. Rellen daarentegen uiten zich expliciet, acuut en als geconcentreerde episodes van geweldadige onrust op specifieke locaties en binnen kort tijdsbestek. Deze ruimtelijk-temporele fenomenen hebben ingrijpende gevolgen: ze vergroten ongelijkheid op het gebied van onderwijs, gezondheidszorg en werkgelegenheid, en ondermijnen zowel de sociale cohesie tussen groepen als het vertrouwen in instituties. Dergelijke patronen duiden op tekortkomingen van het stedelijk bestuur en de planning om interactie en integratie tussen diverse bevolkingsgroepen te bevorderen, waardoor zij hoog op de gemeentelijke agenda's staan. Deze aandacht onderstreept de noodzaak voor een analyse van de wijze waarop woonsegregatie en onlusten zich ruimtelijk voltrekken.

Er blijft een analytische kloof bestaan in het begrijpen van de ruimtelijke structuur van deze fenomenen — de specifieke geografische patronen, configuraties en ruimtelijke relaties die bepalen hoe segregatie en onlusten zich feitelijk voltrekken in stedelijke landschappen. Ondanks de beschikbaarheid van fijnmazige demografische data en geolocatie uit sociale media, ontbreekt het onderzoekers aan zowel de methoden om deze datasets te benutten als aan de empirische toepassingen die nodig zijn om het ruimtelijk inzicht in sociale spanningen te vergroten. De overkoepelende doelstelling is daarom om de ruimtelijke structuur van sociale spanningen in steden bloot te leggen door analytische methoden te ontwikkelen die deze fenomenen kunnen vastleggen op de schaal waarop zij zich daadwerkelijk manifesteren. Deze doelstelling wordt aangepakt via twee onderzoeksstromen: het onderzoeken van de wijze waarop patronen van woonsegregatie zich organiseren in stedelijke landschappen en hoe onlusten zich ruimtelijk-temporeel voltrekken. De eerste stroom omvat de hoofdstukken 2 tot en met 4, waarin wordt onderzocht hoe patronen van woonsegregatie ruimtelijk zijn georganiseerd. Deze hoofdstukken bouwen voort van methodeontwikkeling met eerste empirische bevindingen in de Nederlandse context, naar een temporele analyse van de evolutie van segregatie binnen Nederlandse gemeenten, tot een grootschalig pan-Europees onderzoek naar de ruimtelijke samenhang tussen segregatiepatronen en stedelijke barrières zoals spoorwegen, wegen en rivieren. De tweede stroom bestaat uit hoofdstuk 5, waarin de gedetailleerde ruimtelijk-temporele dynamiek van onlusten in stedelijke wijken wordt

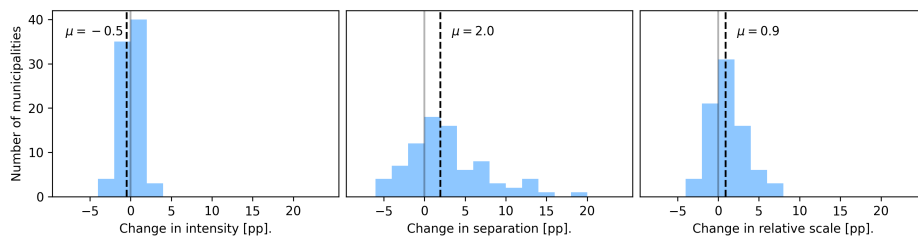
vastgelegd met behulp van computer vision-technieken om visuele data van sociale media te analyseren, met de Franse onlusten rond Nahel Merzouk (juli 2023) als casestudy.

Hoofdstuk 2 onderzoekt de ruimtelijke structuur van woonsegregatie en beschrijft de mate waarin deze varieert tussen verschillende stedelijke contexten. Hiervoor hanteer ik een regionalisatiemethode die demografisch homogene regio's afbakt, en ontwikkel ik een raamwerk om de ruimtelijke structuur van woonsegregatie te karakteriseren aan de hand van drie dimensies: intensiteit — de mate van oververtegenwoordiging van een groep in gesegregeerde regio's ten opzichte van het stedelijk gemiddelde —, separatie — het aandeel van de groep dat in deze gesegregeerde regio's woont — en schaal — de omvang van deze regio's wat betreft bevolkingsaantal. Deze methode is toegepast op Nederlandse 6-positie postcodedata (ruimtelijke eenheden van ongeveer 50 inwoners) om de segregatie van de bevolking met een niet-westerse migratieachtergrond in alle Nederlandse gemeenten te bestuderen. De analyse brengt een aanzienlijke diversiteit aan ruimtelijke patronen tussen Nederlandse gemeenten aan het licht. Er komen zes van de acht theoretisch mogelijke segregatietypen voor, waarbij gemeenten vrijwel de volledige conceptuele ruimte beslaan in plaats van één enkel typisch patroon te vertonen. Figuur 5 illustreert deze driedimensionale karakterisering en laat zien hoe Nederlandse gemeenten variëren in intensiteit, separatie en schaal. Bij het onderzoeken van de relatie tussen deze indicatoren en andere stedelijke variabelen vertonen zowel intensiteit als separatie een positieve samenhang met de omvang van de gemeente, de sociale mix en inkomensongelijkheid. Dit suggereert dat deze verbanden voortkomen uit een gemeenschappelijk onderliggend fenomeen in plaats van onafhankelijke effecten. De absolute schaal van segregatie correleert lineair met het inwoneraantal van de gemeente: grotere steden vertonen proportioneel grotere gesegregeerde regio's in plaats van een groter aantal regio's. Dit wijst erop dat segregatiepatronen in grote gemeenten opgeschaalde versies zijn van de patronen in kleinere gemeenten.

Hoofdstuk 3 bouwt voort op de analyse uit hoofdstuk 2 om ruimtelijke patronen door de tijd heen te volgen. Het onderzoekt de temporele evolutie van segregatiepatronen in Nederlandse gemeenten tussen 2015 en 2020 en hun relatie tot demografische veranderingen. De analyse wijst op een convergentie van segregatiepatronen tussen Nederlandse gemeenten gedurende deze periode. Figuur 6 toont de ruimtelijke verdeling van de residentiële mix en de drie dimensies van woonsegregatie in 2015, wat een referentiekader biedt voor het begrijpen van deze temporele dynamiek. Gemeenten met een initieel lage intensiteit en separatie vertoonden een stijging in deze indicatoren, terwijl gemeenten met historisch hoge waarden — in het bijzonder de grote Randstadsteden — een daling lieten zien, wat resulteerde in een afgenomen variantie op nationaal niveau. De toename van het aandeel van de bevolking met een niet-westerse migratieachtergrond correleert positief met een toename in separatie en schaal, maar niet met veranderingen in intensiteit. Dit suggereert dat een groei van de bevolking met een niet-westerse migratieachtergrond leidt tot een ruimtelijke uitbreiding van gesegregeerde regio's of het ontstaan van nieuwe regio's, in plaats van een toegenomen homogeniteit binnen bestaande gebieden.



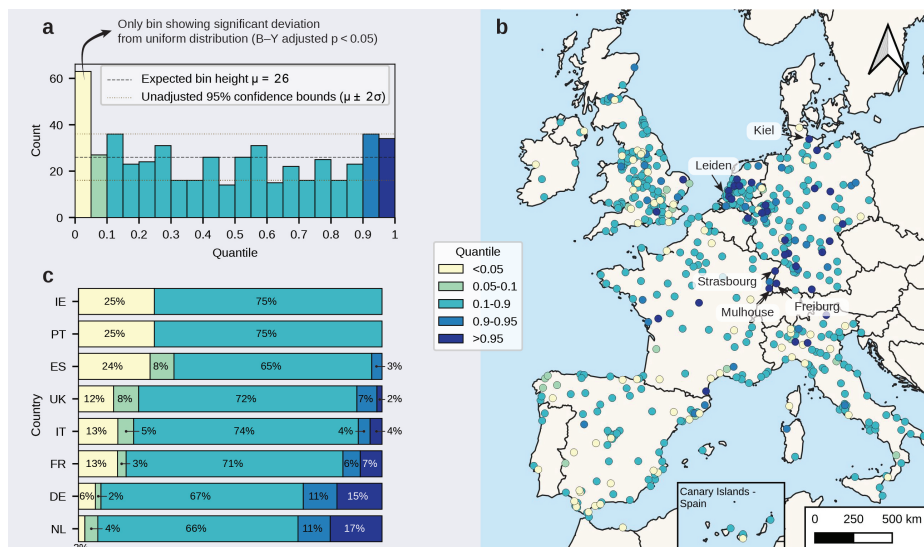
Figuur 5 Relatie tussen segregatie-indicatoren in Nederlandse gemeenten. De separatie en de relatieve schaal worden respectievelijk op de x- en y-as weergegeven, terwijl de kleur van een punt de intensiteit weergeeft. De grootte van de punten is proportioneel aan het inwoneraantal van de gemeente.



Figuur 6 Statistische verdeling van de verandering in intensiteit, separatie en relatieve schaal in procentpunten (pp).

Hoofdstuk 4 onderzoekt of woonsegregatie samenvalt met stedelijke fragmentatie — de opdeling van de stedelijke ruimte door stedelijke barrières zoals spoorwegen, snelwegen en waterwegen — in contexten zonder historisch segregatiebeleid. Hoewel een dergelijke samenhang is gedocumenteerd in Amerikaanse en Zuid-Afrikaanse steden met een verleden van institutionele segregatie, suggereert theorieën dat dit ook op natuurlijke wijze kan voortkomen uit woonvoorkeuren. Het toetsen van deze hypothese in de Europese context vult een empirische leemte op met implicaties voor zowel stedelijke theorievorming als beleid. Dit ontwikkelt hoofdstuk een Monte Carlo-methode om de statistische significantie te toetsen van de ruimtelijke overeenkomst tussen demografische regio's en stedelijke barrières, waarbij de focus ligt op de segregatie van migranten van buiten de EU. Voor elk van de 520 steden in acht West-Europese landen genereert de methode 200 synthetische fragmentatiepatronen die overeenkomen met de daadwerkelijke grootteverdeling van de fragmenten. Ik meet de uitlijning tussen demografische regio's en stedelijke barrières met een clustergeïjkendings-metrick en beoordeel de significantie ervan door de waarde te vergelijken met de uitlijning van demografische regio's op

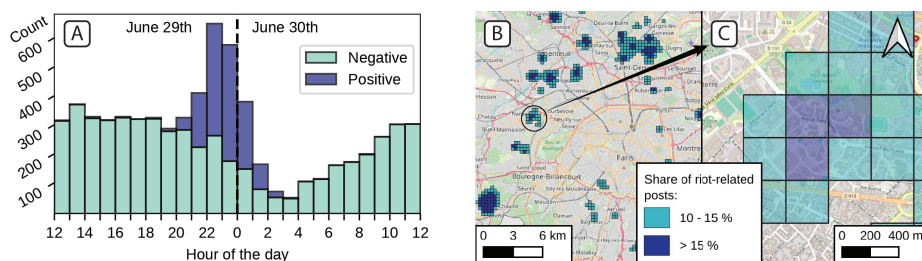
de 200 synthetische partities. De bevindingen laten zien dat er geen sprake is van een systematische uitlijning tussen woonsegregatie en stedelijke fragmentatie in Europese steden — in tegenstelling tot de verwachtingen op basis van de Amerikaanse en Zuid-Afrikaanse context. Veel steden vertonen zelfs minder uitlijning dan op basis van toeval mag worden verwacht. Stedelijke decentralisatie blijkt een significante voorspeller van uitlijning, terwijl de intensiteit van segregatie dat niet is. Er treden belangrijke regionale variaties op: Nederland vertoont een significant sterkere uitlijning dan andere landen, zelfs na correctie voor stedelijke decentralisatie. Deze resultaten suggereren dat stedelijke barrières in Europese contexten zonder expliciet segregatiebeleid over het algemeen niet fungeren als sociale grenzen, hoewel landspecifieke patronen van stedelijke ontwikkeling deze relatie kunnen beïnvloeden. Figuur 7 toont de statistische en geografische verdeling van kwantielrang, wat zowel het algemene gebrek aan systematische samenhang illustreert als de concentratie van steden met een significante uitlijning in Duitsland en Nederland.



Figuur 7 Statistische en geografische patronen van uitlijning tussen woonsegregatie en stedelijke fragmentatie in Europese steden. (a) Verdeling van kwantielrang over alle steden in de dataset. (b) Geografische verdeling van kwantielrang over de steden in de dataset. (c) Verdeling van kwantielrang per land.

Hoofdstuk 5 verschuift de focus naar het in kaart brengen van onlusten — de tweede manifestatie van sociale spanningen die in dit proefschrift wordt onderzocht. In tegenstelling tot de latente, chronische patronen van woonsegregatie, vormen onlusten expliciete, acute gebeurtenissen die in de openbaarheid uitbarsten. Om de dynamiek hiervan te begrijpen, zijn gedetailleerde ruimtelijk-temporele gegevens noodzakelijk. Het vastleggen van dit detailniveau bleek voorheen vaak problematisch, maar sociale

media — waar gebruikers gebeurtenissen in realtime documenteren met nauwkeurige Geolocatie — bieden nieuwe mogelijkheden om het verloop van onlusten te traceren. In dit hoofdstuk onderzoek ik de ruimtelijk-temporele verdeling van onlusten aan de hand van de onlusten rond Nahel Merzouk (2023) als casestudy. Hierbij maak ik gebruik van data van de sociale media-app Snapchat — 107.000 berichten verzameld in Parijs, Marseille, Lyon, Toulouse, Lille, Bordeaux en Grenoble. Ik identificeer aan onlusten gerelateerde inhoud in visuele berichten door een computer vision-model te finetunen op een geannoteerde subset van de data, waarmee een precisie van 77% en een recall van 68% op de testset werd behaald. Toepassing van het model op de volledige dataset laat een sterke ruimtelijk-temporele overeenkomst zien met externe persgegevens op gemeente-dag-niveau. De hoge ruimtelijk-temporele resolutie maakt een ongekennde kartering van de evolutie van onlusten mogelijk — Figuur 8 illustreert de uur-tot-uur ruimtelijke dynamiek in de agglomeratie Parijs. Door gebruik te maken van deze resolutie groepeer ik berichten in afzonderlijke gebeurtenissen, waaruit blijkt dat onlusten zich manifesteren als zeer gelokaliseerde, kortstondige uitbarstingen. Concreet toon ik aan dat 80% van de onlusten minder dan 50 minuten duurt en beperkt blijft tot gebieden kleiner dan 0,3 km².



Figuur 8 A: Temporele verdeling van aan onlusten gerelateerde en niet-gerelateerde berichten in de agglomeratie Parijs tussen 29 en 30 juni. B: Ruimtelijke verdeling van aan onlusten gerelateerde berichten in de agglomeratie Parijs tussen 29 en 30 juni. C: Zoom op een buurt in Rueil-Malmaison.

Deze bevindingen wijzen op verschillende veelbelovende richtingen voor toekomstig onderzoek. Ten eerste zou de analyse kunnen verschuiven van het beschrijven van 'welke patronen er bestaan' naar het verklaren van 'hoe patronen ontstaan', door de processen van woonmobiliteit — individuele verhuizingen naar, uit en binnen wijken — te onderzoeken die segregatie genereren. Ten tweede zou toekomstig onderzoek kunnen uitwijzen waarom segregatie in Nederland en Duitsland sterker samenvalt met stedelijke barrières dan in andere Europese landen, wat kan helpen identificeren of specifieke praktijken in de ruimtelijke ordening dit patroon sturen. Ten derde, hoewel de analyse onlusten karakteriseert als kortstondige en gelokaliseerde uitbarstingen, zou onderzoek naar ruimtelijk-temporele afhankelijkheden tussen gebeurtenissen kunnen onthullen of onlusten zich verspreiden via besmettingsachtige processen of onafhankelijk van elkaar blijven; dit kan leiden tot beter onderbouwde inperkingsstrategieën. Het proefschrift levert verschillende implicaties op voor beleidsmakers. Ten eerste maakt

het monitoren van segregatietrends gerichte interventies mogelijk op plaatsen waar patronen zich beginnen te vormen. De Nederlandse casus illustreert dit: tussen 2015 en 2020 vertoonden randgemeenten van de Randstadsteden en landelijke gemeenten een toename in separatie en schaal, waarmee zij convergeerden naar de niveaus van de grote steden. Dit suggereert dat de prioriteit moet liggen bij middelgrote steden en landelijke gebieden door middel van gemengde woonvormen die sociale woningbouw en de private sector integreren; het aanpakken van segregatie in een vroeg stadium is immers vaak haalbaarder dan het terugdraaien van diepgewortelde patronen. Ten tweede: aangezien segregatie in Europese contexten niet systematisch samenvalt met stedelijke barrières, zouden infrastructurele aanpassingen — zoals het ondergronds brengen van spoorwegen of snelwegen — niet moeten worden gerechtvaardigd op basis van verwachte effecten op de vermindering van segregatie. Ten derde onderstreept dit werk het belang van fijnmazige demografische gegevens voor de analyse van stedelijke sociale fenomenen. Nationale bureaus voor de statistiek en EU-instellingen (met name Eurostat) moeten doorgaan met het openbaar toegankelijk maken van deze data, onder toepassing van anonimiseringstechnieken ter bescherming van de privacy, en EU-brede kaders vaststellen om vergelijkend onderzoek tussen landen mogelijk te maken. Ten slotte vormt de analyse van sociale media, ondanks het onderzoekspotentieel voor het in kaart brengen van onlusten, een riskant pad wanneer deze wordt toegepast voor de identificatie van participanten of predictive policing. Hoewel dit proefschrift de waarde van sociale media aantoonst voor het begrijpen van de dynamiek van onlusten, moeten onderzoekers en praktijkbeoefenaars waakzaam blijven voor mogelijk misbruik ten behoeve van Surveillance of preventieve interventie. Uiteindelijk is inzicht in de ruimtelijke structuur van sociale spanningen — of het nu gaat om chronische patronen zoals segregatie of acute gebeurtenissen zoals onlusten — essentieel voor de ontwikkeling van steden die integratie bevorderen in plaats van verdeeldheid.

Chapter 1

Introduction

1.1 Background

Urban environments bring together social groups with different backgrounds and socioeconomic means — whether defined by education, ethnicity, class, or other characteristics — creating opportunities for enrichment (**Florida, 2002; Jacobs, 1961**). When these differences coincide with inequalities, they generate tensions over competed spatial resources such as housing, public infrastructure, and access to amenities (**Alesina & La Ferrara, 2002; Putnam, 2000; W. Wilson, 1987**). These tensions manifest spatially as residential segregation — defined here as the uneven distribution of social groups in space (**Reardon & O’Sullivan, 2004**) — and as riots — here understood as episodes of collective civil unrest involving violence, vandalism, looting, and clashes with authorities (**Lachman, 1996; Wilkinson, 2009**).

1.1.1 Social tensions in cities

Urban environments bring together social groups with different backgrounds and socioeconomic means — whether defined by education, ethnicity, class, or other characteristics — creating opportunities for enrichment (**Florida, 2002; Jacobs, 1961**). When these differences coincide with inequalities, they generate tensions over competed spatial resources such as housing, public infrastructure, and access to amenities (**Alesina & La Ferrara, 2002; Putnam, 2000; W. Wilson, 1987**). These tensions manifest spatially as residential segregation — defined here as the uneven distribution of social groups in space (**Reardon & O’Sullivan, 2004**) — and as riots — here understood as episodes of collective civil unrest involving violence, vandalism, looting, and clashes with authorities (**Lachman, 1996; Wilkinson, 2009**).

Residential segregation and riots share important characteristics. They both represent social issues that manifest spatially in urban environments, emerging from the aggregation of individual behaviors. At the same time, they differ in the way they manifest of social tensions (**Castells, 1977**). Residential segregation operates as a latent expression, embedding social divisions within spatial patterns without direct confrontation between groups (**Massey & Denton, 1993; Park et al., 1925**). Riots represent explicit manifestations, bringing social conflicts directly into the open through violent confrontations (**Kawalerowicz & Biggs, 2015; Tilly, 2003; Waddington, 2012b**). Beyond this latent-explicit distinction, they differ fundamentally

in their spatiotemporal characteristics: segregation unfolds as large-scale, chronic phenomena evolving over years or decades across broad urban areas (**Logan & Stults, 2011; Massey & Denton, 1993**), while riots emerge as small-scale, acute episodes concentrated in specific locations and compressed timeframes (**Waddington, 2012c**).

1.1.2 Residential segregation

Residential segregation — defined as the uneven distribution of social groups in space (**Reardon & O’Sullivan, 2004**) — represents a latent manifestation of social tensions (**Massey & Denton, 1993; Park et al., 1925**). Segregation can occur along various social characteristics such as ethnicity, income, age, education level, migration background (**Andersson & Malmberg, 2015; Ellis et al., 2018; Marcińczak et al., 2021; Milias & Psyllidis, 2022; Scarpa, 2015**). Not all spatial sorting of social groups is inherently problematic. When spatial clustering reflects unconstrained residential preferences, where individuals freely choose to live near others who share their background or lifestyle, it can be understood as a natural outcome of residential choice (**Tiebout, 1956**). However, when segregation reflects and reinforces inequalities in access to resources and opportunities, it embeds social divisions within spatial patterns with substantial consequences for the city’s population (**Ludwig et al., 2012; Nieuwenhuis et al., 2020; Owens, 2018; Tóth et al., 2021**).

Extensive research has documented how residential segregation reinforces existing inequality between groups along several aspect of life including educational achievement, health outcomes, and employment opportunities (**Ludwig et al., 2012; Nieuwenhuis et al., 2020; Owens, 2018; Tóth et al., 2021**), while also contributing to social fracture (**Farber et al., 2014; Laurence et al., 2019; Semyonov & Glikman, 2008; van der Laan Bouma-Doff, 2007**). Inequalities widen through neighborhood effects — when households with similar challenges are geographically clustered, individual hardships compound into neighborhood-level effects that create additional barriers beyond what any single household would face in isolation (**Chetty et al., 2016; Coleman, 1988; Sampson et al., 2002**). This means that residents face not only their own household’s disadvantages but also spillover effects due to concentrated disadvantages in the entire neighborhood. For instance, research demonstrates that students from disadvantaged families in segregated areas achieve lower standardized test scores than would be predicted by their family income alone, with neighborhood concentration of disadvantage amplifying individual barriers to educational success (**Owens, 2018**). These compounding inequalities, which, combined with limited interactions between segregated groups, fracture broader social cohesion (**Farber et al., 2014; Laurence et al., 2019; Semyonov & Glikman, 2008; van der Laan Bouma-Doff, 2007**). The magnitude of these consequences has made residential segregation a policy priority for national and local authorities worldwide, from the United Nations’ New Urban Agenda to municipal integration strategies in major cities (**Gemeente Amsterdam, 2017; Mayor of London, 2021; the United Nations, 2017; Ville de Paris, 2021**).

The severity of these impacts has driven research efforts to understand what

generates segregation patterns. This requires distinguishing between two analytical levels: the structural conditions that create the potential for segregation, and the mechanisms through which these conditions translate into observable spatial patterns. Structural conditions set the constraints and opportunities that determine where different groups can realistically live, while mechanisms describe the processes through which residents, landlords, and institutions respond to those constraints, ultimately producing observable segregation patterns. Several structural conditions shape this environment. Inequality in income or wealth creates systematic differences in the residential choices available to different groups (**Haandrikman et al., 2023; Reardon & Bischoff, 2011; San Millán et al., 2025; Watson, 2009**). Discrimination operates alongside inequality, as prejudicial attitudes and behaviors constrain where certain groups feel welcome or are able to settle (**Denton, 2006; MacDonald et al., 2018**). Finally, the spatial design legacy of cities, where different types of housing have been constructed and in what spatial concentrations, creates a physical opportunity structure that shapes where different groups can realistically live (**Arbaci, 2007; Bickford & Massey, 1991; Skifter Andersen et al., 2016**).

Three mechanisms concentrate similar groups in the same areas, each activated by the structural conditions described above. First, in some contexts, past policies and urban planning have intentionally established segregation through measures such as redlining in the US (**Ananat, 2011; Chyn et al., 2022**) or apartheid in South Africa (**Ballard & Hamann, 2021; Maharaj, 2020**), creating enduring spatial patterns that persist despite the formal termination of these policies. Second, segregation arises through self-selection, where groups' varying preferences and constraints shape their residential choices (**Banzhaf & Walsh, 2008; Brueckner et al., 1999; Tiebout, 1956**). Neighborhoods with desirable amenities such as good schools, parks, or transport connections command higher housing prices (**Black, 1999; Conway et al., 2010; Debrezion et al., 2007**), making them accessible primarily to higher-income groups, while areas with fewer amenities remain more affordable to lower-income groups. Third, homophily — the tendency for individuals to prefer living near others who share their cultural background, language, or social characteristics — further concentrates similar groups in the same areas (**Clark, 1991; Fossett, 2006; Massey & Denton, 1988; Schelling, 1971**); for instance, groups that have experienced hostility or exclusion in certain neighborhoods may develop preferences that steer them toward areas where they feel safer or more welcome. Segregation results from the interaction between these structural conditions and the mechanisms they activate.

Understanding how these various mechanisms result in segregation patterns requires systematic approaches for measuring the spatial structures of segregation. A substantial body of literature develops the methodological foundations for measuring and analyzing segregation patterns. Studies typically rely on census data or administrative records that provide population characteristics disaggregated by geographic units such as census tracts, postal codes, or administrative neighborhoods (**Massey & Denton, 1988; Reardon & O'Sullivan, 2004**). The measurement of segregation phenomena

has traditionally focused on quantifying the intensity of spatial separation between different groups. Conventional approaches employ grid-based metrics such as the dissimilarity index (**Duncan & Duncan, 1955**) and spatial information theory (**Theil & Finizza, 1971**) indices that reduce complex spatial patterns to single numerical indicators. These traditional measures treat geographic units as discrete entities, with some recent studies incorporating spatial dependency through techniques like spatial moving averages to account for relationships between neighboring areas (**Andersson et al., 2018; Clark et al., 2015; Costa & Valk, 2018; Lan et al., 2020; Marcińczak et al., 2021; Östh et al., 2015**). Research has also examined the topology and spatial structure of segregation patterns (**Ballard & Hamann, 2021; Kramer, 2017; Roberto & Korver-Glenn, 2021**). These efforts usually employ qualitative approaches, such as visual inspection of maps and descriptive accounts of spatial configurations.

1.1.3 Riots

As an explicit manifestation of social tensions, riots are episodes of violent social unrest, characterized by vandalism, looting, and clashes with law enforcement (**Lachman, 1996; Wilkinson, 2009**). These events unfold as spontaneous, episodic phenomena that emerge rapidly across urban areas (**Waddington, 2012c**). Riots are not without political significance. Historically, they have functioned as vehicles for raising attention to grievances and catalyzing social and political change, representing a form of contentious politics through which marginalized groups assert claims that conventional channels fail to address (**Kawalerowicz & Biggs, 2015; Tilly, 2003; Waddington, 2012b**). However, their economic and social costs are substantial and readily quantifiable, encompassing property damage, business disruption, and infrastructure destruction that can reach millions in damages within single episodes (**Cicchelli et al., 2007; Home Office, 2011; Sénat, 2024**). Furthermore, their effects extend beyond their immediate disruption, affecting cohesion between different groups and trust between citizens and institutions (**Briggs, 2012; Nägel & Lutter, 2023**). These events also present acute challenges for crisis management and urban governance, testing the capacity of local authorities to respond effectively while avoiding escalation. Recent examples demonstrate the persistent occurrence and the violence of such events across countries, including the UK riots of 2011, and the French riots of 2005 and 2023 (**Kawalerowicz & Biggs, 2015; Murray, 2006; Sénat, 2024; Waddington, 2012b**).

These substantial economic and social costs of riots have prompted research to investigate their root causes. Inequality, poverty, and lack of prospects which generate accumulated grievances, have been mentioned by some researchers as possible root causes in relation to the Paris 2005, London 2011, and Stockholm 2013 riots (**Body-Gendrot, 2013; Kawalerowicz & Biggs, 2015; Lapeyronnie, 2006**). There is however lack of empirical research for underpinning said relations. Relative deprivation in amenities and services further compounds these conditions, as populations in disadvantaged neighborhoods face not only individual hardship but also inferior access to the resources that shape life outcomes (**Body-Gendrot, 2016; Newburn, 2016**;

Schierup et al., 2014; Till, 2013). This is presumably exacerbated by distrust toward institutions, particularly law enforcement, creating persistent tensions between authorities and marginalized groups (**Gurr, 2015; Holdo & Bengtsson, 2020; Wilkinson, 2009).**

While structural conditions define the landscape of vulnerability within which riots become possible, they do not themselves cause riots. Understanding how these conditions translate into actual unrest requires examining the mechanisms through which latent tensions become open conflict, the processes through which individuals, in specific circumstances, move from grievance to action. A symbolic triggering event, typically a specific incident perceived as emblematic of broader injustices, crystallizes accumulated grievances and ignites unrest, such as the accidental death of Zyed Benna and Bouna Traore in the North-East suburb of Paris in 2005, the shooting of Mark Duggan in London in 2011, and the shooting of Lenine Relvas-Martins in Stockholm (**Holdo & Bengtsson, 2020; Koff & Duprez, 2009; Waddington, 2012a).** Once a triggering event ignites initial unrest, individual participation is shaped by interdependent decision-making processes. **Granovetter, 1978's** threshold model of collective behavior captures how participation cascades as crowd size grows: as more individuals join, the perceived cost of participation declines since the probability of individual identification and punishment decreases, progressively lowering the threshold at which joining becomes rational for others. Simultaneously, what **Jasper, 1998** describes as reactive emotions; transitory, context-specific feelings formed and reinforced within collective action itself; amplify and validate individual grievances, transforming private frustration into shared moral outrage and further motivating participation. Once ignited, riots may escalate and spread through contagion-like processes, propagating across neighborhoods and cities (**Bohstedt, 2013; Bonnasse-Gahot et al., 2018; Granovetter, 1978).**

To study both the structural causes of riots and their dynamics from triggering events to widespread unrest, researchers require methodological approaches capable of capturing these phenomena across space and time. A substantial body of literature has developed around these two analytical objectives, examining first how underlying social conditions relate to riot occurrence (**Baudains, Braithwaite & Johnson, 2013; Malmberg et al., 2013),** and second how unrest spreads and evolves once triggered (**Berestycki et al., 2015; Bonnasse-Gahot et al., 2018; Pires & Crooks, 2017).** Research investigating the structural causes of riots focuses on mapping relationships between neighborhood characteristics and unrest patterns. Studies examine associations between riot locations and factors such as socioeconomic deprivation, demographic composition, and histories of previous incidents, seeking to understand why certain areas experience higher riot intensity than others (**Baudains, Braithwaite & Johnson, 2013; Malmberg et al., 2013).** This approach treats riots as spatially embedded phenomena, analyzing how the structural conditions manifest in observable geographic patterns of civil unrest. Complementing this analysis of underlying causes, research examines how riots spread and evolve from initial triggering events to widespread unrest. Studies employ epidemiological and diffusion models to understand how civil unrest propagates across urban areas over time (**Baudains, Johnson & Braithwaite, 2013; Berestycki et al.,**

2015; Bonnasse-Gahot et al., 2018; Davies et al., 2013; Pires & Crooks, 2017).

1.2 Emerging fine-resolution data sources for urban analysis

Recent advances in data sources and processing expand researchers' potential to study residential segregation and riots. Fine-resolution demographic data enable the identification of the spatial structure of residential segregation which can be used to compare its alignment to urban planning context. Geolocated social media data reporting urban events have potential for mapping riots at much finer scale than traditional data sources as press reports.

1.2.1 Fine-resolution demographic data for mapping segregation

The increasing availability of fine-resolution demographic data opens new analytical possibilities for research on residential segregation (**Centraal Bureau voor de Statistiek, 2021; van Leeuwen, 2020**). By fine resolution, I refer to demographic data available at increasingly smaller spatial scales - for example, spatial units containing several thousands people (such as US census tracts), several hundreds people (such as US census blocks available since 1990), 50 people (such as Dutch postcodes available since 2015), or even household-level data (such as Dutch or Swedish microdata) (**Centraal Bureau voor de Statistiek, 2021; United States Census Bureau, 2025; van Leeuwen, 2020**). Traditional coarse spatial resolution has been associated with challenges such as the Modifiable Areal Unit Problem (MAUP), where research findings such as segregation indices depend on the both the resolution and the spatial layout of the input data (**Openshaw, 1981; Petrović et al., 2018**). Coarser resolutions tend to aggregate heterogeneous populations within larger spatial units, potentially obscuring important small-scale demographic variations. The zoning effect occurs when the way spatial units are delineated at a given scale alters statistical results, as different boundary definitions can group populations differently and change observed patterns. Finer-resolution data reduce these effects.

Spatial granularity strengthens development of approaches that capture the relation between segregation and urban development. An example is regionalization, the process of grouping spatial units into continuous demographically-homogeneous regions, which can be particularly valuable for mapping spatial structure of segregation (**Duque et al., 2007**). It can be used to study how demographic regions align spatially with specific urban planning context, and can inform targeted interventions (**G. Galster et al., 2006**). With coarse resolution data, regionalization approaches work with already aggregated, heterogeneous units (such as census tracts that contain both affluent and low-income areas), thereby compromising the detection of true underlying demographic transitions. As resolution becomes finer, spatial units are more likely to represent spatially uniform demographic contexts, which facilitates the identification of coherent demographic regions that reflect actual population patterns on the ground.

1.2.2 Social media data for mapping rioting events

Social media platforms have created an entirely new category of data source that can capture urban social phenomena with fine spatial and temporal granularity (**Bazzaz Abkenar et al., 2021; George et al., 2021; Hernández-Fuentes & Monnier, 2022; Panagiotopoulos et al., 2016; Prabhu et al., 2021; Procter et al., 2013**). Users document events with precise geolocation information and timestamp data, creating a distributed network of observations that provide comprehensive coverage of social events at fine spatiotemporal resolution (**Bazzaz Abkenar et al., 2021; George et al., 2021; Hernández-Fuentes & Monnier, 2022; Panagiotopoulos et al., 2016**). This detailed documentation capability is particularly valuable for understanding rapidly evolving phenomena such as riots, where traditional data sources often provide only sparse and coarse spatial and temporal information.

Social media platforms vary in their analytical suitability for mapping riots based on two key aspects: functional use and data structure. Regarding functional use, platforms serve distinct purposes: debate-based platforms like Twitter/X facilitate discussion, content-creation-based platforms like Instagram emphasize visual content sharing, while live-reporting-based platforms like Snapchat focus on real-time event documentation (**Alhabash & Ma, 2017**). For riot mapping, live reporting platforms offer greater analytical value because they capture events with precise spatiotemporal coordinates that correspond to actual incident locations and timing, whereas discussion-oriented platforms may document commentary that occurs in different locations and timeframes than the events themselves (**Barrie & Frey, 2021**). Data structure represents the second aspects of a platform's suitability, as platforms differ substantially in how readily their content can be processed at scale. Platforms with rich structured metadata, such as hashtags and geolocation tags, facilitate automated processing and large-scale analysis, while platforms emphasizing textual or visual content require more intensive interpretation processes that limit analytical scalability. These considerations explain why large-scale empirical studies of mass social events have predominantly focused on Twitter/X, where relevant posts can be systematically retrieved through hashtag analysis (**Alsaedi et al., 2017; Nicoletti et al., 2021; S. L. Wilson, 2017**). Insofar, these approaches have enabled researchers to achieve spatial resolution at the municipal level and temporal resolution at the daily scale, matching the granularity of traditional administrative data sources.

1.3 Problem statement

The substantial social and economic consequences of residential segregation and urban riots — from reinforced inequalities to material and human cost — underscore the critical importance of understanding these phenomena for both urban theory and policymaking. Recent advances have produced fine-resolution demographic datasets and geolocated social media data that capture urban events. Yet, significant gaps persist in understanding these phenomena at the spatial scales where they actually manifest. Specifically, these gaps consist of methodological limitations — where analytical approaches fail to match the spatial sophistication of available data sources — resulting in substantive knowledge

gaps about how these urban social tensions actually operate in space. We lack empirical understanding of the spatial structure and extent of segregated areas within cities, how they relate to urban infrastructure and planning contexts, and how riots unfold and spread across neighborhoods with fine spatiotemporal precision. These methodological and empirical limitations constrain both theoretical advances in understanding urban social tensions and the development of spatially-targeted policy interventions. For instance, without understanding how transportation network contributes to segregation patterns, policymakers face costly urban planning decisions without evidence of their effectiveness in reducing segregation. Similarly, limited understanding of riot dynamics constrains the development of effective containment and deescalation strategies. The following two subsections examine how these limitations specifically constrain research on residential segregation and urban riots - phenomena that, while both rooted in social tensions, differ fundamentally in their spatial scale, temporal dynamics, and analytical requirements.

1.3.1 Limited spatial understanding of residential segregation

Most common studies on residential segregation treat geographic units as discrete entities, discarding the spatial structure that defines segregation as a geographic phenomenon (Massey & Denton, 1988; Reardon & Firebaugh, 2002; Theil & Finizza, 1971). While the literature considers spatial dependency through techniques such as spatial moving averages (Andersson et al., 2018; Brown & Chung, 2006; Clark et al., 2015; Costa & Valk, 2018; Lan et al., 2020; Marcińczak et al., 2021; Östh et al., 2015), a methodological gap remains in mapping the spatial structure of residential segregation. This perpetuates an empirical gap in our understanding of where segregated areas are located within cities, what their spatial extent encompasses, how they evolve, or how they relate to other spatial features in the city like physical barriers (e.g. highways, railways, or rivers). Addressing this gap would improve the understanding of the processes that generate and sustain segregation patterns. Such theoretical advances would, in turn, open new possibilities for policy development, allowing policymakers to design spatially-targeted interventions that address segregation at the neighborhood and district scales where patterns actually manifest.

1.3.2 Limited spatial analysis of urban riots

Research on urban riots suffers from significant spatiotemporal limitations, with riots typically studied at the municipality level, day-by-day (Aubert et al., 2023; Bonnasse-Gahot et al., 2018; Malmberg et al., 2013; Nicoletti et al., 2021), despite being local events with sub-daily dynamics. This mismatch between the spatiotemporal scale of riots and the coarse spatial resolution of analysis prevents researchers from accurately modeling the dynamics that characterize these events. For instance, a riot may consist of discrete incidents — acts of looting on specific streets, arson at particular intersections, or confrontation with police — with potential spatiotemporal dependencies, while traditional approaches aggregate these into a single municipality-level observation of riot intensity, thereby obscuring patterns like spatial clustering, temporal sequences,

and cascading effects that would inform the dynamics of riot propagation and escalation. The primary cause is reliance on data sources such as press reports offering data at municipality-level resolution, day-by-day (Aubert et al., 2023; Bonnasse-Gahot et al., 2018). Attempts debate-based social media platforms like Twitter for riot mapping (Alsaedi et al., 2017; Nicoletti et al., 2021) have only equalized this resolution because such platforms primarily capture post-hoc discussion rather than real-time observation, with users often commenting on events they witnessed earlier or did not witness at all (Barrie & Frey, 2021). In contrast, live-reporting-based social media platforms like Snapchat present untapped potential for capturing riots at much finer spatial and temporal resolution. However, researchers have been unable to fully exploit this resource due to the analytical challenges posed by unstructured visual content requiring human interpretation. This represents a critical methodological gap: developing scalable approaches to identify riot-related content from unstructured data sources. Resolving this methodological gap would enable fine-grain spatiotemporal analysis of riots.

1.4 Research objectives

The overarching objective is therefore to reveal the spatial structure of social tensions in cities — how residential segregation patterns organize across urban landscapes and how riots unfold spatiotemporally — by developing analytical methods capable of capturing these phenomena at the scales where they actually manifest. This objective is operationalized through four research questions:

RQ1: How to characterize the spatial structure of residential segregation and to what extent patterns differ across urban contexts?

RQ2: How does the spatial structure of residential segregation patterns evolve over time and relate to demographic change?

RQ3: To what extent the spatial structure of residential segregation follows urban barriers?

RQ4: How to identify riot-related content from visual social media data, and what does this reveal about the spatiotemporal nature of riots?

These research questions form two research streams that address the overarching objective through examining different manifestations of social tension: residential segregation (RQ1-3) and riots (RQ4). The first stream (RQ1-3) develops and applies spatial analysis methods for residential segregation, progressing from foundational method development with initial empirical findings in the Dutch context (RQ1), to temporal extension examining segregation evolution across Dutch municipalities (RQ2), to large-scale cross-European analysis focused specifically on spatial alignment between segregation and urban barriers (RQ3). The second stream (RQ4) employs computer vision techniques to analyze visual social media data, examining the spatiotemporal patterns of riots. Importantly, this thesis does not aim to determine when segregation levels become socially problematic or the extent to which riots are beneficial or harmful,

the objective is strictly to reveal how these phenomena organize spatially.

1.5 Contributions

This dissertation makes methodological contributions for understanding the spatial structure of social phenomena in urban environments, and generates empirical insights through large-scale applications of these methods.

1.5.1 Methodological contributions

The research advances spatial analysis methods by developing approaches that capture the spatial structure of urban social phenomena. Traditional methods often ignore or inadequately represent the spatial aspects of social processes. The methodological contributions enable social phenomena to reveal their natural spatial organization and dynamics.

- **Adaptive regionalization framework:** Development of data-driven approaches that construct meaningful spatial units, applicable across different urban contexts
- **Multi-dimensional segregation measurement:** Introduction of intensity, separation, and scale indicators that measure different aspects of the spatial structure of residential segregation, moving beyond traditional uni-dimensional metrics
- **Spatial alignment quantification:** Novel methodology for measuring the degree of correspondence between different spatial partitions using Monte Carlo simulations, enabling systematic testing of hypotheses about relationships between social and physical urban structures
- **Computer vision for social media analysis:** Pioneer application of computer vision techniques to identify riot-related content from unstructured visual social media data

1.5.2 Empirical contributions

The dissertation provides substantial empirical insights through large-scale applications of the developed methods. These empirical contributions not only validate the methodological approaches but also generate new knowledge about urban social phenomena across different European contexts.

- **Characterization of the diversity of segregation patterns:** Comprehensive analysis of all Dutch municipalities revealing fine diversity in segregation patterns, with six of eight conceptual segregation types observed in practice
- **Longitudinal segregation dynamics:** First comprehensive analysis tracking evolution of segregation patterns across 82 Dutch municipalities (2015-2020),

revealing convergence in segregation patterns between large and highly segregated cities and mid-size less-segregated towns

- **Cross-European comparative analysis:** Large-scale study of 520 cities across 8 European countries demonstrating no general alignment between residential segregation and urban fragmentation
- **Fine-resolution urban unrest mapping:** Unprecedented fine-grained spatiotemporal analysis of riot dynamics at 300×300 meter resolution and hourly intervals across 7 cities in France

Chapter 2

Characterizing residential segregation in cities using intensity, separation, and scale indicators

Abstract

Residential segregation is known to exacerbate inequality across groups in many aspects of people's life such as education achievements, well-being, and health condition. While residential segregation is known to vary across cities, we lack understanding of the diversity of segregation patterns and the urban characteristics that shape these different manifestations. This chapter addresses this knowledge gap by developing a framework for characterizing segregation patterns along three distinct dimensions: intensity, separation, and scale. When applied to the Netherlands to study segregation of population with a migration background, the framework reveals substantial diversity in segregation patterns across urban contexts, challenging assumptions of uniform spatial structure across cities.

Section 2.1 outlines the limitations of existing segregation measurement approaches and establishes the theoretical foundation for identifying and characterizing the spatial structure of segregation. Section 2.2 presents the proposed methodology, introducing the three-dimensional characterization framework and the agglomerative clustering approach for delineating spatially segregated regions from small census units. Section 2.3 presents the case study and the data. Section 2.4 presents the empirical results, demonstrating the diversity of segregation patterns across Dutch municipalities and how patterns relate to city characteristics. Section 2.5 concludes this chapter with a discussion of the implications for segregation research and policy. A data story summarizing the findings of this research is available [here](#) link.

2.1 Introduction

Residential segregation, here understood as the uneven distribution of social groups in space (**Reardon & O'Sullivan, 2004**), is a persistent problem in many cities in the world

This chapter is based on the journal article: Spierenburg, L., van Cranenburgh, S., Cats, O. (2023). Characterizing residential segregation in cities using intensity, separation, and scale indicators. *Computers, Environment and Urban Systems*. 103, 101990.

(**Tammaru et al., 2015; Wang et al., 2018**). Occurring along one or several social characteristics — such as income, religion, or migration background — segregation often exacerbates inequality across groups in many aspects of people's life, notably in education achievements, well-being, and health condition (**Ludwig et al., 2012; Nieuwenhuis et al., 2020; Owens, 2018; Tóth et al., 2021; Williams & Collins, 2001**). Consequently, numerous local and national authorities consider residential segregation a priority issue (**Gemeente Amsterdam, 2017; Mayor of London, 2021; the United Nations, 2017; Ville de Paris, 2021**). In order to adequately tackle residential segregation, these authorities need tools for characterizing it and for understanding its underlying determinants. A body of literature addresses these two needs, proposing indicators to quantify residential segregation patterns, and relating these patterns to city characteristics such as city size, social mix, or income inequality (**Durlauf, 1996; Gordon & Monastiriotis, 2006; Krupka, 2007; Massey & Denton, 1988; Scarpa, 2015; Theil & Finizza, 1971; Veneri et al., 2021**).

When characterizing segregation patterns, policymakers should account for the multi-dimensionality of the phenomenon (**Massey & Denton, 1988; Petrović et al., 2018**). Widely studied dimensions include: the degree to which groups are over-represented in certain regions, the proportion of people living in regions in which their group is over-represented, and the spatial extent of these regions; respectively called segregation intensity, separation, and scale in our paper (**Andersson et al., 2018; Clark et al., 2015; Fowler, 2016; Lan et al., 2020; Olteanu et al., 2019; Östh et al., 2015; Reardon & O'Sullivan, 2004**). Capturing all these dimensions independently is crucial for two reasons. First, they contribute separately to the segregation experienced by the population. For example, for the same intensity of residential segregation, larger spatial scale may result in higher segregation at school, as distance plays an important role in the allocation of children to schools (**Gramberg, 1998; Taylor & Gorard, 2001**). Second, policies addressing residential segregation may not necessarily affect all dimensions equally.

Several studies have introduced frameworks capturing segregation intensity, separation, and scale, but they all share the shortcoming of measuring separation and scale indirectly, i.e. as a function of the intensity indicator. For instance, the scale indicators proposed by **Olteanu et al., 2019** and **Lan et al., 2020** are derived from the spatial entropy index, which would also be used for measuring segregation intensity (**Theil & Finizza, 1971**). When comparing cities, the variation observed in the different indicators is thus partly governed by the index they are based upon, which inhibits the robustness of a cross-sectional study. Measuring directly intensity, separation, and scale would allow to better disentangle these dimensions and to perform a meaningful comparison of segregation patterns across cities. This would unlock a great, yet untapped potential to study the different faces of segregation in cities, and to explore how certain segregation patterns associate with city characteristics.

While several studies propose metrics for intensity, separation, and scale in the literature; a method for measuring these dimensions separately is lacking. We

address this methodological gap and then apply the proposed approach to all Dutch municipalities. Furthermore, we make two substantive contributions. First, we characterize segregation patterns in the Netherlands and explore similarities and differences across Dutch municipalities. Second, we demonstrate how the approach can be used to relate segregation patterns to several city characteristics, such as the city population, social mix, and income inequality, which have often been associated with segregation — see for example **Gordon and Monastiriotis, 2006; Krupka, 2007; Moreno-Monroy and Veneri, 2018; Musterd et al., 2015; Natale et al., 2018.**

In this study, we specifically investigate the spatial segregation of people with a non-western migration background in the Netherlands. Segregation between Dutch natives and the foreign-origin population is now a major issue for local and national authorities, as it is deemed to hamper integration and social mobility (**Hartog & Zorlu, 2009; Tselios et al., 2015; Vervoort, 2012; Zorlu & Mulder, 2008**). Analyzing segregation from a multi-dimensional perspective is particularly suited in this context, as high segregation intensity associated with large scale is deemed to hamper integration (**Tselios et al., 2015**).

2.2 Method

To study segregation, this work considers the city fabric as an ensemble of segregated and mixed regions forming a certain segregation pattern. We propose to characterize and quantify residential segregation patterns by means of the following three metrics: intensity, separation, and scale. Intensity and separation relate to the spatial distribution of groups across regions, whereas scale relates to the size of the segregated regions. The proposed indicators are described and formally introduced in subsection 2.2.1. Before one can measure these indicators, it is necessary to first delineate segregated and mixed regions. To this end, we use a regionalization method to demarcate regions in which the potential to encounter individuals from each group of interest (potential exposure to each group) is uniform. We then identify a region as segregated if the potential exposure to a group in the region is significantly larger than average in the perimeter of study (subsection 2.2.2). The computation of the potential exposure is described in subsection 2.2.3. Figure 2.1 below illustrates the different steps involved in the analysis, taking the municipality of Leiden as an example. The map on the left depicts the share of the group of interest in the population residing in each spatial unit (using raw demographic data). The middle map displays the exposure levels in each spatial unit, and the map on the right displays the regions where the potential exposure to the group of interest is significantly larger than the municipality average. Finally, intensity, separation, and scale can be used to summarize the characteristics of the segregated regions delineated on the right map in figure 2.1.

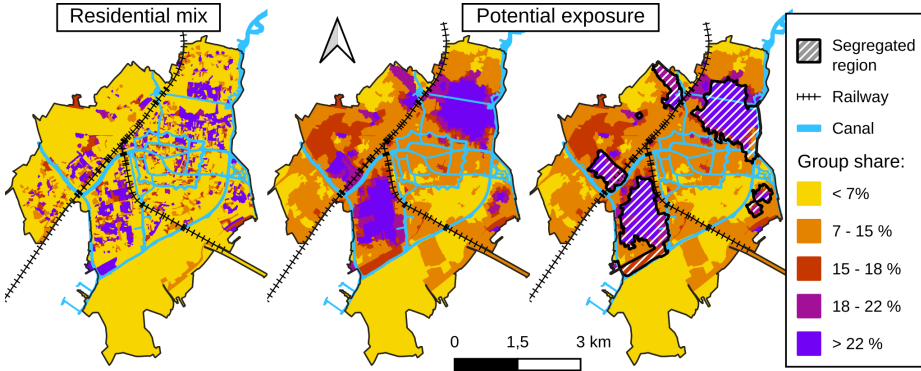


Figure 2.1 Delineation of segregated regions in Leiden. The color indicates the proportion of individuals from the group of interest among the population either residing in the spatial unit (left map), or able to reach it (center and right map). The right map displays the regions identified as segregated.

2.2.1 Indicators

This work proposes three indicators to characterize segregation of a group k in a city: intensity, separation, and scale. Together, they can express a wide diversity of segregation patterns. This subsection defines each metric formally and demonstrates the patterns they can capture using a simple example (figure 2.2). We assume here that segregated regions have already been identified and are available as input for subsequent analysis, as in the right map of figure 2.1 (see subsection 2.2.2 for more details). This study considers two groups, one of them being the group of interest. A region R is an ensemble of spatial units. It has one of the three following statuses, indicated by a variable s : the potential exposure to the group of interest is significantly lower than the city average ($s = -1$), higher than the city average ($s = 1$), or mixed ($s = 0$). The potential exposure to group k estimates the share of individuals from group k in the total population that one would encounter in a given space (spatial unit or region), in percentage terms.

Intensity

The segregation intensity of group k is the difference between the potential exposure to group k in all segregated regions $r \in U_{s=1}$ and the city average μ_k , expressed in percentage points (equation 2.1). In this equation, n_R is the population of region R , $n_{U_{s=1}}$ is the population living in the ensemble of regions $U_{s=1}$, and π_{Rk} is the potential exposure to group k in region R . The intensity is therefore the average exposure in regions that are overexposed to group k , weighted by their population, compared to the city average. More details on how to obtain π_{Rk} — which reflects the extent to which region R is exposed to group k — are provided in subsection 2.2.3.

$$\phi_k = \sum_{R \in U_{s=1}} \frac{n_R}{n_{U_{s=1}}} \pi_{Rk} - \mu_k \quad (2.1)$$

Separation

The separation of group k is the proportion of individuals from this group living in a region that is overexposed to group k (equation 2.2). It indicates the degree to which individuals from the group of interest live in a segregated region.

$$\chi_k = \frac{n_{k,U_{s=1}}}{n_{k,tot}} \quad (2.2)$$

Scale

The scale is the median size of segregated regions, in terms of population size. Half of the people living in a region where group k is over-represented reside in a region that is more (or equally) populated than the scale indicator. It can be expressed in absolute terms (equation 2.3), in number of inhabitants, or relative to the city population (equation 2.4).

$$\psi_k = n_{med\{U_{s=1}\}} \quad (2.3)$$

$$\tilde{\psi}_k = \frac{\psi_k}{n_{tot}} \quad (2.4)$$

Characterizing segregation patterns

The combination of these three distinctive indicators can cover a wide variety of residential segregation patterns. Figure 2.2 displays eight hypothetical cities exhibiting different segregation patterns, and shows the intensity, separation, and scale for each setting in a three-dimensional plot. In addition, we also display for each setting the corresponding spatial entropy index H , a widely used 1-dimensional indicator for residential segregation (**Theil & Finizza, 1971**). The example is crafted so as to showcase two levels for each indicator (high and low) and demonstrate that all combinations between the indicators are in principle possible. For instance, situation A2 represents a segregation pattern with high intensity, low scale, and low separation.

These two figures highlight the advantages of using three indicators over the H index alone. In the example, the H index can clearly differentiate segregation patterns where both separation and intensity are low (situations A1 and B1), from patterns where they are both high (situations C2 and D2): the H index is lower than 3 % in the first case and higher than 15 % in the second case. However, it does not allow to differentiate situations where separation is high and intensity is low (e.g. C1), from situations where separation is low and intensity is high (e.g. A2): the H index is the same for situations C1 and A2, even though the two segregation patterns are remarkably different. Moreover, the H index value does not provide any information on the scale. Finally, it does not provide an

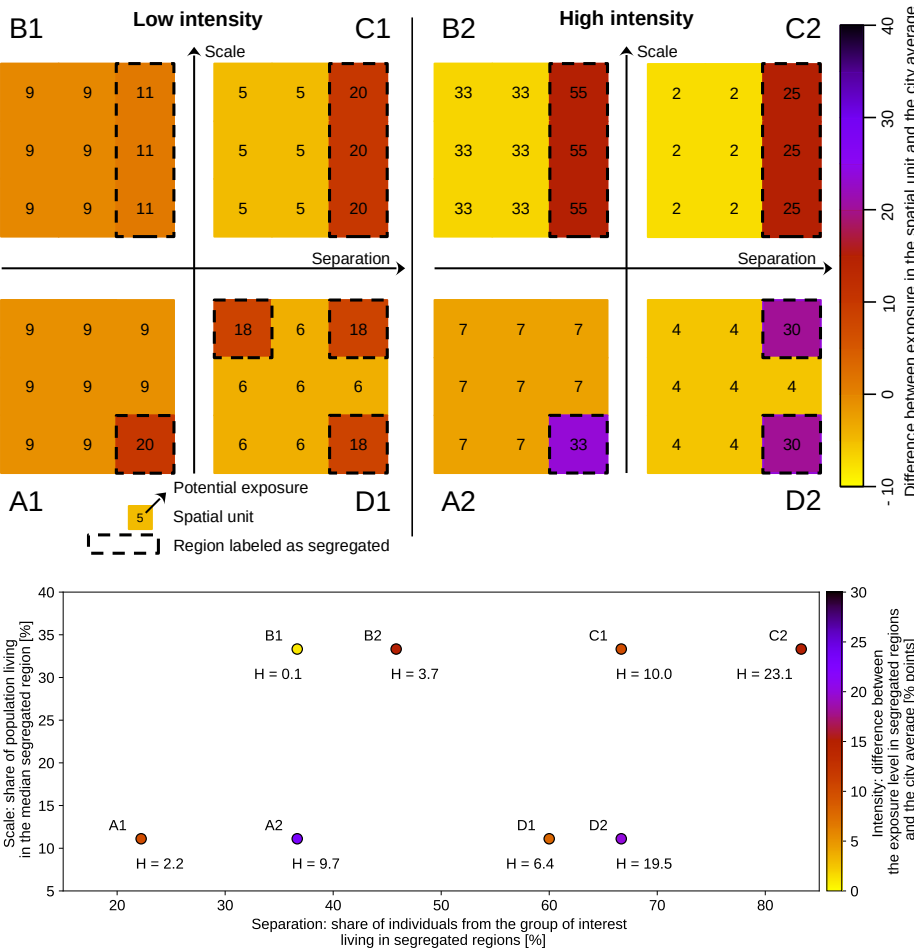


Figure 2.2 Eight conceptual examples of cities with nine equally populated spatial units (top) demonstrating the diversity of segregation patterns that the intensity, separation, and scale indicators can capture in a three dimensional plot (bottom). In the conceptual examples, the value in a spatial unit indicates the potential exposure to the group of interest, while the color represents the difference between the potential exposure and the city average. Finally, regions overexposed to the group of interest are delineated by the dashed black borders. In the plot, the color represents the difference between the average potential exposure in these regions and the city average. In all settings except for setting B2, the average exposure in the city is set to 10 % (setting B2 requires a higher city average in order to manifest itself). In the plot, the H value shown below each dot is the spatial entropy index corresponding to each segregation pattern.

intuitive representation of the segregation pattern. An H index of 4 % for a city is hard to interpret for a policymaker if it is not benchmarked against other cities. The metrics

proposed in this work explicitly target these drawbacks. They differentiate all eight proposed cases and can be intuitively communicated to policymakers. For instance, a separation index of 50 % means that half of the people from the group of interest live in a region considered as overexposed to that group. Moreover, the different metrics are complementary, providing information on different aspects of urban segregation thereby making them suited for a meaningful cross-city comparison.

2.2.2 Delineation of spatially segregated regions

The indicators proposed in subsection 2.2.1 require a meaningful division of the urban fabric into segregated and mixed regions as input. This work uses a regionalization method, namely agglomerative clustering to aggregate contiguous spatial units into regions that are homogeneous in terms of potential exposure to the group of interest (subsection 2.2.2). Subsequently, regions in which the potential exposure is significantly larger than the city average are labeled as segregated (subsection 2.2.2).

Agglomerative clustering

We use agglomerative clustering to group spatial units into larger regions (**Theodoridis & Koutroumbas, 2008**). In the initialization phase, all spatial units are considered as independent clusters. Then, one merges clusters (hereafter called regions) iteratively. Only adjacent regions may be merged. For each iteration, one considers the dissimilarity between each pair of regions (in terms of potential exposure), and merges the two most similar regions. The similarity between regions is determined by the Ward distance, shown in equation 2.5 measuring the increase in the within-region variance when two of them are merged (**Müllner, 2011**). In equation 2.5, ν_i is the number of spatial units in region i .

$$d(i \cup j, k) = \sqrt{\frac{(\nu_i + \nu_k) \cdot d(i, k)^2 + (\nu_j + \nu_k) \cdot d(j, k)^2 - \nu_k \cdot d(i, j)^2}{\nu_i + \nu_j + \nu_k}} \quad (2.5)$$

Figure 2.3 provides an example of agglomerative clustering. At first, all spatial units (A, B, C, D , and E on the left side of the figure) are treated as individual regions. The dissimilarity $d(i, j)$ between each pair (i, j) of these initial regions is the difference between the exposure level in the two spatial units. The most similar regions are A and E , but they cannot be merged because they are not adjacent. Instead, A and B are merged. Then the distance between the newly formed region $A \cup B$ and any other region k is computed using formula 2.5. We repeat the procedure until a stopping criterion is satisfied. The dendrogram in the right of figure 2.3 summarizes the successive merges, along with the corresponding increase in within-region variance.

When aggregating spatial units together, one should ensure that spatial units are adjacent, so that regions are spatially continuous. For that, we use a connectivity matrix to ban merging operations that would result in spatially discontinuous regions. The

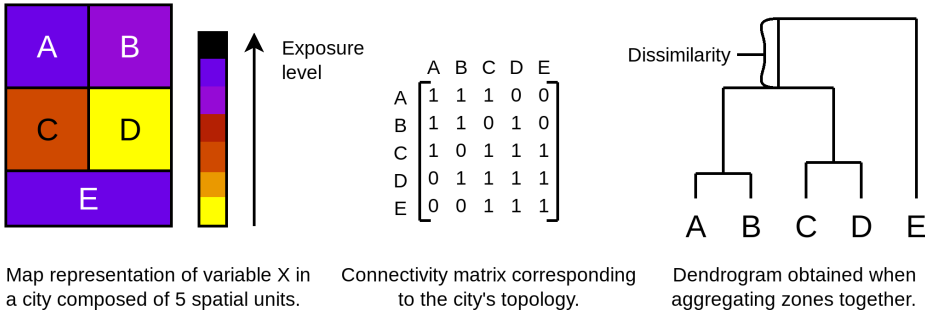


Figure 2.3 Example of agglomerative clustering applied to an illustrative city (left). The connectivity matrix (center) indicates the spatial units that are adjacent. The dendrogram (right) depicts the successive merging operations between regions.

connectivity matrix Ω for a city with ν initial spatial units is a $\nu \times \nu$ matrix, in which component ω_{ij} is one if i is adjacent to j , and zero otherwise (see center of figure 2.3).

The agglomerative process is stopped when the increase in within-region variance exceeds a certain threshold. Like in any clustering procedure, one needs to find a suitable threshold that will strike a good balance between identifying many small and homogeneous regions or too few large and heterogeneous regions. If the threshold is set too high, the algorithm aggregates regions that do not have comparable exposure levels. If the threshold is set too low, the algorithm misses aggregating regions that should belong to the same larger region. We tune this parameter empirically by testing a wide range of values, and investigating the consistency of detected regions.

Labeling a region as segregated

The average exposure in region R is computed using equation 2.6. In this equation, the potential exposure in spatial unit i is weighted by the population n_i residing in this spatial unit (variable θ_i represents the weight of spatial unit i in the region). We label region R as segregated if the average exposure in region $\bar{y}_R$ is significantly larger than the average exposure μ if the groups were to be randomly distributed in space. This is defined by referring to a difference of more than one standard deviation $\sigma_{\bar{y}_R}$, away from the average exposure in the city μ (see equation 2.7). In this equation, the region R is segregated when s is 1 (overexposure) or -1 (underexposure). When s is 0, the region is considered as mixed.

$$\overline{y_R} = \frac{\sum_{i \in R} n_i y_i}{\sum_{i' \in R} n_{i'}} = \sum_{i \in R} \theta_i y_i \quad (2.6)$$

$$s = \begin{cases} -1 & \text{if } \frac{\overline{y_R} - \mu}{\sigma_{\overline{y_R}}} \leq -1 \\ 0 & \text{if } -1 < \frac{\overline{y_R} - \mu}{\sigma_{\overline{y_R}}} < 1 \\ 1 & \text{if } \frac{\overline{y_R} - \mu}{\sigma_{\overline{y_R}}} \geq 1 \end{cases} \quad (2.7)$$

Adjacent regions having the same value for s are merged.

2.2.3 Potential exposure

We compute the potential exposure in each spatial unit before aggregating them into homogeneous regions. The potential exposure to a given group in a spatial unit aims at representing the potential to encounter the group in the spatial unit, relative to the potential to encounter any group. One could also aggregate units using the residential mix (left map in figure 2.1). However, the residential mix is arguably less representative of the segregation experienced by individuals: if one spatial unit highly populated by a group is surrounded by units deserted by this group then the segregation experienced is most likely lower than what the residential mix would have indicated. Furthermore, the potential exposure also has a smoother distribution (center map in figure 2.1), which is more suited for aggregating units into regions (Spierenburg et al., 2022).

We use an accessibility metric to quantify the ease with which people from each group can reach the centroid of any given spatial unit. People able to reach the spatial unit are weighted using the walking time between their place of residence and the spatial unit based on the fundamental principle of human mobility patterns proposed by **Schlöpfer et al., 2021**: the further away people live, the less likely they are to visit the spatial unit (subsection 2.2.3). Then, the potential exposure to a certain group in a spatial unit corresponds to the share of that group in the population able to reach this spatial unit (subsection 2.2.3).

Travel impedance

For a given destination spatial unit, we first determine the origin spatial units located within an acceptable walking distance from it. The shortest walking distance between any spatial unit and any other spatial unit is computed using the street layout. The walking time is calculated from the walking distance, using a walking speed of 4.5 km/h. Then, inhabitants living in the origin spatial units are weighted based on the walking time and the travel impedance function described by equation 2.8.

$$w(t) = \begin{cases} 1 & \text{if } 0 \leq t[s] < 60 \\ \frac{3600}{t^2} & \text{if } 60 \leq t < 1200 \\ 0 & \text{if } t \geq 1200 \end{cases} \quad (2.8)$$

The travel impedance shown in equation 2.8 is derived from the work of **Schläpfer et al., 2021** providing a law to model the visitation pattern of individuals. This law is expressed in equation 2.9. $\rho_i(t, f)$ is the influx of visitors coming to place i , living at travel time t^1 from i and visiting i at frequency f . α_i is a constant that depends on the destination place i , it relates to the attractiveness of i . η is set to 2 (derived empirically). The number of instances in which someone living at walking time t visits i at frequency f is $\rho_i(t, f) \cdot f$. Next, the number of instances in which someone living at a walking time t visits i is obtained from equations 2.10 and 2.11. We use this relation to model the number of visits of inhabitants from one spatial unit to another as a function of the walking time separating the two units (equation 2.11). We set a cut-off of 20-minutes travel time to limit the number of shortest paths between spatial units to be computed, i.e. we stop exploring a path if the duration exceeds 20 minutes. All travel times shorter than 1 minute are set to 1 minute. The constant c_i does not affect the potential exposure indicator (see subsection 2.2.3). under the assumption that the attractiveness is constant across groups.

$$\rho_i(t, f) = \frac{\alpha_i}{(t \cdot f)^\eta} \quad (2.9)$$

$$w_i(t) = \int \rho_i(t, f) f df = \frac{\alpha_i}{t^2} \int \frac{1}{f} df \quad (2.10)$$

$$w_i(t) = \frac{c_i}{t^2} \quad \text{where } c_i = \alpha_i \int \frac{1}{f} df \quad (2.11)$$

Potential exposure indicator

The likelihood that people from group k living in spatial unit j visit spatial unit i is denoted by $w(t_{ij})$. The likelihood that people from group k visit destination i , noted q_{ik} is therefore the sum of likelihood over all origin spatial units (equation 2.12). Then, π_{ik} the potential exposure to group k in spatial unit i is the share of visits from group k to i in all visits to spatial unit i . It can be observed that constant c_i does not affect the potential exposure indicator since it appears in both the numerator and the denominator of equation 2.13.

$$q_{ik} = \sum_j \frac{c_i \cdot n_{jk}}{t_{ij}^2} = c_i \cdot \sum_j n_{jk} / t_{ij}^2 \quad (2.12)$$

$$\pi_{ik} = q_{ik} / \sum_{k'} q_{ik'} = \frac{c_i \cdot \sum_j n_{jk} / t_{ij}^2}{c_i \cdot \sum_{k'} \sum_j n_{jk'} / t_{ij}^2} \quad (2.13)$$

¹In the work of **Schläpfer et al., 2021**, the law relates the influx $\rho_i(t, f)$ to the cartesian distance r rather than travel time t between two places.

2.3 Case study and datasets

2.3.1 Case study

The method proposed in section 2.2 can be used to investigate segregation along any dimension such as income, age, or migration background; within a perimeter of study such as a city or a functional urban area (**Dijkstra et al., 2019**). In the Netherlands, segregation of residents with a migration background persists and hampers their integration to the rest of the population (**Hartog & Zorlu, 2009; Tselios et al., 2015; Vervoort, 2012; Zorlu & Mulder, 2008**). In this paper, we therefore characterize residential segregation of the population with a non-western migration background in every Dutch municipality. The definition of the migration background used by the Centraal Bureau voor de Statistiek is as follows **van Leeuwen, 2020**. Inhabitants are grouped into three categories: individuals from Dutch descent (both parents were born in the Netherlands), individuals with a western migration background (Europe, North America, New Zealand, Australia, Japan, and Indonesia), and individuals with a non-western migration background. Individuals with a migration background originate from another country than the Netherlands or have one of their parents originating from another country than the Netherlands. We take Dutch municipalities, also called *gemeenten*, as perimeters of study, as they are local authorities responsible for urban planning and housing development of a well-demarcated area.

After characterizing residential segregation patterns for every Dutch municipality, we relate the observed patterns to three characteristics often associated with segregation: number of inhabitants, social mix, and income inequality (**Gordon & Monastiriotis, 2006; Krupka, 2007; Moreno-Monroy & Veneri, 2018; Musterd et al., 2015; Natale et al., 2018**).

2.3.2 Datasets

Demographic data are retrieved from the Centraal Bureau voor de Statistiek (**van Leeuwen, 2020**). The spatial units are the 6-digits postcodes. These units are around $100 \times 100 \text{ m}^2$ large and populated by around 50 inhabitants in dense urban areas. This work uses data from 2017. The left map in figure 2.1 displays the population mix in each spatial unit.

Data on the street network are also required to compute walking times between spatial units. These are retrieved from the **OpenStreetMap, 2021** database, using the OSMnx library in Python to extract the street network in every municipality (**Boeing, 2017**).

In this work, we relate segregation patterns to the cities' population, social mix, and income inequality. The city population and the city's social mix are retrieved from the demographic dataset. Data on income inequality in Dutch municipalities are retrieved from the **Centraal Bureau voor de Statistiek, 2019**. The population in Dutch municipalities ranges from 1,000 to 900,000 inhabitants, and around 80 of them are

populated by more than 50,000 inhabitants. The share of individuals with a non-Western migration background ranges from 1 to 38 % of the municipality population, with a median of 5 %. Finally, Dutch municipalities show relatively low income inequality compared to the rest of OECD countries (Veneri et al., 2021).

2

2.3.3 Reflection on the modifiable areal unit problem

The demographic data used in this analysis are provided for spatial units defined by the data supplier. Regionalization processes are generally subject to the modifiable areal unit problem (MAUP) when using such data (Openshaw, 1981). Figure 2.4 illustrates the impact of the MAUP on the outcome of a regionalization process. This figure depicts a city inhabited by two household groups, A (purple) and B (yellow). The data supplier provides demographic data using a certain partition of space into spatial units, such as grid 1 or grid 2 (top and bottom left maps in figure 2.4). In these two examples, each spatial unit (black squares) contains four households. The middle top and bottom maps of figure 2.4 represent the spatial demographic datasets resulting from the two proposed partitions of space, where the color of a spatial unit indicates the share of households from group A residing in it. The data supplier would provide such a dataset. In these two maps, the hatched areas represent regions identified as segregated when clustering spatial units on demographic data. In this case, the outcome of the regionalization method is sensitive to the grid used to partition space: regions obtained using data from grid 1 are different from the ones obtained using data from grid 2.

Our method partially mitigates this issue, as we cluster spatial units on potential exposure instead of demographic data. The top and bottom right map in figure 2.4 represent the potential exposure to group A in each spatial unit, using data from grid 1 and data from grid 2. The potential exposure can be seen as a spatial moving average of demographic data, which filters out small-scale differences between data from grid 1 and data from grid 2 (Spierenburg et al., 2022). Consequently, regions obtained by clustering spatial units on potential exposure are less sensitive to the MAUP than the ones obtained by clustering units based on raw demographic data. In figure 2.4, regions identified as segregated are more consistent on the right-hand maps than on the middle maps.

The size of spatial units also affects the outcome of a regionalization process. More specifically, the minimum size of regions delineated by our method is constrained by the spatial resolution of the demographic dataset ($100 \times 100 \text{ m}^2$ in our case study). A better resolution could be obtained using the micro-data from the Dutch statistics institute, providing data at the household level (Centraal Bureau voor de Statistiek, 2021). However, such data is not openly available. Our use of the postcode data enables the reproducibility of the results.

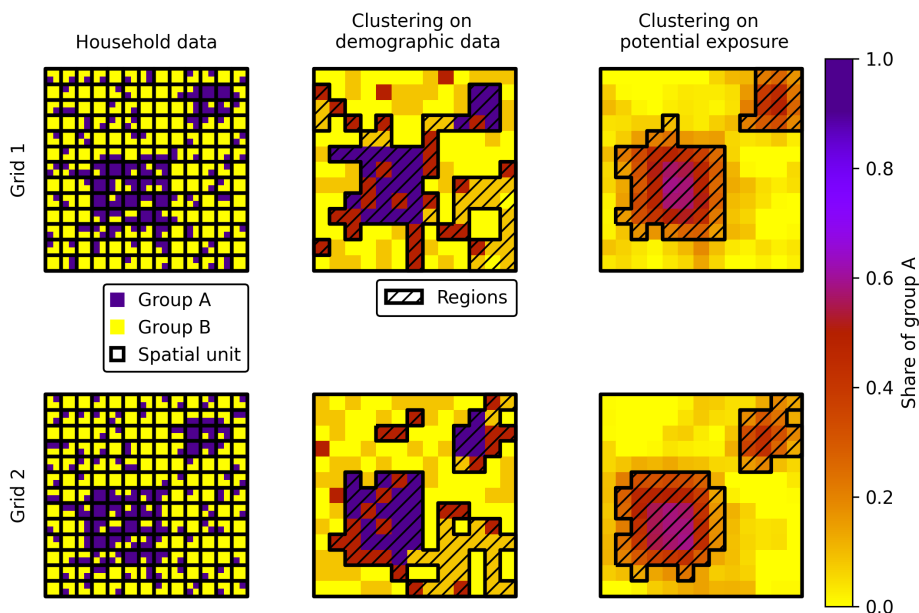


Figure 2.4 Clustering spatial units into regions based on demographic data (middle) and on potential exposure (right), using two different definitions of spatial units (grid 1 and grid 2 on the left).

2.3.4 Code and data produced

The computational workflow supporting this publication is executed via a set of 11 python scripts. We have published the code (MIT license), along with the data produced at each stage of the computational workflow (CC-BY-4.0 license) on the following platform: <https://doi.org/10.4121/c.6237276>.

2.4 Results

The method described in section 2.2 proposes three components of residential segregation: intensity, separation, and scale. The combination of these components allows a multi-faced identification of similarities and differences in the segregation patterns that prevail across municipalities. In this section, we showcase this potential by exploring the variety of segregation patterns in all Dutch municipalities² (subsection 2.4.1), and relating the observed patterns to municipality characteristics often associated with segregation in the literature (subsection 2.4.2) (Gordon & Monastiriotis, 2006; Krupka, 2007; Moreno-Monroy & Veneri, 2018; Musterd et al., 2015; Natale et al., 2018).

²Municipalities defined as of 2017, excluding Baarle-Nassau, a municipality with erratic borders between the Netherlands and Belgium.

2.4.1 Characterization of segregation patterns in the Netherlands

Residential segregation can potentially exhibit a wide range of spatial patterns, as suggested in the hypothetical example provided in figure 2.2, motivating the need to study segregation as a multidimensional phenomenon. In the following, we employ the three proposed measures of segregation to examine how those segregation patterns manifest themselves across Dutch municipalities. This enables determining which ones of the conceptual examples from figure 2.2 are actually encountered in practice, and identify related trends among Dutch municipalities. Figure 2.5 plots the distributions of every indicator across Dutch municipalities, whereas figure 2.6 shows how they relate to one another. Figure 2.7 summarizes the three indicators for all municipalities in one plot, reproducing figure 2.2 that was used in the simple example provided in subsection 2.2.1.

Descriptive statistics

Figure 2.5, depicts the distributions of each segregation indicator across Dutch municipalities. Most Dutch municipalities (75 %) show low intensity (lower than 10 percentage points above the municipality exposure average), and yet a high level of separation (more than 50 % of the group of interest live in segregated regions). Dutch municipalities do not show particular trends regarding the scale of segregation relative to their size. Small, medium, and large-scale segregation occur in relatively even proportions.

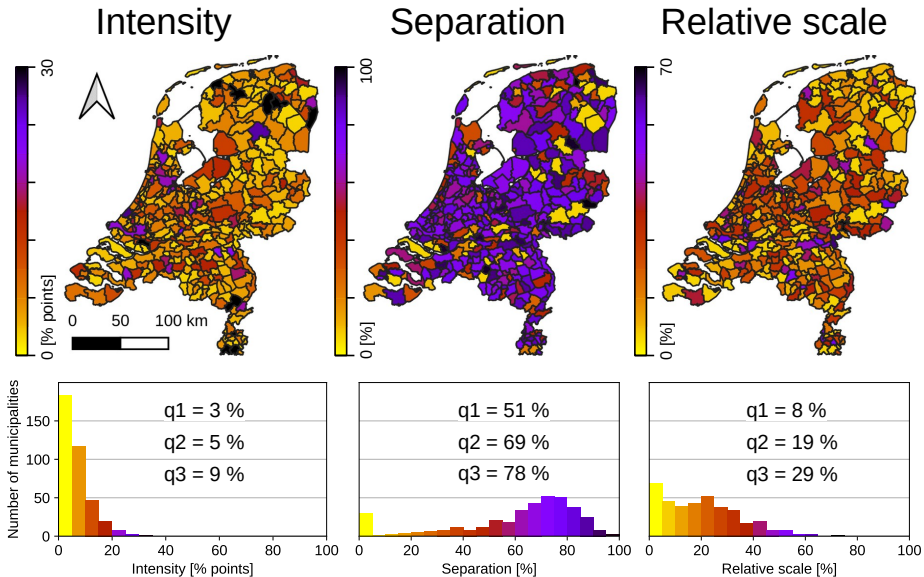


Figure 2.5 Distribution of the intensity, separation, and relative scale indexes. The color represents which interquartile the observations belong to for a given indicator.

The shapes of the distributions of intensity, separation, and relative scale show high

differences when considering skewness and kurtosis. Intensity is right-skewed with a long tail: the skewness is positive (1.4), and the excess kurtosis is large (2.2). The relative scale is also right-skewed (skew of 0.6) with low kurtosis (-0.1). Separation is in contrast left-skewed (skew of -1.2) with low excess kurtosis (0.7). Summarizing residential segregation with a uni-dimensional indicator conceals the disparity in the distributions of the three individual dimensions, and this result already motivates the need to study this phenomenon as multi-dimensional. The maps in figure 2.5 do not directly display any clear geographic trend in relation to the three indicators.

Figure 2.6 displays how the three indicators relate to one another. Relative scale and separation are strongly and positively correlated ($\rho = 0.71$), whereas intensity exercises no to limited negative Pearson correlation with separation ($\rho = -0.09$) and relative scale ($\rho = -0.22$), respectively ³. The spatial entropy index combines intensity and separation, it is low (respectively high) when intensity and separation are low (respectively high). This indicator confounds situations where both indicators have diverging values (high separation low intensity, and low separation high intensity), as shown in subsection 2.2.1. Yet, our results show that such situation happens in practice. High separation does not necessarily come with high intensity, and conversely (no positive Pearson correlation observed), questioning the robustness of such indicator combining these two dimensions.

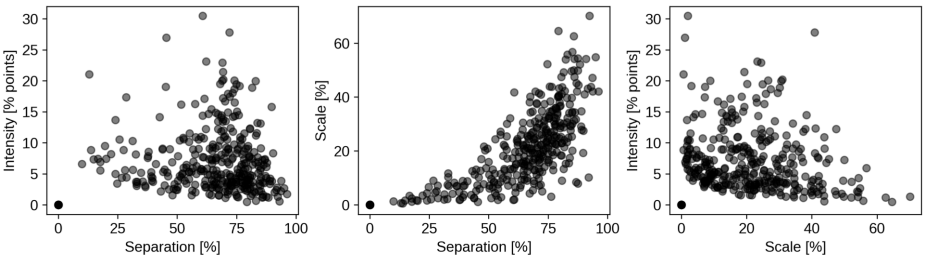


Figure 2.6 Relation between the three different segregation indicators.

The diversity of segregation patterns in Dutch municipalities

Dutch municipalities exhibit a rich diversity of segregation patterns. Figure 2.7 displays the three indicators together in a single plot, along with three examples of municipalities characterized by substantially different segregation patterns. Some municipalities like Losser at the bottom left of the figure show low intensity, low separation, and low relative scale, some at the top right like the Hague show high intensity, high separation, and high relative scale, while some others like Alkmaar show high intensity, high separation, and

³We have removed 9 municipalities with extremely high intensity (greater than 0.3) when computing these correlation coefficients: Gulpen Wittem, Eijsden Margraten, Laarbeek, Schinnen, Tynaarlo, Binnenmaas, Noordenveld, Vlagtwedde, het Bildt. In these municipalities, regions labeled as segregated are asylum shelters and prisons. The social mix observed in these regions is not due to segregation. These outliers show extreme values for the intensity (more than 40 %), as the share of individuals with a migration background is lower than 2 %, in these municipalities.

low relative scale. These examples illustrate that all three dimensions add information when investigating segregation in the Netherlands.

2

Importantly, this result highlights the deficiencies of a uni-dimensional indicator like the spatial entropy index which masks differences in features of segregation that remain uncovered. The spatial entropy index does not capture the scale, and collapses intensity and separation indicators into a single value. A segregation pattern expressing high intensity and low separation can have the same value for the spatial entropy index as another pattern with low separation and high intensity (see figure 2.2 in subsection 2.2.1), while our empirical results in figures 2.6 and 2.7 demonstrate that both patterns exist in reality, and that intensity and separation are uncorrelated ($\rho = -0.09$). This underscores the need to consider intensity and separation independently.

Figure 2.7 conveys another finding about residential segregation in the Netherlands. Municipalities do not converge towards any particular pattern. Six of the eight conceptual examples illustrated in figure 2.2, are observed in the Netherlands, and each of which prevails in a number of municipalities. Dutch municipalities only share one common characteristic: they do not exhibit the combination of a low separation level and large relative scale level (situations *B1* and *B2* in figure 2.2), hence there is no observation in the top left quadrant of figure 2.7. Notwithstanding, Dutch municipalities show little similarities and express high variability across all dimensions considered. Therefore, we investigate in the subsequent subsection 2.4.2 whether, despite the variety of segregation patterns observed, municipalities sharing common characteristics also manifest similar patterns. For instance, figure 2.7 suggests that the most populated municipalities are characterized by a higher separation level (none of the large municipalities appears in the left half of the plot).

2.4.2 Association with municipality characteristics

Subsection 2.4.1 showed that residential segregation takes various configurations in the Netherlands, and that no general segregation pattern prevails in Dutch municipalities. However, these municipalities also differ in many other aspects such as their population size, demography, or economy. We next turn into examining the association between the observed segregation patterns and municipality characteristics. In particular, we investigate this by relating intensity, separation, and relative scale to three characteristics often associated with segregation in the literature: municipality size, share of the group of interest in the municipality population (hereafter called social mix), and income inequality (Gordon & Monastiriotis, 2006; Krupka, 2007; Moreno-Monroy & Veneri, 2018; Natale et al., 2018). In this work, municipality size is measured using the total population living in the municipality, the social mix is the proportion of residents with a non-Western migration background living within the municipal borders, and income inequality is represented by the local Gini coefficient. Figure 2.8 plots the three segregation indicators against the three municipal characteristics considered.

Larger, more mixed, and more unequal municipalities show higher levels of

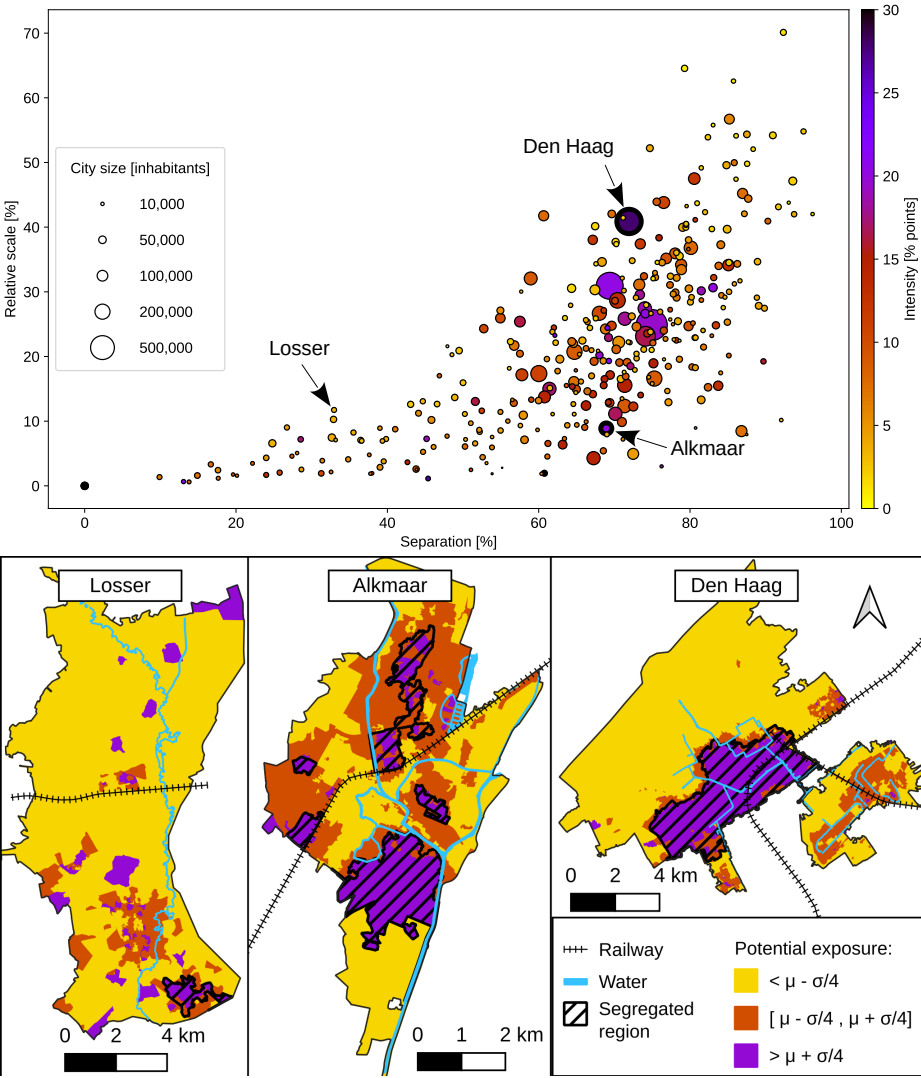


Figure 2.7 Relation between segregation indicators in Dutch municipalities (top plot), along with three observed segregation patterns (bottom three maps). The separation and the relative scale are shown on the x and y axes, respectively, while the color of a point depicts the segregation intensity. The size of points is proportional to the municipality population. The three maps display the exposure levels per spatial unit, along with the identified segregated regions in the municipalities of Losser, Alkmaar, and Den Haag. The color indicates the value of the exposure in a spatial unit compared to the municipality average μ based on the standard deviation σ .

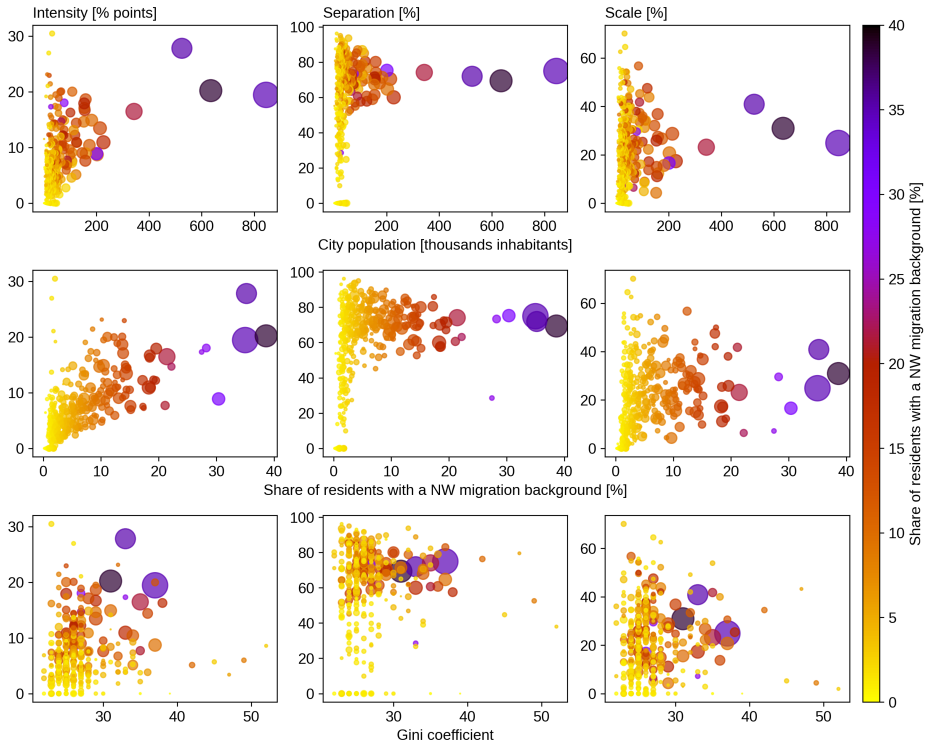


Figure 2.8 Relation between intensity (left column), separation (middle column), relative scale (right column) indicators and municipality size (top row), social mix (middle row), and Gini coefficient (bottom row). The color of an observation represents the social mix. The size of each observation is proportional to the municipality population.

segregation intensity and separation. Intensity is positively correlated with all municipal characteristics: population ($\rho = 0.42$), social mix ($\rho = 0.64$), and the Gini coefficient ($\rho = 0.16$). Even though separation shows lower Pearson correlation values with these three characteristics (0.15, 0.30, and 0.01, respectively), a clear elbow-shape relation can be observed in figure 2.8. However, the association of intensity and separation with municipality population, social mix, and income inequality seem to be the manifestations of a similar underlying phenomenon. First, these characteristics are themselves correlated. Larger municipalities tend in the Netherlands (as well as elsewhere, see [Natale et al., 2018](#) and [Castells-Quintana et al., 2020](#)) to be more mixed and more unequal. Second, the relations between intensity (respectively separation) and municipality population, social mix, and income inequality follow the same linear (respectively elbow-shape) trend in figure 2.8. Hence, researchers exploring the relation between one of these characteristics and segregation should be particularly prudent and control for a potential omitted variable bias.

We find that the scale of segregation relative to the municipality population does not correlate with any of the municipal characteristics considered in figure 2.8. Conversely, the absolute scale, represented by the number of inhabitants living in the median segregated region (see subsection 2.2.1), is remarkably linearly associated with the municipality population (right plot in figure 2.9). The Pearson correlation coefficient is 0.93. One could expect some correlation, as the indicator for the absolute scale has an upper bound that is constrained by the municipality population: it cannot be larger than the total population. Such dependence would usually result in a correlation. However, the relation does not necessarily have to be linear, it could have been sub-linear, if larger municipalities were to be characterized by more segregated regions, rather than larger regions. Such a situation would have resulted in a concave shape of the right plot in figure 2.9. The fact that the right plot in figure 2.9 shows a clear linear relationship between the absolute scale of segregation and the municipality size, suggests that large municipalities are (proportionally) larger versions of small municipalities, as far as the scale of segregation is considered.

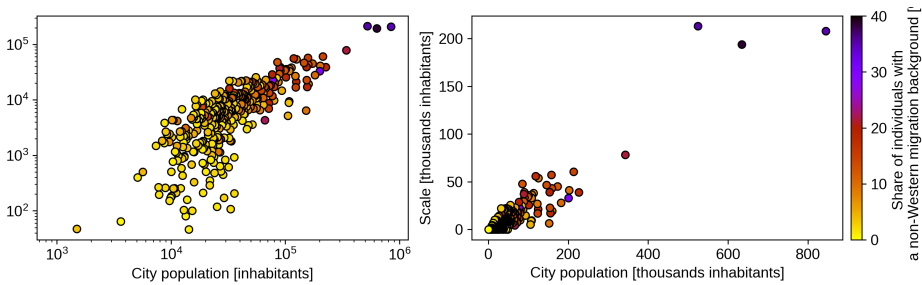


Figure 2.9 Relation between the absolute scale of segregation (population in the median segregated region), and municipality population. The color of an observation represents the share of residents with a non-Western migration background. As the population in Dutch municipalities is such that there are many small municipalities and few large municipalities, we plot the relation in a linear-linear (right) and a log-log graph (left).

2.5 Conclusion

In this paper, we propose a method to characterize residential segregation along three dimensions: intensity, separation, and scale. This method quantifies each dimension separately, which allows to discern every combination of the three indicators. We demonstrate it in a Dutch case study where we could successfully identify a wide variety of residential segregation patterns, and relate these patterns to city characteristics.

We reveal two trends in residential segregation patterns in the Netherlands. First, we observe that both intensity and separation levels associate with municipality size, income inequality, and migrant share. While this result is in line with past research (**Gordon & Monastiriotis, 2006; Natale et al., 2018; Veneri et al., 2021**), it also suggests that the

association of segregation with city size, migrant share, and income inequality are three faces of the same phenomenon. These three variables are correlated in the Netherlands, and all of which are similarly and positively associated with segregation intensity (linear relation) and separation (concave relation). Second, the absolute scale of segregation is highly correlated with municipality size. This implies that larger municipalities have larger segregated regions rather than more segregated regions, when compared to smaller municipalities.

Our study provides two arguments to prefer a multi-dimensional over a uni-dimensional analysis of residential segregation. First, uni-dimensional metrics, such as the spatial entropy index, do not capture the spatial scale, while our result shows that scale varies substantially across Dutch municipalities, which is consistent with the results of **Petrović et al., 2018**. Second, such uni-dimensional metric usually combines intensity and separation, which are in practice negatively correlated in the Netherlands.

In addition to providing a more complete characterization of residential segregation, our indicators are easier to comprehend than the spatial entropy index, which cannot be interpreted directly (**Theil & Finizza, 1971**). Our indicators relate to actual quantities such as the proportion of individuals living in segregated regions for the separation index. This feature makes them more understandable not only for policymakers, but also for a broader audience and thereby may stimulate a more evidence-based public debate.

The full potential of the method is yet untapped. Further research could pursue the association of segregation patterns with geographic, demographic, and urban characteristics; investigate other social dimensions such as education or income; evaluate how segregation patterns evolve over time; or assess how they vary across countries. Exploring these research directions would enrich the understanding of segregation, and, hopefully, assist local and national authorities in designing more inclusive cities.

Chapter 3

Evolution of residential segregation patterns in the Netherlands between 2015 and 2020

Abstract

While the previous chapter revealed substantial diversity in the spatial structure of segregation patterns across urban contexts, little is known about how these patterns evolve temporally and how they respond to demographic change. This chapter addresses this knowledge gap and investigates the temporal evolution of segregation patterns across 82 Dutch municipalities between 2015 and 2020. The analysis reveals a convergence of segregation patterns, with secondary towns with initially low intensity and separation experienced an increase in those indicators while major cities with historically high segregation intensity and separation experienced saw a decrease in those indicators. In addition, we find a positive association between the increase in the share of population with a non-western migration background and the spatial growth of segregated regions.

The remaining part of this study is organized as follows. Section 3.1 introduces the issue and elaborates on the societal importance of the work. Section 3.2 reviews related work. Section 3.3 describes how the method developed in chapter 2 is extended to track segregation patterns in time. Section 3.4 introduces the case study and the data used. Section 3.5 presents the results. Section 3.6 elaborates on the possible causes for the patterns observed and discusses potential future research directions. A data story summarizing the findings of this research is available [here](#) link.

3.1 Introduction

Residential segregation, defined here as the spatial clustering of communities in distinct residential areas, reduces opportunities for inter-groups interactions while exacerbating inequality (Dong et al., 2020; Farber et al., 2014; Laurence et al., 2019; Levy & Razin, 2019; Morales et al., 2019; Semyonov & Glikman, 2008; Tóth et al., 2021). It

This chapter is based on the journal article: Spierenburg, L., van Cranenburgh, S., Cats, O. (2024). *Evolution of residential segregation patterns in the Netherlands between 2015 and 2020*. Cities, 150, 105038.

can occur along different social dimensions, such as ethnicity, income, education, or migration background. In the Netherlands, residential segregation is particularly pressing issue for populations with a non-western migration background, as it negatively affects their life outcomes, in particular educational achievements and naturalization (**Leclerc et al., 2022; van der Greet & Fortuijn, 2017; van der Laan Bouma-Doff, 2007**). As a frequent subject of public discourse, it garners attention from public authorities who regularly express their commitment to mitigate it (**Gemeente Amsterdam, 2017; Gemeente Den Haag, 2021**). A thorough examination of the temporal changes in residential segregation patterns is crucial. It deepens our understanding of its dynamics, aids in the formulation of targeted policies for specific areas needing intervention, and enables the evaluation of the effectiveness of past policies.

Residential segregation can be characterized along three dimensions, 1) intensity, the extent to which the population of interest is over-represented in certain regions of city, 2) separation, the share of this group living in these regions, and 3) scale, the size of these regions (**Spierenburg et al., 2023**). Importantly, segregation is not static but a dynamic phenomenon, evolving over time along these three dimensions as people relocate and new immigrants settle in (**Boschman & van Ham, 2015; Kauppinen & van Ham, 2019; Zorlu & Mulder, 2008**). For instance, an increase in separation with a stable scale might occur if relocation patterns give rise to new social enclaves without affecting the size of existing one. Alternatively, in cities where a particular group predominantly moves out of neighborhoods where they are a minority, intensity would increase with no change in scale. Conversely, in expanding cities, scale may increase while the intensity remains steady if the segregated regions grow in size as the city expands, without any alteration in the group proportions within each region. Monitoring residential segregation through these three dimensions — intensity, separation, and scale — enables the identification of emerging patterns that might not be evident when only considering uni-dimensional indicator such as the dissimilarity or the entropy index (**Reardon & O'Sullivan, 2004**). Multidimensional frameworks are therefore increasingly favored for investigating the evolution of residential segregation (**Lan et al., 2020; Nielsen & Hennerdal, 2017; Sleutjes et al., 2019**).

In the Netherlands, the research by **Sleutjes et al., 2019** used such a multidimensional framework to assess segregation intensity across various scales in the Amsterdam metropolitan region. Their findings revealed that residential segregation remained stable between 2003 and 2014, both in terms of scale and intensity. While this study offers valuable insights at the regional level, the demographic and housing contexts in the Amsterdam metropolitan region differ significantly from those in other Dutch urban areas. For instance, in 2020, the proportion of the population with a non-western migration background was substantially higher than the country average, standing at 24 % compared to 14 % nationwide (**Centraal Bureau voor de Statistiek, 2020**).

To date, the understanding of the evolution of residential segregation patterns at the country level remains limited. More specifically, little is known about potential differences between dynamic regions such as Amsterdam and other areas, and on

potential associations between demographic shifts in cities and changes in residential segregation. This lack of a solid comparative analysis across Dutch municipalities has several negative implications. First, public debates on residential segregation and immigration in the Netherlands is often devoid of empirical evidence. Second, there is a pressing need for authorities to base their policy decisions and resource allocations on robust quantitative data. Third, a thorough grasp of the factors driving changes in residential segregation hinges on having a substantial amount of observational data. Our study aims to fill this knowledge gap, by addressing the following research questions:

How residential segregation patterns along migration background evolved in the Netherlands from 2015 to 2020?

1. What are the geographical trends in the evolution of residential segregation patterns?
2. How do the changes in residential segregation patterns relate to the increase in the population with a non-western migration background?

To this end, we identify residential segregation patterns in the Netherlands and assess how they have evolved between 2015 and 2020. We focus on the segregation of the population with a non-western migration background. Particular attention is given to this group in the Dutch context, as individuals from this group are subject to income inequality and lower educational achievement; and residential segregation contributes to these issues (Albada et al., 2021; Baldwin Hess et al., 2018; Erisen & Kentmen-Cin, 2017; Gracia et al., 2016; OECD, 2018; Thijssen et al., 2021; van de Werfhorst & Heath, 2019).

The remaining part of this study is organized as follows. We describe the method implemented in section 3.3, introduce the case study and the data used in section 3.4, present the results in section 3.5, and conclude our analysis in section 3.6.

3.2 Literature review and theoretical background

A large body of the literature on residential segregation is dedicated to the assessment of the phenomenon. In this section, we review the most commonly used approaches to ground our work in relation to existing studies.

3.2.1 Segregation indicators

Past studies propose a large variety of indicators for residential segregation. The two most established indicators are the dissimilarity and the entropy indexes, both quantifying the extent to which different groups live separated from each other (Duncan & Duncan, 1955; Theil & Finizsa, 1971). Despite being widely used by practitioners and policymakers, they are often criticized for failing to capture critical aspects of

segregation, such as the spatial scale of segregation (**Petrović et al., 2018; White, 1983; Wong, 2004**). The spatial scale of segregation is deemed to have detrimental consequences on the potential for inter-group interactions, as a larger scale implies fewer opportunities for social interactions (**Farber et al., 2014**). This limitation has led to the development of more comprehensive frameworks integrating spatial scale among other additional dimensions (**Brown & Chung, 2006; Feitosa et al., 2007; Fossett, 2017; Massey & Denton, 1988**).

3

Among the different multi-dimensional frameworks developed, the one proposed by **Massey and Denton, 1988** has been particularly influential. They assess residential segregation along 5 dimensions: Evenness, exposure, concentration, centralization, and clustering. Most current multi-dimensional frameworks still consider dimensions related to evenness and clustering. Evenness is *"the differential distribution of two social groups among areal units in a city"*, higher values indicating larger separation of groups (**Massey & Denton, 1988**). Clustering is *"the extent to which areal units inhabited by minority members adjoin one another, or cluster, in space"*, larger values implying larger distance between groups (**Massey & Denton, 1988**). In many studies, clustering is referred to as spatial scale (**Lan et al., 2020; Olteanu et al., 2019; Petrović et al., 2018**).

3.2.2 Comparative analysis

Comparative analyses examine several case studies to unravel general trends in a phenomenon (**Robinson, 2011; Ward, 2010**). They are extensively used in urban science studies as they allow to build theory from the overarching pattern observed (**Nijman, 2007; Storper & Scott, 2016**). Several studies have conducted comparative analyses to assess the evolution of residential segregation by systematically measuring segregation indicators in a set of cities for several time periods (**Bellman et al., 2018; Chodrow, 2017; Delmelle, 2017; Farrell & Lee, 2011; Lan et al., 2020; Zwiers et al., 2018**). For instance, in their comparative studies of neighborhood change across the largest U.S. metropolitan areas, **Farrell and Lee, 2011** and **Delmelle, 2017** identified a sharp decline in the white population in the presence of a large increase in the population from minority groups between 1980 and 2010, a phenomenon coined as tipping point by (**Schelling, 1969**). In the Netherlands, **Zwiers et al., 2018** observed a consistent stability of residential segregation along migration background in the four largest municipalities between 1999 and 2013.

3.2.3 Data-driven approaches

A recurring limitation in the literature is the reliance on predetermined spatial structures, due to the use of administrative boundaries or distance-based grouping of spatial units (**Clark et al., 2015; Ellis et al., 2018; Lan et al., 2020; Nielsen & Hennerdal, 2017; Wright et al., 2014**). Such approaches may not fully capture the evolution of segregation patterns, which often transcend these predefined spatial structures (**Chodrow, 2017**). To address this limitation, a body of literature proposes approaches to identify the spatial structure of residential segregation from the data, moving away from fixed

layout (Chodrow, 2017; Cottrell et al., 2017; Kirkley, 2022; Olteanu et al., 2020; Sousa & Nicosia, 2022; Spierenburg et al., 2023). These approaches typically involve constructing demographically homogeneous regions, a process called regionalization, before quantifying residential segregation along one or several dimensions. This first step aims to maximize within-region homogeneity and between-regions differences, given certain exogenous parameters (e.g. the number of regions). Such parameters are usually tuned by the analyst based on prior knowledge (Chodrow, 2017; Cottrell et al., 2017; Olteanu et al., 2020). Yet, this process is tedious and arbitrary, especially when the parameters are tuned for each city and each time period in the dataset. Therefore, state-of-the-art approaches strive to simplify the parameter tuning stage (Kirkley, 2022; Spierenburg et al., 2023). For instance, in the approach of Spierenburg et al., 2023, a single exogenous parameter needs to be tuned once, and the method can be applied to any other city from the same dataset.

3.2.4 Theoretical framework

This study aims to assess the evolution of residential segregation patterns in the Netherlands using a comparative analysis. In addition to the 4 largest cities considered by Zwiers et al., 2018, we include suburban and secondary towns, as these municipalities are deemed to exhibit radically different patterns than urban cores (Hochstenbach & Musterd, 2018). We consider 82 municipalities in the analysis. This requires the quantification of residential segregation along several dimensions for a large number of cities and several time periods. Therefore, we adopt the theoretical framework presented by Spierenburg et al., 2023, for its agnosticism towards predetermined spatial structure and minimal parameter tuning requirements.

Spierenburg et al., 2023's approach involves identifying demographically homogenous regions within cities, focusing on a specific group of interest. These regions are categorized as under-representing, over-representing, or mixed in relation to the group of interest. Segregation is then analyzed along three dimensions: intensity, separation, and scale. **Intensity** is the extent to which the group of interest is over-represented in regions labelled as such. **Separation** measures the share of the group of interest living in regions in which they are over-represented. **Scale** is the spatial extent of segregated regions.

These indicators can be mapped against the framework proposed by Massey and Denton, 1988. Evenness in Massey and Denton, 1988 is disaggregated into two dimensions: intensity and separation. This allows us to differentiate a situation where a small part of the population of interest is strongly over-represented in certain regions (low separation, high intensity), from a situation where a large part of the population of interest is slightly over-represented in certain regions (high separation, low intensity) (Spierenburg et al., 2023). Clustering in Massey and Denton, 1988 is equivalent to scale in Spierenburg et al., 2023.

3.3 Method

We adopt the approach of **Spierenburg et al., 2023**, and adapt it so as to obtain consistent regions over several periods of time. The method of **Spierenburg et al., 2023** consists of three steps. First, they filter out noise in the demographic data using a spatial moving average (see middle maps in figure 3.1 and section 3.3.1). Second, they delineate regions that are homogeneous in terms of demographics. This step is called regionalization (sections 3.3.4 and 3.3.5). Third, they summarize the observed segregation patterns along three dimensions: intensity, separation, and scale (section 3.3.6). In this work, we perform two additional steps **before** performing the regionalization, to ensure that the variable of interest is comparable across time periods. We normalize the variable of interest across cities and time periods (section 3.3.2), then we apply a statistical transformation to limit the influence of extreme values in the regionalization process (section 3.3.3). The normalization, handling of extreme values, and regionalization are illustrated in the right maps of figure 3.1.

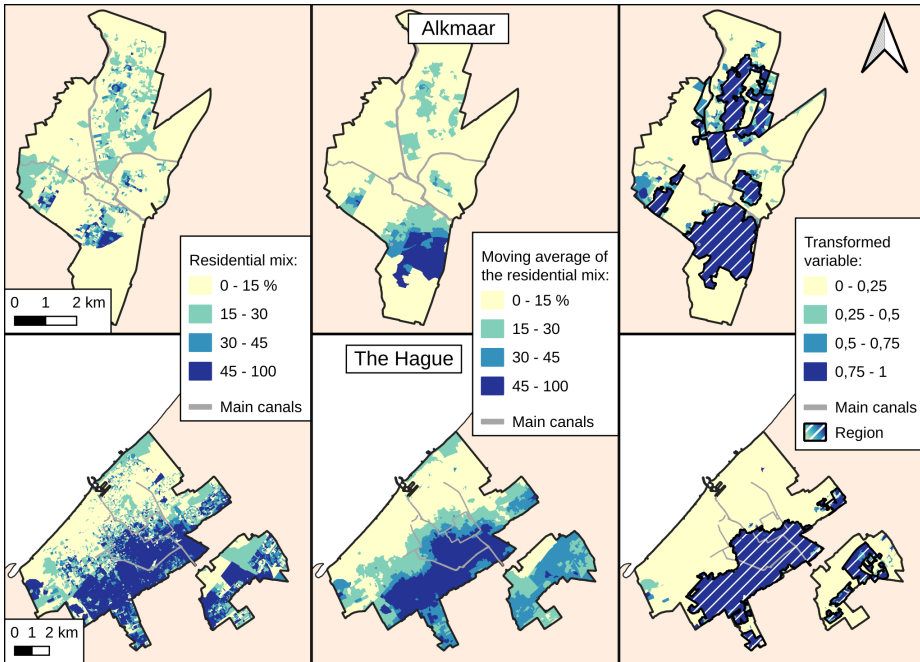


Figure 3.1 Map representation of the regionalization method in the cities of Alkmaar and the Hague in 2015. The left maps represent the raw data, being the share of individuals with a non-western migration background—called residential mix—per spatial unit. The maps in the middle show the spatial moving average of that residential mix. The maps on the right represent the transformed variable (normalization and handling of extreme values) and regions obtained after the regionalization process (for clarity, this map displays only regions over-representing the population with a non-western migration background).

3.3.1 Spatial moving average

There are small-scale local discrepancies in the data (see left maps in figure 3.1). We use a spatial moving average to filter local discrepancies (see equation 3.1), while preserving larger-scale patterns in the data (see middle maps in figure 3.1). In equation 3.1, variable x in unit j is weighted by the coefficient c_{ij} in the spatial average y_i . The weight c_{ij} depends on the total population n_j living in j , and the walking time t_{ij} between i and j , in seconds. The weight c_{ij} increases with the population of unit j and the spatial proximity of j and i . In this case study, x_j is the proportion of individuals with a non-western migration background living in unit j . This spatial moving average can also be seen as the potential to encounter an individual with a non-western migration background in the spatial unit, relative to the potential to encounter anyone, regardless of its group (Spierenburg et al., 2023). In the following, we also name it **potential exposure**, as do Spierenburg et al., 2023.

$$y_i = \frac{\sum_j c_{ij} \cdot x_j}{\sum_j c_{ij}} \text{ where } c_{ij} = \begin{cases} n_j & \text{if } 0 \leq t[s] < 60 \\ 3600 \cdot n_j / t_{ij}^2 & \text{if } 60 \leq t < 1200 \\ 0 & \text{if } t \geq 1200 \end{cases} \quad (3.1)$$

3.3.2 Normalization

The share of the population with a non-western migration background varies significantly from a time period to another and from a municipality to another. For instance, in 2015, it was 12 % in Alkmaar and 35 % in the Hague. The value for the spatial moving average per spatial unit in these two cities are not directly comparable. A value of 15 % in a spatial unit would imply an over-representation of the population with a non-western migration background in Alkmaar, whereas, in the Hague, it would imply an under-representation of that group (see middle maps in figure 3.1). We normalize the variable of interest to enable the comparison of values across cities and different time periods. To this end, for each spatial unit, we measure the deviation of the moving average y_i to its theoretical value μ , considering a case in which groups are randomly allocated in space (see equation 3.2). In equation 3.2, μ is the proportion of individuals with a non-western migration background in the city for the time period considered. It is also the expected value of y_i . the value σ_{y_i} is the standard deviation of y_i if groups were to be randomly distributed in space. The normalization z_i can be seen as the scaled difference of the moving average y_i in i to its theoretical value in a random allocation of groups. z_i is positive if unit i over-represents the group of interest and negative if it under-represents the group of interest.

$$z_i = \frac{y_i - \mu}{\sigma_{y_i}} \quad (3.2)$$

3.3.3 Statistical transformation

The normalization process described above introduces extreme values and regionalization processes tend to overfit extreme values. Numerous spatial units show a potential exposure deviating significantly from the value yielded by the random allocation of social groups. We compress the extreme values resulting from the normalization process using a sigmoid function bounded between 0 and 1. We use the cumulative distribution function of the standard normal Gaussian (see equation 3.3). Therefore, the transformed value can be seen as the probability that a random allocation would result in a lower potential exposure than the one observed in the spatial unit considered. If the potential exposure to individuals with a non-western migration background is significantly larger — respectively smaller — than the city average, the probability will be close to 1 — respectively 0 —. The right maps in figure 3.1 illustrate this transformation.

$$p_i = P(z_i \leq Z_i) = \int_{-\infty}^{Z_i} \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} dz_i \quad (3.3)$$

3.3.4 Regionalization

A regionalization task consists in delineating regions that are homogeneous according to either a variable or a set of variables from spatial units. In this work, we use spatially-constrained agglomerative clustering. We apply it to the statistical transformation described in section 3.3.3. In the initialization phase of this algorithm, all units are considered as individual regions. Then, regions are iteratively merged together, minimizing the within-regions variance (**Ward Jr., 1963**). The agglomerative process is spatially constrained, meaning that only adjacent regions can be merged. We stop aggregating regions when the within-cluster sum-of-squares exceeds a certain threshold, tuned empirically.

3.3.5 Labeling regions

After delineating regions that are homogeneous in terms of demographics, we classify them into three categories: 1) regions over-representing individuals with a non-western migration background ($s = 1$), 2) regions under-representing individuals with a non-western migration background ($s = -1$), 3) mixed regions ($s = 0$). We use the same criterion as in **Spierenburg et al., 2023**. For each region, we compute the average potential exposure to individuals with a non-western migration background $\bar{y}_R$ of all spatial units pertaining to that region R , weighted by the units' population (see equation 3.4). We also compute the theoretical standard deviation $\sigma_{\bar{y}_R}$ of the average potential exposure in the scenario where groups are randomly allocated in space. Then, if the observed average potential exposure is more than two standard deviations away from its theoretical average μ , we label the region as either over- or under-representing individuals with a non-western migration background (depending on the sign), otherwise, the region is labeled as mixed (see equation 3.5). The regions highlighted in the right

maps of figure 3.1 are the ones that over-represent the population with a non-western migration background.

$$\overline{y_R} = \frac{\sum_{i \in R} n_i y_i}{\sum_{i' \in R} n_{i'}} \quad (3.4)$$

$$s = \begin{cases} -1 & \text{if } \frac{\overline{y_R} - \mu}{\sigma_{\overline{y_R}}} \leq -2 \\ 0 & \text{if } -2 < \frac{\overline{y_R} - \mu}{\sigma_{\overline{y_R}}} < 2 \\ 1 & \text{if } \frac{\overline{y_R} - \mu}{\sigma_{\overline{y_R}}} \geq 2 \end{cases} \quad (3.5)$$

3.3.6 Residential segregation indicators

To characterize residential segregation patterns, we adopt the indicators of **Spierenburg et al., 2023**, which characterize the observed pattern along the dimensions of intensity, separation, and scale. **Intensity** is the extent to which individuals with a non-western migration background are over-represented in the regions labeled as over-representing them ($s = 1$ in equation 3.5). It is the difference between the potential exposure to individuals with a non-western migration background in these regions and the city average. **Separation** corresponds to the proportion of individuals with a non-western migration background experiencing segregation. We measure it as the proportion of individuals with a non-western migration background living in regions in which they are over-represented. **Scale** is the spatial extent of regions in which individuals with a non-western migration background are over-represented. We measure it using the median size of regions over-representing individuals with a non-western migration background, in population terms. We express it in percentage terms, relatively to the total city population.

3.4 Data and case study

3.4.1 Dataset

We use open data from the Netherlands National Bureau of Statistics to measure the demographic composition of neighborhoods (**van Leeuwen, 2020**). The spatial units employed in this study are the 6-digits postcodes, with each unit covering a surface smaller than 100x100 square meters in urbanized areas (see figure 3.2), enabling a fine-resolution examination of demographic trends. We consider two demographic groups: 1) individuals with a non-western migration background, and 2) the rest of the population which includes individuals with a western migration background and individuals without a migration background. The non-western migration background is defined based on the criteria provided by the National Bureau of Statistics, encompassing individuals born abroad or having at least one parent born abroad in countries excluding European and North American countries, along with Japan, Indonesia, Australia, New Zealand (**Centraal Bureau voor de Statistiek, 2016**). The dataset does not disaggregate

further the migration background which prevents from studying the segregation by country of origin. Moreover, the data values are rounded to the closest 5 for absolute values and to the closest 10 % for percentage values in the raw dataset, and they are not disclosed if less than 10 people live in the spatial unit. Data are available and consistent across all years between 2015 and 2020. We therefore investigate the evolution of residential segregation patterns between 2015 and 2020.

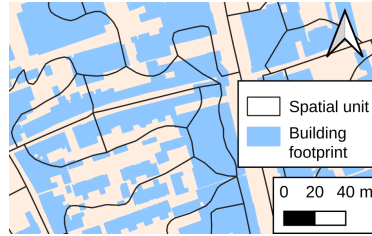


Figure 3.2 Spatial units (6-digits postcodes) in the demographic data used.

3.4.2 Situation in 2015

The composition of the population with a non-western migration background and the patterns of segregation vary significantly from a city to another, influenced by differences in economy, demography, and urban planning. In the Netherlands as of 2015, 12 % of the population had a non-western migration background. However, this figure showed considerable variation across municipalities, ranging from as low as 1 % to as high as 37 %. Our analysis focuses on municipalities with larger populations, specifically those exceeding 50,000 inhabitants, which includes 82 municipalities in total. Less populated municipalities typically have a smaller share of population with a migration background and are less dense. The census data are less reliable in these situations, as they are not disclosed in sparsely populated spatial units (see 3.4.1). We therefore choose to filter out less populated municipalities from our analysis. We consider the municipality boundaries of 2020 (van Leeuwen, 2020). There are some clear geographic trends related to the share of the population with a non-western migration background across Dutch municipalities. In the densely populated and urbanized Randstad¹, individuals with a non-western migration background represented 26 % of the population, compared to 7 % in the rural provinces in the North of the Netherlands (Groningen, Friesland, and Drenthe)², see top maps in figure 3.3. Moreover, within the Randstad itself, there is a notable gap between the four largest urban cores (Amsterdam, Rotterdam, the Hague,

¹The Randstad does not have a formal definition, in this work, we include the following municipalities: Alphen aan den Rijn, Amstelveen, Amsterdam, Capelle aan den IJssel, Delft, Gouda, Haarlem, Haarlemmermeer, Katwijk, Krimpenerwaard, Lansingerland, Leiden, Leidschendam-Voorburg, Nieuwegein, Pijnacker-Nootdorp, Rotterdam, Schiedam, Stichtse Vecht, the Hague, Utrecht, Velsen, Vijfheerenlanden, Vlaardingen, Westland, Woerden, Zaanstad, Zoetermeer

²Considering only the municipalities with more than 50,000 inhabitants in the Randstad and in Noord-Nederland.

and Utrecht) and the neighboring suburban towns regarding the share of population with a non-western migration background (see top right map in figure 3.3).

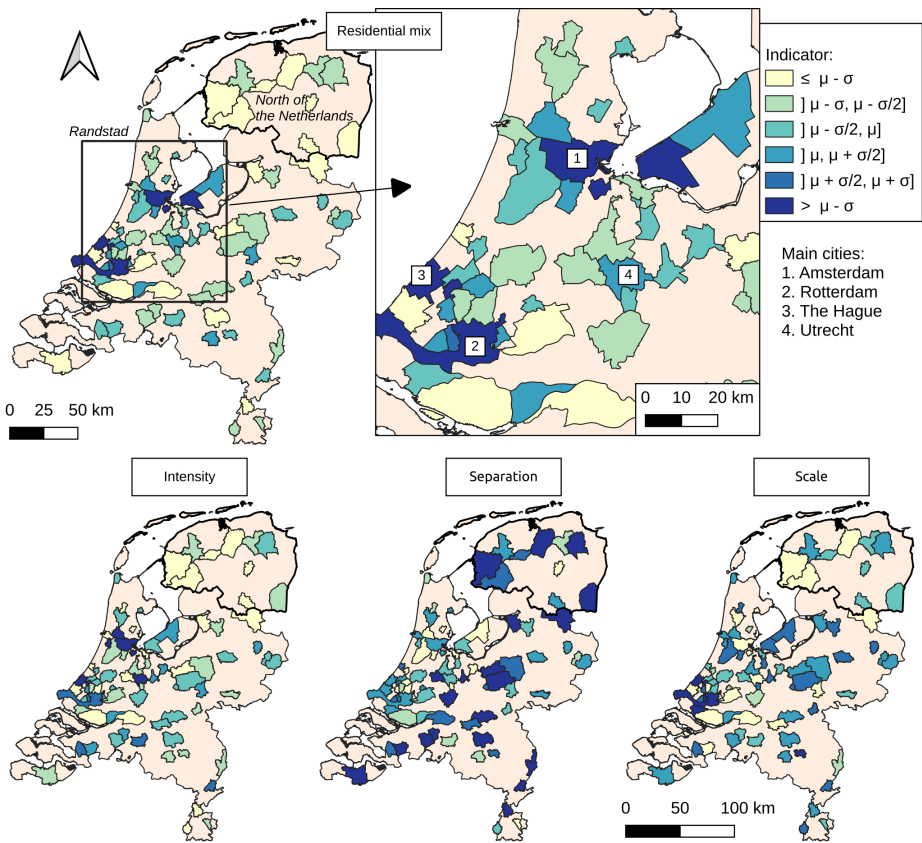


Figure 3.3 Spatial distribution of the proportion of the population with a non-western migration background (called residential mix), segregation intensity, separation, and scale in 2015.

Residential segregation patterns exhibited clear geographic trends in 2015. Urban cores in the Randstad showed much higher intensity than towns in the province of Noord-Nederland (bottom-left map in figure 3.3). Intensity in 2015 was actually strongly correlated with the share of the population with a non-western migration background (Pearson correlation of 0.76). Meanwhile, the degree of separation was lower in the Randstad compared to the other municipalities, as illustrated in the middle map at the bottom of figure 3.3). Scale did not exhibit any particular spatial trend.

3.5 Results

After presenting the overall evolution of residential segregation patterns between 2015 and 2020 in subsection 3.5.1, we organize our analysis in three parts. First, in subsection 3.5.2, we uncover trends across municipalities based on their location (cities from the Randstad compared to other municipalities) and city type (urban cores compared to suburban towns). Second, in subsection 3.5.3, we observe a convergence in segregation patterns across Dutch municipalities. Third, in subsection 3.5.4, we investigate the relation between the rise in the proportion of individuals with a migration background and changes in segregation patterns.

3.5.1 Descriptive statistics

We first investigate the distribution of the change in the residential segregation patterns between 2015 and 2020 (see figure 3.4). There is no clear uniform trend in the evolution of segregation intensity, separation, and scale across Dutch municipalities between 2015 and 2020. On average, intensity decreased by 0.5 percentage point (pp), with a standard deviation of 0.6 pp across municipalities. The change in intensity is centered around 0, and is positive for 52 % of the 82 municipalities considered in our analysis. Separation increased for most municipalities (73 %, with an average increase of 2 pp), but exhibits considerable heterogeneity. Few municipalities show a substantial increase in separation (exceeding 10 pp), while the others show a limited increase or even a decrease in separation. This distribution is skewed to the right, with a standard deviation exceeding the mean. The evolution of the relative scale follows a similar trend to that of separation. It increases for 67 % of the municipalities, the average evolution is 0.9 pp while the standard deviation is 1.2 pp. These results underscore the heterogeneity in the evolution of residential segregation patterns across municipalities. In the subsequent sections, we examine the trends across different municipality characteristics, e.g. geographic location and demographic shift.

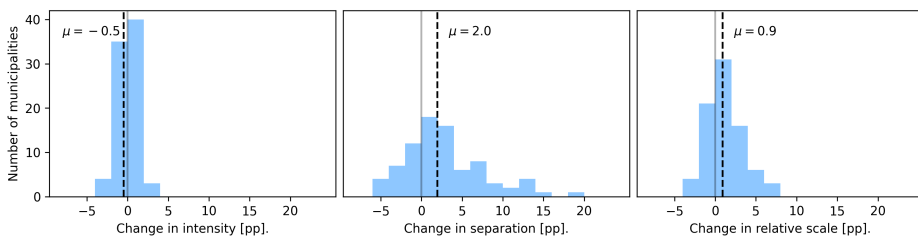


Figure 3.4 Statistical distribution of the change in intensity, separation, and relative scale in percentage points (pp).

3.5.2 Geographical trends

While we do not observe overarching trends in the evolution of segregation across Dutch municipalities, we do identify specific trends associated with geographic characteristics between 2015 and 2020. Regarding intensity, we find a notable decrease in cities in the Randstad region, particularly in larger municipalities, reflecting a potential trend towards a more balanced distribution of individuals with a non-western migration background (map A and chart D in figure 3.5). Interestingly, municipalities outside of the Randstad exhibit an opposite trend, marked by an upswing in segregation intensity. As for separation, we observe a contrasting evolution between core and suburban municipalities within the Randstad. Separation decreased in the urban cores of the Randstad (i.e. Amsterdam, Rotterdam, the Hague, Utrecht) while it increased in suburban town surrounding these cores (map B in figure 3.5). We do not observe any geographic trend regarding the evolution in the relative scale of residential segregation (map C in figure 3.5).

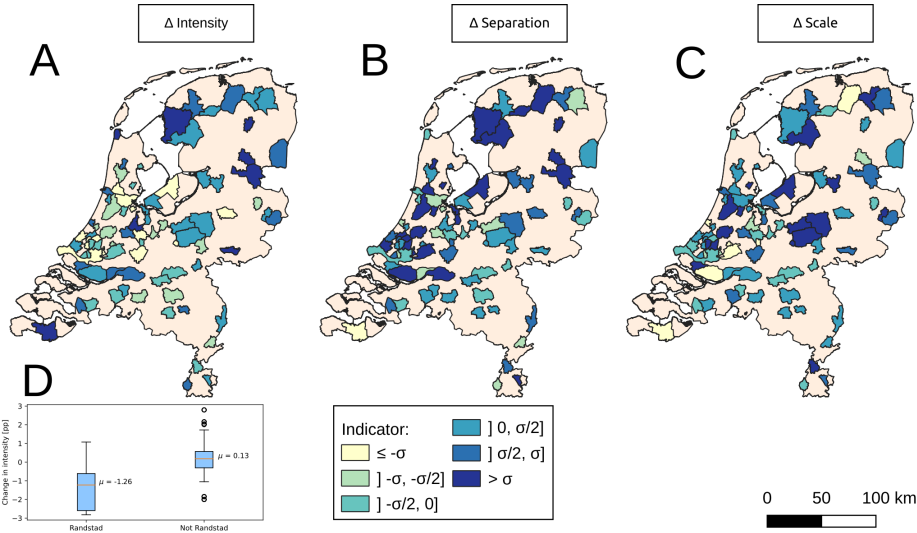


Figure 3.5 Evolution of segregation patterns along intensity, separation, and scale between 2015 and 2020 in Dutch municipalities.

3.5.3 Convergence of segregation patterns

Our analysis reveals a noteworthy convergence of residential segregation patterns along intensity and separation among Dutch municipalities between 2015 and 2020. Regarding intensity, we observe a reduction in the dispersion of the distribution in 2020 compared to 2015 (top left histogram in figure 3.6). The standard deviation of intensity across municipalities decreased from 5.7 pp in 2015 to 4.8 pp in 2020. Notably, the municipalities where intensity was previously the highest experienced a decrease.

Conversely, we observe an increase in intensity in municipalities where intensity was initially the lowest (bottom left plot in figure 3.6). For separation, we also observe a decrease in the distribution's dispersion in 2020 compared to 2015 (top center plot in figure 3.6). The standard deviation of separation across municipalities reduced from 6.8 pp in 2015 to 6.1 pp in 2020. As for intensity, the largest increases in separation occurred in municipalities where separation was initially the lowest (bottom center plot in figure 3.6). In contrast, the indicator of relative scale does not demonstrate any specific trend, indicating that the spatial distribution of segregated regions remained relatively stable over the five-year period (top and bottom right plots in figure 3.6).

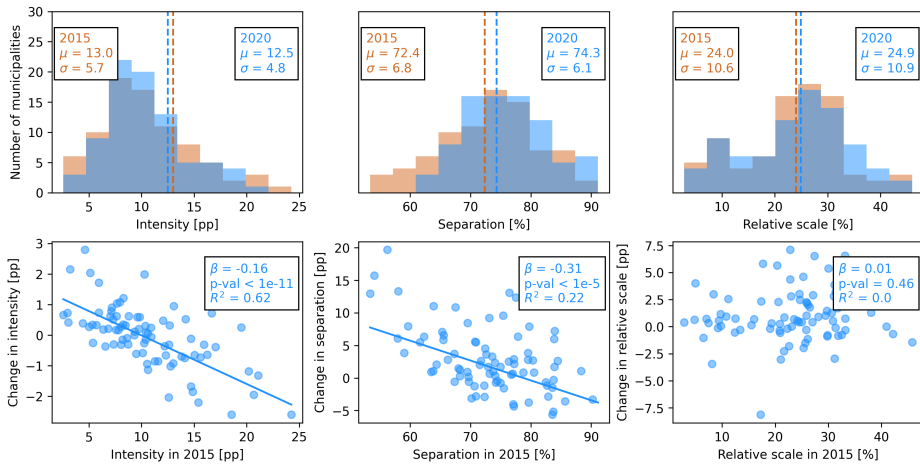


Figure 3.6 Evolution in the distribution of intensity, separation, and relative scale indicators between 2015 and 2020 (top), and relation of the change in the indicators to their levels in 2015 (bottom). pp stands for percentage points.

Regression to the mean could exacerbate the convergence observed. Regression to the mean occurs when extremely high or low measurements, influenced by random variation, tend to move closer to the average in later assessments. For instance, if high or low levels of intensity and separation in 2015 were a product of random variation, we could expect to see these values return closer to the mean by 2020. In practice, randomness in the measurement is weak. The correlation between intensity in 2015 and intensity in 2020 is 0.98 (0.79 in the case of separation), higher correlations suggesting weaker random variations. Moreover, regression to the mean falls short of explaining the reduced variance in intensity and separation over these periods, as illustrated in figure 3.6.

3.5.4 Relation between share of population with a non-western migration background and change in segregation patterns

Between 2015 and 2020, the share of the population having a non-western migration background increased in all municipalities considered by 1.9 percentage points (pp) on average. This upward trend varied, with increases ranging from a modest 0.7 to a significant 6.7 percentage points. This increase is notably higher in suburban towns surrounding urban cores in the Randstad region (see left map in figure 3.7).

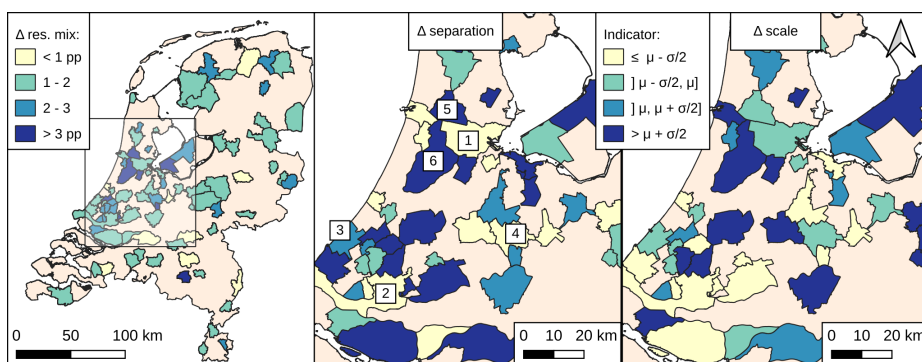


Figure 3.7 Left: Geographic distribution in the evolution of the proportion of population with a non-western migration background (called residential mix), in percentage point (pp). Middle and right: evolution of separation and scale, focusing on the Randstad area. Largest municipalities 1: Amsterdam, 2: Rotterdam, 3: the Hague, 4: Utrecht. Other municipalities mentioned: 5: Zaanstad, 6: Harlemermeer.

Surprisingly, we observe a negative correlation between the change in intensity and the share of the population with a non-western migration background — also called residential mix in this section — in 2015. Municipalities with lower residential mix in 2015 experienced an increase in intensity, while those with higher residential mix saw a decrease in intensity and the relationship appears to be proportional. This nuanced finding reveals that the strong correlation between intensity and residential mix in 2015 (subsection 3.4.2) is weakening between 2015 and 2020. Additionally, there is no significant correlation between the increase in residential mix and the change in intensity, indicating that municipalities with higher increases in the share of the population with a non-western migration background did not necessarily experience larger changes in intensity.

The change in separation and scale is positively associated with the increase in the share of population with a non-western migration background. For instance, in the Randstad region, urban cores experienced a low increase in residential mix and a decrease in separation and scale, meanwhile the high increase in residential mix in the suburbs came with an increase in separation and scale.

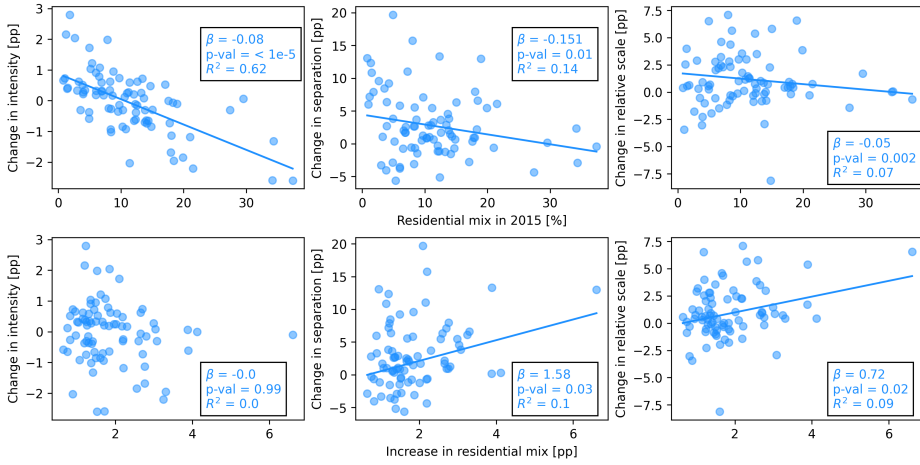


Figure 3.8 Change in the intensity, separation, and scale indicators between 2015 and 2020 related to the proportion of the population with a non-western migration background in 2015 (top) and to the increase in this proportion between 2015 and 2020 (bottom). Percentage points are noted pp in this figure.

Hence, the high increase in the population with a non-western migration in suburban towns in the Randstad did not come with an increase in intensity — social enclaves have not become more homogeneous —. Instead, either the scale of segregation increased (such as in Harlemermeer), or new segregated regions emerged (such as in Zaanstad).

3.6 Discussion and conclusion

This study focuses on the 82 largest municipalities in the Netherlands, covering 60 % of the Dutch population. We choose to exclude rural municipalities due to a limitation associated with the census data used. The dataset provides open data at a highly granular resolution, yet variables on the migration background are (1) not disclosed when less than 5 people live in a spatial unit, (2) rounded to the closest 10 % otherwise (**van Leeuwen, 2020**). In cities with lower density or with lower populations with a non-western migration background, that population may be under-reported. We have filtered out these cities by focusing only on the largest municipalities. This limitation could be addressed by using the microdata from the **Centraal Bureau voor de Statistiek, 2021**.

Our investigation yielded several key findings. First, we do not observe any overarching trend in the Netherlands, as residential segregation neither uniformly improved nor worsened across municipalities. Second, we observe a convergence in residential segregation patterns along segregation intensity and separation between 2015 and 2020. Third, the evolution of segregation patterns associates with the change in the population with a non-western migration background relative to the city population. The

higher the increase in the share of population with a non-western migration background, the higher the increase in separation and scale, while intensity does not correlate with the change in city demographics.

In the following subsections, we hypothesize on potential causes for our findings, identify limitations in our approach, and suggest directions for further research.

3.6.1 Interpretation of the patterns observed

The observed decrease in intensity in the Randstad may be partly attributed to the phenomenon of gentrification (**Janssen et al., 2023**). Affordable neighborhoods, where households with a non-western migration background are often over-represented, become increasingly attractive to young professionals seeking affordable housing in the large urban cores. Consequently, this influx of new residents from diverse backgrounds contributes to a decrease in segregation intensity. However, it is important to recognize that this trend might be temporary. Gentrification tends to drive up property prices and rents, exerting pressure on lower-income households — including those with a non-western migration background — to move out to other affordable neighborhoods (**Hochstenbach & Musterd, 2018**). As a result, this transitional process is viewed as a shift between two distinct segregation patterns by many researchers (**Atkinson et al., 2011; Zuk et al., 2018**).

Another plausible explanation for this could be the influence of public actions and urban planning priorities. Municipalities with higher intensity and separation levels might have segregation issues at the forefront of their agendas (**Gemeente Amsterdam, 2017; Gemeente Den Haag, 2021**). Consequently, they may implement targeted urban planning measures to address residential segregation, which could, if successful, lead to a reduction in these indicators over time. Conversely, municipalities with lower intensity and separation levels might have different priorities, resulting in fewer explicit measures to combat segregation.

3.6.2 Limitation and further research

One limitation of this study is the relatively short time span considered for assessing the evolution of residential segregation patterns. The analysis is conducted over a five-year period, spanning from 2015 to 2020, during which the Netherlands has seen a sharp increase in non-western migration. While this timeframe allows us to capture notable changes in segregation trends, it may not fully account for longer-term dynamics, especially given that urban development can span over a more extended period. While open data on demographics exist from 2010, the variables are consistent only from 2015. Access to data from earlier years, which could be obtainable through the **Centraal Bureau voor de Statistiek, 2021**, would enable the construction of a more extensive temporal dataset. Investigating trends over a more extended period could reveal historical patterns and help in understanding the long-term effects of policies and socio-economic changes on residential segregation in the Netherlands.

This study focuses on the 82 largest municipalities in the Netherlands, covering 60 % of the Dutch population. We do not investigate rural municipalities. This limitation is due to the census data used. The dataset provides open data at a highly granular resolution, yet variables on the migration background are (1) not disclosed when less than 5 people live in a spatial unit, (2) rounded to the closest 10 % otherwise (**van Leeuwen, 2020**). In cities with lower density or with lower population with a non-western migration background, that population may be undersampled. We have filtered out these cities by focusing only on the largest municipalities. Using the microdata from the **Centraal Bureau voor de Statistiek, 2021** would allow to address that limitation.

The population with a non-western migration background is in itself inherently diverse, encompassing variations in education levels, income, country of origin, and migrant generation (**Boterman et al., 2021; Custers & Engbersen, 2022**). Subgroups defined along these dimensions within the broader category may experience residential segregation differently, both in terms of the level of segregation they encounter and the impact it has on their life outcomes. It is important to exercise caution against committing the fallacy of division, wherein observations made on the group as a whole may not necessarily apply uniformly to its subgroups. For instance, a recent immigration of highly-skilled migrants moving in neighborhoods where Dutch natives are over-represented would decrease residential segregation of migrants as a whole, while the situation would remain unchanged for low-skilled migrants. The data used in this work do not allow to disaggregate the population with a non-western migration background into smaller subgroups. Further research could segment the analysis using subcategories such as income and education levels, using the microdata provided by the **Centraal Bureau voor de Statistiek, 2021**.

Finally, while we could identify trends in the evolution of residential segregation across cities, we have left aside trends within cities. In practice, residential segregation patterns may vary within municipalities. For instance, intensity could increase in certain regions while simultaneously decreasing in others within the same urban area. Exploring such regional variations within municipalities could offer a more comprehensive understanding of the underlying dynamics driving residential segregation. The regionalization method used in this work enables the investigation of trends within cities. However, we believe that a meaningful analysis of within-city trends necessitates context-specific information, such as related to the local housing stock, historical urban development, and existing policies. Therefore, we recommend research investigating within-city trends to focus on selected case studies and incorporate contextual information.

Chapter 4

Misalignment between residential segregation of non-EU migrants and urban barriers varies across Western European cities

Abstract

In this chapter, we investigate whether the spatial structure of residential segregation aligns with physical barriers like railways, motorways, and waterways. Previous research has shown that such alignment exists in cities with past segregation policies in the US and South Africa, but this relationship remains untested in other contexts. Here we test this relationship in 520 cities across eight Western European countries, looking at segregation along non-EU migration background. We find no systematic alignment between residential segregation and urban barriers across European cities. We identify important regional variations, with cities in Germany and the Netherlands showing stronger alignment than other countries. These findings suggest that urban barriers do not generally function as separators of demographic groups in European contexts, with country-specific urban development potentially influencing the observed regional differences.

The remaining part of this study is organized as follows. Section 4.1 introduces the issue and reviews related work in other contexts. Section 4.2 shows no systematic alignment between residential segregation patterns and railways, motorways, and waterways; and assesses regional differences. Section 4.3 discusses the potential reasons related to the patterns observed. Section 4.4 presents the data and method used in this research.

4.1 Introduction

Residential segregation, defined as the uneven distribution of social groups in space, has been consistently linked to disparate outcomes in education, employment, and health

This chapter is based on the journal article: This chapter is based on the journal article: Spierenburg, L., Ralon-Santizo, E., van Cranenburgh, S., Cats, O. (in press). Misalignment between residential segregation of non-EU migrants and urban barriers varies across Western European cities. *Scientific reports*.

for disadvantaged populations (**Nieuwenhuis et al., 2020; Owens, 2018; Reardon & O'Sullivan, 2004; Tóth et al., 2021; van der Grefte & Fortuijn, 2017**). These neighbourhood effects suggest that the spatial concentration of social groups influences individual life chances through mechanisms such as social isolation, reduced access to social capital, and local environmental quality (**Chetty et al., 2016; G. C. Galster, 2011; Hedman & Van Ham, 2011**). Understanding the mechanisms that produce and perpetuate this phenomenon is therefore crucial for addressing its societal consequences. Research has long documented how residential segregation patterns emerge alongside urban development (**Ananat, 2011**). In particular, scholars have found that residential segregation often overlaps with urban fragmentation—i.e., the partition of urban space through urban barriers such as railways, motorways, and waterways—where urban barriers act as frontiers between social groups (**Iyer & Pryce, 2024; Mahajan, 2024; Tóth et al., 2021**).

4

The phenomenon that residential segregation aligns with urban fragmentation has been extensively demonstrated in the United States and South Africa, where urban barriers clearly separate social groups, epitomized by the common expression "wrong side of the tracks" (**Aiello et al., 2025; Ananat, 2011; Ballard & Hamann, 2021; Kramer, 2017, 2018; Noonan, 2005; Roberto & Korver-Glenn, 2021**). In these countries, alignment between residential segregation and urban fragmentation was policy-driven, with urban barriers serving as tools for enforcing social separation through explicit segregation policies like redlining in the US and apartheid in South Africa (**Chyn et al., 2022; Egede et al., 2023; Joshi et al., 2024; Maharaj, 2020**). Several studies suggest that alignment between residential segregation and urban fragmentation could also occur without explicit separation policies, from people's residential choice patterns (**Anciaes et al., 2016; Mitchell & Lee, 2014; Tóth et al., 2021**). Residential choices drive segregation through two primary mechanisms: Tiebout sorting and homophily. Tiebout sorting involves households 'voting with their feet', i.e. by selecting neighborhoods that offer a specific bundle of public goods, services, and housing costs (**Tiebout, 1956**). These preferences, as well as budget constraints, often align with demographics, and groups may sort into different urban zones (**Banzhaf & Walsh, 2013; Tronstine, 2020**). Since urban barriers like motorways or railways often delineate administrative or service boundaries, Tiebout sorting can result in a systematic alignment between residential segregation and physical fragmentation. Residential choices also drive segregation through homophily, i.e. the preference to be around people similar to oneself (**Schelling, 1971**). Importantly, this process does not require mutual preference for separation: if one group seeks distance from dissimilar others, the other group may be displaced through market mechanisms such as rising rents (**Hwang & Ding, 2020**). Segregation could then align with urban fragmentation organically under homophily, where urban barriers act as separators, one group moving out from one side of the barrier and moving in to the other side.

Beyond their role as physical dividers, urban barriers can also function as focal points for the formation of homogeneous demographic corridors. According to the

sector model of urban growth (**Hoyt, 1939**), residential patterns often emerge in wedges along transport infrastructure rather than in concentric circles. In this view, barriers are associated with externalities that influence land value on both sides of the infrastructure. For instance, railways and motorways often generate negative externalities such as noise and air pollution, which may concentrate disadvantaged populations (**Cohen & Coughlin, 2008; Theebe, 2004**). Conversely, waterways can represent positive environmental amenities that attract affluent groups to both banks (**Luttik, 2000; Schaeffer et al., 2016**). In both scenarios, the barrier acts as an attractor or repulsor for specific social groups, resulting in spatial configurations where social frontiers do not overlap with the physical infrastructure, but rather run parallel to it.

The alignment between residential segregation and urban fragmentation is documented in the US and South African contexts, primarily focusing on ethnicity and segregation policies. However, empirical evidence is lacking on whether this relationship is generalizable to Western Europe, where segregation typically occurs along income and migration background due to moving patterns and social housing allocation rather than explicit segregation policies. Addressing this gap involves testing whether such alignment can emerge in the European context, without segregation policies from organic residential preferences and market mechanisms. Practically, it would clarify for European policymakers whether removing urban barriers is an effective tool for reducing residential segregation along migration background.

We address this knowledge gap by quantifying how residential segregation patterns overlap with the fragmentation of the urban space across Europe. Our study encompasses 520 cities with populations exceeding 50,000 across eight Western European countries: Germany, France, Ireland, Italy, the Netherlands, Portugal, Spain, and the United Kingdom. We use the high-resolution gridded dataset of migrant populations developed by the EU's Joint Research Centre (**Alessandrini et al., 2017**). We focus specifically on the residential segregation of non-EU migrants and analyze whether residential segregation aligns with urban fragmentation more than would be expected by chance.¹ Our methodological framework employs a Monte Carlo approach, in which we generate 200 synthetic fragmentation patterns for each city and compare the actual spatial overlap against this null distribution. Figure 4.1 summarizes our analysis approach. We identify demographically homogeneous regions using a regionalization method that aggregates grid cells based on demographic data (**Spierenburg, van Cranenburgh & Cats, 2024**). Urban fragments are defined as contiguous groups of cells bounded by railways, motorways, waterways, or uninhabited areas. We test the statistical significance of the overlap between the demographic partition and the fragmentation partition by comparing the observed value against 200 synthetic fragmentation partitions.

Our results suggest that the role of urban barriers as social frontiers, a phenomenon well-documented in contexts with explicit segregationist policies, does not necessarily generalize to the context of migration-based segregation in Western European.

¹The United Kingdom was part of the European Union in 2011, the time at which the data was produced.

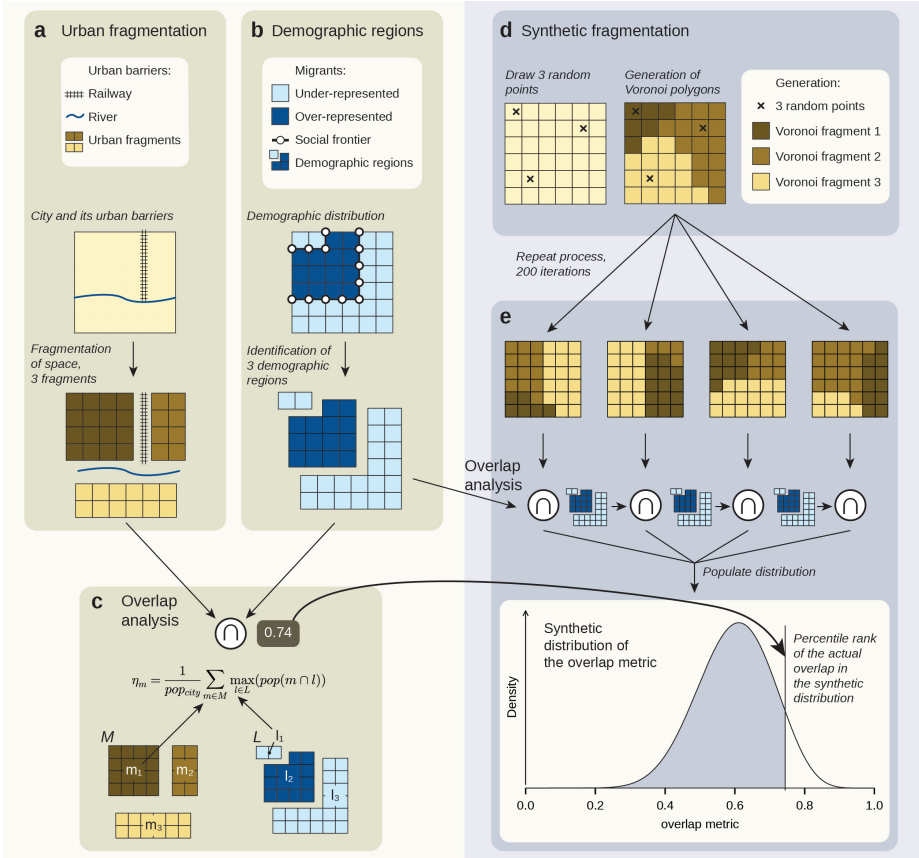


Figure 4.1 Quantifying the statistical significance of the overlap between residential segregation and urban fragmentation at the city level. (a) Fragmentation of urban space by urban barriers into three fragments. (b) Spatial partitioning of space into demographically homogeneous regions. (c) Quantifying the overlap between residential segregation and urban fragmentation, using the purity score. (d) Generating 200 synthetic urban fragmentation partitions, using Voronoi polygons. (e) Computing the overlap between residential segregation and each synthetic fragmentation partition, generating the statistical distribution and evaluating the quantile rank of the actual overlap in the distribution of the synthetic overlaps.

Specifically, we find that there is no uniform trend in the alignment between residential segregation patterns and urban fragmentation across European cities. Instead, our results reveal significant spatial heterogeneity: while a subset of cities exhibits alignment, a statistically significant number of cities—particularly in Spain, Italy, and the United Kingdom—show less alignment than would be expected by chance. These regional variations are further confirmed by our regression analysis, which shows that cities in Germany and the Netherlands exhibit significantly stronger alignment between barriers

and segregation compared to other countries in our sample, even after controlling for city-specific characteristics. These findings suggest that the relationship between urban barriers and social frontiers is neither universal nor random, but rather influenced by country-specific urban development patterns. Future research should further investigate the mechanisms behind this regional divergence, particularly how urban barriers may serve as dividers of distinct development and housing units in some contexts while acting as demographic corridors in others.

4.2 Results

Our analysis of 520 cities across eight European countries reveals two key findings regarding the spatial relationship between residential segregation patterns and urban fragmentation. First, contrary to expectations based on findings reported for North American cities, we do not observe a uniform alignment between residential segregation and urban fragmentation across European cities (subsection 4.2.1). Second, the cities that do exhibit significant alignment are predominantly concentrated in the Netherlands and Germany, suggesting important regional variations in this relationship (subsection 4.2.2).

4.2.1 Absence of a uniform alignment between residential segregation and urban fragmentation across European cities

For each city in our dataset, we measure the alignment between residential segregation and urban fragmentation, and assess its statistical significance by assessing how the metric ranks in 200 Monte Carlo simulations. If the metric ranks higher than 85% of the simulations, the quantile rank is 0.85. To assess the extent to which residential segregation aligns with urban barriers in the European context, we investigate the statistical distribution of quantile ranks for all 520 urban areas in our sample (Figure 2a). In this distribution, the x-axis represents the quantile rank of each city, while the y-axis shows the frequency of cities falling into each rank. Here, urban barriers refer to railways, motorways, waterways, and uninhabited areas that fragment urban space, while social frontiers denote the boundaries between demographically distinct regions. Under the null distribution, which asserts that there is no relation between residential segregation and urban fragmentation, each bin has approximately the same number of observations, i.e. for instance 5% of the cities should have a quantile rank higher than 0.95. If residential segregation aligns with urban fragmentation across European cities, we would expect an over-representation of cities in the right portion of the histogram, particularly in the highest quantile rank bin (0.95-1.00), indicating that urban barriers align with social frontiers to a degree unlikely to occur by chance. The observed distribution reveals a pattern that is predominantly uniform with a slight L-shape, characterized by higher counts in the first (0.00-0.05) bin compared to the rest (Figure 4.2 a). Importantly, in the vast majority of the cities that we investigate ($N=486$, 93.5%), residential segregation patterns show no significant alignment to urban fragmentation, with their quantile rank falling outside the top 5% of the synthetic distribution.

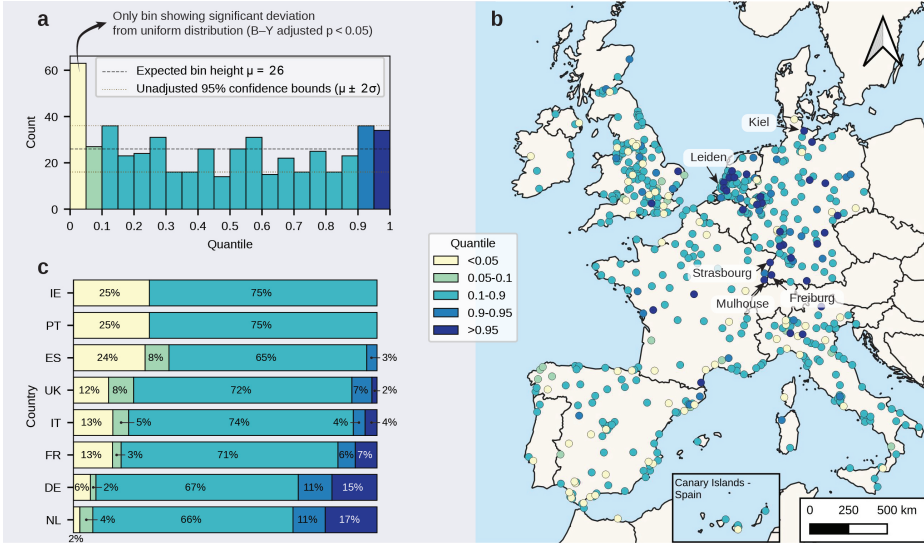


Figure 4.2 Statistical and geographical patterns of alignment between residential segregation and urban fragmentation across European cities. (a) Distribution of quantile ranks across all cities in the dataset. Each bar shows the number of cities falling within a specific quantile rank bin, while the dashes horizontal line indicates the expected count per bin under a null scenario where the alignment between residential segregation patterns and urban fragmentation is not substantially different from what would be expected by chance (no systematic relationship exists between residential segregation and urban fragmentation). The dotted horizontal lines indicate significance thresholds without applying the Benjamini-Yekutieli correction. This correction accounts for the inherent dependence structure of the data. (b) Geographic distribution of quantile ranks across cities in the dataset, cities mentioned in the paper are annotated on the map. (c) Distribution of quantile ranks across countries. Map produced with QGIS 3.34.1 and edited with Adobe Illustrator 2025.

To assess whether the observed distribution pattern deviates significantly from what would be expected under random chance, we use a formal statistical testing framework that accounts for multiple comparisons using **Benjamini and Yekutieli, 2001** adjusted p-values (Figure 4.2 a). Our statistical test reveals that only the lowest quantile rank bin (0.00-0.05) shows a statistically significant deviation from expected counts, with cities being overrepresented in this category. This finding suggests a pattern that is opposite to the expected alignment between urban barriers and social frontiers, with a small subset of European cities exhibiting less alignment between urban barriers and social frontiers than what would be expected by chance. This result is contrary to the expectation that urban barriers delineate segregation patterns across European urban contexts, highlighting a fundamental difference in patterns observed in their North American counterparts.

To shed more light on this result, we showcase the results from two cities that exhibit clear disagreement between residential segregation and urban fragmentation: Lyon

(France) and Birmingham (United Kingdom), shown in figure 4.3. In both cities, we observe distinct spatial patterns in the demographic distribution. While Birmingham exhibits a center-periphery pattern with migrant populations being overrepresented in the center (figure 4.3 a), Lyon displays a pronounced East-West division (figure 4.3 b), with migrants predominantly residing in eastern districts. Despite the presence of clear social frontiers, our analysis reveals minimal alignment with urban barriers in both contexts. In Lyon, the *boulevard périphérique* (D383) follows a similar East-West orientation as the social frontier; nevertheless, rather than marking the frontier between demographic regions, it runs through the core of the area over-representing migrants, functioning more as a "spine" for this region rather than as a dividing barrier. Similarly, Birmingham's center contains dense infrastructure forming a central node in a radial network, yet these physical elements do not delineate social frontiers between areas over-representing migrants and the rest of the city. These two cities are illustrations of the statistical pattern observed across our sample, demonstrating that even in cities with pronounced segregation patterns, urban barriers do not necessarily systematically define social frontiers in the European context.

4

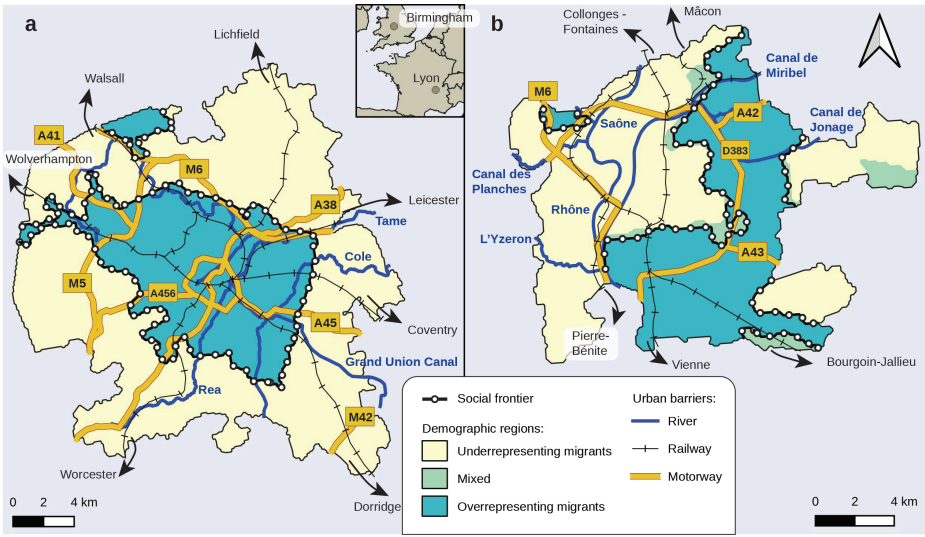


Figure 4.3 Spatial layout of social frontiers and urban barriers in Birmingham, United Kingdom (a) and in Lyon, France (b). Selected highways, rivers and railways are displayed to enhance readability. Map produced with QGIS 3.34.1 and edited with Adobe Illustrator 2025.

4.2.2 Greater association between residential segregation and urban fragmentation in Germany and the Netherlands than in other countries

We identify geographical differences in the extent to which social frontiers align with urban barriers throughout Europe (Figure 4.2). While we do not observe a consistent alignment between residential segregation patterns and the urban fragmentation phenomenon in our sample, urban areas exhibiting such alignment are predominantly concentrated in Germany and the Netherlands. There exists stronger alignment between residential segregation patterns and urban fragmentation in a subset of cities in Germany (15 % of cities) and the Netherlands (17 % of cities). This pattern suggests that the relationship between urban barriers and social frontiers may be influenced by country-specific urban development strategies. Yet, even in Germany and the Netherlands, cases exhibiting significant correspondence between urban barriers and demographic frontiers constitute a minority of cities in these countries. Interestingly, the French region of Alsace, where cities like Strasbourg and Mulhouse were administered by Germany between 1871 and 1918, also shows significant alignment between urban fragments and demographic regions.

To formally assess the potential significance of these country-specific effects, we conduct a beta regression analysis where the dependent variable is the quantile rank of a city's purity score. This model allows us to control for two city-level variables that might explain these geographical differences (**Ferrari & Cribari-Neto, 2004**). Specifically, we aim to measure whether higher levels of segregation or decentralization are positively associated with the alignment between residential segregation patterns and urban barrier. We control for segregation intensity using the dissimilarity index (subsection 4.4.3), as the alignment between segregation patterns and urban fragmentation might be less observable in cities where segregation is less pronounced. We also control for decentralization, hypothesizing that multicentric urban structures may be more likely to exhibit barrier-frontier alignment because physical barriers can serve as natural separators between distinct development centers, whereas monocentric cities with radial growth patterns tend to extend across barriers rather than being bounded by them. We use a normalized average distance from city center for measuring decentralization (see subsection 4.4.3). Figure 4.4 displays the distribution of segregation intensity and decentralization across the countries in our sample. Spain and Italy exhibit the highest levels of segregation intensity, while Portugal and France show the lowest, consistent with findings from **Benassi et al., 2020** using the same data and dissimilarity indicator. Germany and the Netherlands exhibit moderate levels of segregation intensity relative to other countries in the sample. The decentralization indicator, on the contrary, does show a similar geographic pattern to that of the alignment between physical barriers and social frontiers, with Germany and the Netherlands showing the highest levels of spatial decentralization, while Spain and Portugal display the most centralized urban structures.

The results of the regression are presented in table 4.1. The regression model estimation results show that Dutch cities have a significant positive association (average

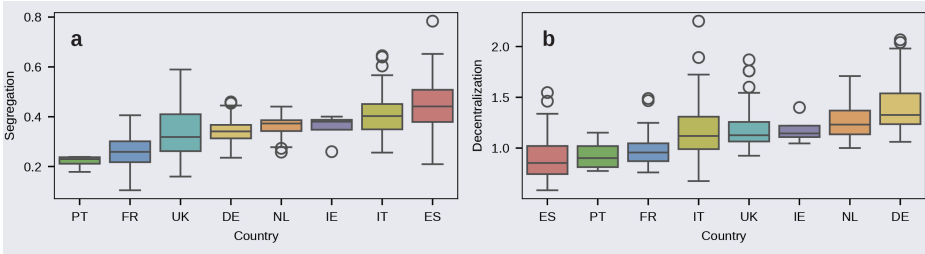


Figure 4.4 Cross-country variation in residential segregation intensity and urban spatial structure. (a) Distribution of residential segregation across urban areas per country, measured using the dissimilarity index. (b) Distribution of the decentralization indicator across urban areas per country. The countries are arranged from left to right by ascending median.

4

quantile rank of 0.61 $p=0.008$), even after controlling for decentralization, while Germany’s association approaches but does not reach the 5% significance threshold (average quantile rank of 0.56 $p=0.064$). In contrast, we observe significant negative association in Spain (average quantile rank of 0.34 $p<0.001$), the United Kingdom (average quantile rank of 0.41 $p=0.001$), and Italy (average quantile rank of 0.43 $p=0.03$), reinforcing our earlier findings. These geographic disparities suggest systematic country-level differences besides decentralization in how residential segregation patterns relate to urban barriers.

	coef (logit scale)	p-value	Predicted average
Segregation	-0.1042	0.869	-
Decentralization	0.5759	0.022	-
Portugal	-0.9711	0.093	0.27
Spain	-0.6844	0.000	0.34
Ireland	-0.5049	0.399	0.38
the United Kingdom	-0.3605	0.001	0.41
Italy	-0.2979	0.03	0.43
France	-0.0645	0.687	0.48
Germany	0.2393	0.064	0.56
the Netherlands	0.4567	0.008	0.61

Table 4.1 Regression model estimation results. The model includes country-specific fixed effects, allowing us to identify countries where the alignment is significantly different from the null distribution. A positive coefficient implies a positive association of the variable with the quantile rank. The variables for segregation and decentralization are centered before the regression, subtracting their respective mean. The average predicted quantile rank per country is measured using the transformation $\text{logit}^{-1}(\alpha_c)$ and provided in the last column of the table.

Examining our control variables, we find that segregation intensity is not statistically significant, indicating that we cannot attribute the lack of observed alignment to an

insufficient level of segregation. Conversely, decentralization shows a significant positive effect ($p=0.023$), suggesting that the spatial organization of cities plays an important role in the alignment between urban barriers and social frontiers. This finding aligns with the distinction between monocentric cities prevalent in countries like Spain and France versus the more multicentric urban agglomeration structures characteristic of Germany and the Netherlands (figure 4.4). The latter configuration may facilitate a more systematic division of space compared to the predominantly radial development patterns of monocentric cities.

To illustrate our statistical findings with concrete examples, we conduct a qualitative analysis of Amsterdam (Figure 4.5). Amsterdam is one of the cities exhibiting the highest quantile rank values (quantile rank = 1) in our dataset, indicating exceptional alignment between residential segregation patterns and urban fragmentation. In Amsterdam, five main regions have an over-representation of migrants: Nieuw-West, Noord, Oost, along the Beneluxbaan, and Zuidoost. Crucially, these regions are delineated with remarkable correspondence to the city's waterways. In the western part of the city, a succession of canals (from Westlandgracht to Westelijk Marktkanaal) clearly separates a region over-representing migrants in the West from an area under-representing migrants in the East of the canals (figure 4.5 b). Similarly, in Amsterdam Noord (figure 4.5 c), the major canal of Ringsloot creates an almost perfect boundary between the northern area over-representing migrants and the southern area under-representing migrants. Similar striking overlap (quantile rank > 0.95) occurs in several Dutch and German cities, including Leiden (quantile rank = 0.97), Freiburg (quantile rank = 0.97), and Kiel (quantile rank > 0.995) (figure 4.2).

4.3 Discussion

Our analysis of 520 European cities demonstrates that the alignment between urban fragmentation and residential segregation is not homogeneous in the European context, but instead varies across national urban development patterns. This finding represents a significant departure from previous research, which has largely focused on contexts where explicit segregation policies were historically implemented, such as cities affected by redlining in the United States and the apartheid in South Africa (**Grannis, 2009; Joshi et al., 2024; Maharaj, 2020; Noonan, 2005; Roberto & Hwang, 2017**). Our findings demonstrate that the same pattern does not always emerge when such explicit policies are absent. While households may indeed seek separation and sort themselves into homogeneous neighborhoods, this sorting does not exhibit a consistent relationship with urban barriers across Europe, with some regions showing significant alignment while others show a marked lack of it.

In fact, we observe the opposite in a significant number of cities where residential segregation aligns less with urban fragmentation than expected under the null hypothesis. This suggests that when groups sort residentially near urban barriers, other factors may outweigh homophilic preferences for separation from dissimilar groups. Urban barriers

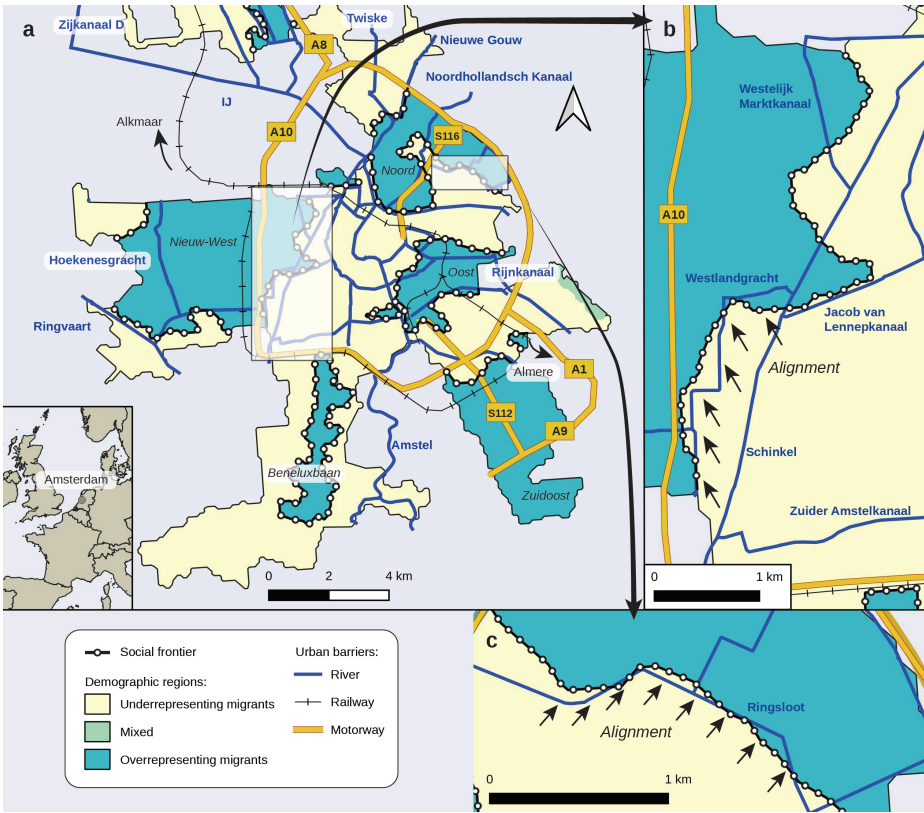


Figure 4.5 Spatial layout of social frontiers and urban barriers in Amsterdam, the Netherlands. (a) Entire map of Amsterdam. (b) Zoom over Amsterdam West. (c) Zoom over Ringsloot in Amsterdam Noord. Selected highways, rivers and railways are displayed to enhance readability. Map produced with QGIS 3.34.1 and edited with Adobe Illustrator 2025.

may generate effects that either attract or repel certain population groups. For instance, areas adjacent to motorways experience higher noise and air pollution on both sides, making these locations less desirable regardless of which side of the barrier they occupy, thus potentially concentrating disadvantaged populations throughout the corridor rather than separating different groups. Conversely, rivers with scenic waterfront views may attract advantaged populations to both banks, creating concentrations of similar groups on either side of the barrier. In both cases, urban barriers create similar conditions on both sides, and these attraction or repulsion effects may have a stronger influence on residential choice than any preference to use barriers as a way to separate from dissimilar groups.

Our methodological approach to generating synthetic urban partitions represents a deliberate trade-off between realism and statistical rigor that directly impacts our results.

The way synthetic partitions are defined affects the distribution of synthetic overlap scores against which we compare observed patterns. Creating more realistic or plausible synthetic partitions would require incorporating additional spatial rules and assumptions, which could also introduce potential bias through circular reasoning. Assumptions about plausible urban layouts could artificially produce the alignment (or misalignment) patterns we seek to detect. Therefore, we chose a plain approach that controls only for the statistical properties directly affecting our overlap metric—specifically, the number of urban fragments and their size distribution, since more evenly distributed fragment sizes naturally yield better overlap scores. By avoiding additional spatial assumptions, we essentially treat each city as a blank canvas and redraw urban frontiers randomly while preserving only the size characteristics of the actual fragmentation pattern.

4

We acknowledge that our relative classification of demographic regions is designed to identify spatial discontinuities rather than absolute population thresholds. A region overrepresenting migrants in a city with a 2% migrant share differs substantively from such a region in a city with a 20% share. However, for the purpose of testing whether urban barriers function as social frontiers, this relative approach is appropriate. We are testing whether physical infrastructure aligns with the sorting process of a city, regardless of its absolute demographic composition, to see if barriers serve as the 'hinge' points where a city's demographic distribution shifts.

4.3.1 Limitations

Several limitations of our study should be acknowledged, particularly regarding the social categories used to measure sorting. While the North American and South African literature focuses predominantly on race as the primary axis of social distance, the generative processes of residential segregation in the European context often revolve around a combination of migration background and income (**Musterd et al., 2017**). Our study focuses on the former, as non-EU migration background (which includes native-born individuals with non-EU parentage) represents a key dimension of residential segregation in European cities. However, this category is internally diverse, conflating groups with different socio-economic characteristics. The lack of standardized, high-resolution income data across the 520 cities in our sample prevents further disaggregation of the results. Future research may investigate whether income-based segregation follows a similar or rather a distinct spatial logic which may align differently with urban barriers than the migration-based patterns observed here.

Moreover, despite efforts to build standardized data across countries, notable methodological differences remain that could influence our results (**Alessandrini et al., 2017**). These differences include variations in spatial resolution, where some countries like the Netherlands provide data natively at 100×100m resolution, while others, such as France, require upscaling from coarser administrative zones (approximately 300×300m) that were subsequently disaggregated into smaller grid cells. From a statistical perspective, larger original zones tend to have migrant shares closer to the city average due to their larger sample sizes, which reduces the extreme values that drive

segregation patterns. Additionally, the operational definition of migration background varies significantly across countries: for instance, a person born in the Netherlands to parents from outside the EU who acquires Dutch citizenship retains their migration background classification, whereas in France, the same individual would no longer be classified as having a migration background upon acquiring French nationality as an adult. These differences in definition and resolution directly impact measurements of segregation intensity. Our alignment analysis is less sensitive to these variations since we focus on spatial layout patterns rather than absolute migrant shares, and the demographic regions we identify are typically orders of magnitude larger than even the coarsest spatial resolution in our dataset. Finally, our data is from 2011, and segregation patterns may have changed in the subsequent years, particularly given the significant migration flows to Europe since the mid-2010s.

4.3.2 Recommendations for further research

The underlying mechanisms driving the country-specific patterns we observe remain unclear and warrant further investigation in future research. In particular, understanding why Dutch and German cities exhibit stronger alignment between urban barriers and social frontiers compared to other European countries could provide valuable insights into how urban planning traditions and infrastructure development influence residential segregation patterns. To uncover these mechanisms, future research could investigate other spatial features that might overlap with segregation patterns in European cities. For instance, urban barriers might delineate housing development patterns, where plots of land bounded by railways, motorways, or waterways were developed with distinct housing types or at different time periods, potentially creating areas with varying affordability and accessibility that indirectly influence residential sorting patterns. The methodological framework developed in this study—comparing observed spatial overlap against synthetic null distributions—could be readily adapted to assess alignment between demographic patterns and partitions based on housing quality, concentration of social housing, building age, or access to amenities. Such approaches might reveal more consistent patterns that explain the spatial distribution of social groups in European cities, where segregation clearly exists but follows a different spatial logic than the barrier-bounded patterns observed in North American contexts.

4.4 Methods

Our study investigates whether residential segregation patterns align with urban fragmentation across European cities. We employ a Monte Carlo approach to test the statistical significance of the observed spatial overlap between demographic regions and urban fragments defined by urban barriers. For each of the 520 cities in our sample, we generate 200 synthetic fragmentation patterns and compare the actual spatial alignment against this null distribution. This methodology allows us to determine whether urban barriers act as social frontiers more than would be expected by chance.

4.4.1 Data sources and preparation

Our analysis requires two datasets: demographic data to identify patterns of residential segregation and geographic data to define urban barriers. The following subsections detail our data sources and the preparation steps we undertook to create standardized, comparable spatial partitions of residential segregation and urban fragmentation across all 520 cities in our sample.

Demographic dataset

This study used a dataset mapping the spatial distribution of migrants across urban areas in eight European countries (France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain, and the UK) derived from harmonized 2011 Census and registry data (Alessandrini et al., 2017). The dataset was created by the Joint Research Centre (JRC) of the European Commission by processing ad hoc extractions from National Statistical Institutes, organizing population data by citizenship and/or country of birth into a uniform spatial grid with cells of 100×100 meters. The dataset is called *Data for Integration (D4I)* and is available upon request at https://knowledge4policy.ec.europa.eu/migration-demography/data-integration-d4i_en. The population was classified into two categories: migrants from outside the EU and the rest (including non-migrants and EU migrants). The definition of migrants varied across countries, based on citizenship criteria in Italy and France, and on country of birth in the remaining countries (Germany, Ireland, the Netherlands, Portugal, Spain, and the UK). All methods were carried out in accordance with relevant guidelines and regulations.

Study area

We define the spatial extent of urban areas through population density analysis. First, we calculate the average population density within a 400-meter radius of each cell, applying a smoothing function that weighs nearby populations inversely to the square of the distance. The 400-meter radius was selected based on a qualitative assessment of different smoothing parameters to create cohesive urban areas while preserving meaningful local variation in density. Cells exceeding a density threshold of 1,000 people per km^2 are classified as high-density areas. This conservative threshold accounts for edge effects, which cause the moving average calculation to yield densities approximately half those of central areas in peripheral cells. Adjacent high-density cells are then aggregated into contiguous urban areas, and only areas with a total population exceeding 50,000 inhabitants are retained, consistent with the established definition of urban cores in the literature (Dijkstra et al., 2019). We define the study area using a convex hull around each urban area to ensure we capture any urban barriers that might exist at the periphery of settlements. This methodological approach ensures our analysis focuses on densely populated urban environments while maintaining consistency across the 520 cities in our sample.

Regionalization method for constructing demographic regions

We identify homogeneous demographic regions using the regionalization method developed in **Spierenburg, van Cranenburgh and Cats, 2024**. This method applies a spatial moving average to migrant proportion data to filter out small-scale variations. It then uses spatially-constrained agglomerative clustering, which ensures only adjacent cells merge into regions, maintaining spatial contiguity throughout. To classify these regions, we measure the extent to which the region-wide mean of the spatially-smoothed migrant proportion $\overline{\rho_R}$ differs from the city average μ . This mean is calculated by averaging the cell-level spatial moving averages within each region, weighted by the cells' population. We evaluate this difference using the Z-score in equation 4.1, where $\sigma_{\overline{\rho_R}}$ represents the theoretical standard deviation of $\overline{\rho_R}$ under a random spatial allocation of groups in space. Following **Spierenburg, van Cranenburgh and Cats, 2024**, we apply a $2\sigma_{\overline{\rho_R}}$ threshold (corresponding to a threshold $p < 0.05$): regions exceeding $2\sigma_{\overline{\rho_R}}$ are labeled as *overrepresenting migrants*, those below $-2\sigma_{\overline{\rho_R}}$ as *underrepresenting migrants*, and the remainder as *mixed*. The result is a set of contiguous demographic regions for each city, with each region representing an area of relatively homogeneous migrant population characteristics, enabling direct comparison with urban fragments defined by urban barriers.

$$\text{A region is } \begin{cases} \text{underrepresenting migrants} & \text{if } \frac{\overline{\rho_R} - \mu}{\sigma_{\overline{\rho_R}}} \leq -2 \\ \text{mixed} & \text{if } -2 < \frac{\overline{\rho_R} - \mu}{\sigma_{\overline{\rho_R}}} < 2 \\ \text{overrepresenting migrants} & \text{if } \frac{\overline{\rho_R} - \mu}{\sigma_{\overline{\rho_R}}} \geq 2 \end{cases} \quad (4.1)$$

Identification of urban barriers and creation of urban fragments

We extract urban barriers from the **OpenStreetMap, 2021** database, which include railways, motorways, and waterway. We then systematically partition the demographic grid: each contiguous group of cells that is fully separated from other groups by one or more urban barriers constitutes an urban fragment. This fragmentation process creates a spatial partition that we then compare against the independently derived demographic regions to assess potential alignment between urban barriers and social frontiers.

4.4.2 Quantifying alignment between social frontiers and urban barriers

To assess whether residential segregation patterns align with urban fragmentation beyond what would be expected by chance, we employ a two-step approach. First, we develop a purity score metric to quantify the spatial overlap between demographic regions and urban fragments within each city. Second, we establish statistical significance by comparing observed alignment scores against a null distribution generated through Monte Carlo simulations using synthetic urban partitions.

Measuring the alignment between two spatial partitions using the purity score

We measure the alignment between residential segregation patterns and urban fragmentation using a purity score metric η (equation 4.2). For each urban fragment m (constructed in subsection 4.4.1) in the set M of all urban fragments, we identify the demographic region l (constructed in subsection 4.4.1) in the set of all demographic regions L that contains the largest share of the fragment's population. The overall purity score sums these maximum population overlaps across all fragments M , normalized by the total population pop_{city} .

$$\eta_m = \frac{1}{pop_{city}} \sum_{m \in M} \max_{l \in L} (pop(m \cap l)) \quad (4.2)$$

4

The purity score can be interpreted as a measure indicating how well one can reconstruct demographic regions using urban fragments as building blocks. It ranges from 0 to 1, with 1 implying that demographic regions can be perfectly reconstructed from urban fragments. This metric is calculated for each city as well as for each of the synthetic partitions.

Generation of synthetic partitions

We generate synthetic urban partitions by randomly placing seed points within the city area, matching the number of fragments observed in the actual urban structure. We then apply Voronoi tessellation, assigning each population grid cell to its nearest seed to form distinct urban fragments. We use transformed Gaussian random fields to draw these random points, providing flexible control over spatial patterns. A Gaussian random field $Z(x, y)$ assigns random values to spatial locations (x, y) , where any finite collection of field values follows a multivariate normal distribution. The field is characterized by its covariance function:

$$\text{Cov}[Z(x_k, y_k), Z(x_{k'}, y_{k'})] = \sigma^2 \exp\left(-\frac{h^2}{\lambda^2}\right) \quad (4.3)$$

$$P(x) = \frac{[Z(x, y) + |\min(Z)|]^\zeta}{\sum_k [Z(x_k, y_k) + |\min(Z)|]^\zeta} \quad (4.4)$$

where h is the distance between points, λ is the length scale parameter controlling spatial correlation, and ζ is the power exponent. Equation 4.3 defines how field values $Z(x, y)$ at different locations covary—larger λ values produce smoother, more clustered patterns while smaller λ values create more dispersed arrangements. The power transformation in equation 4.4 converts the field into probabilities for seed placement, where higher ζ values increase clustering around field maxima.

To ensure that synthetic structures accurately match the size distribution of actual urban fragments, we implement a two-stage process as follows.

First, we generate 500 fields with varying λ and ζ and construct corresponding synthetic partitions, optimizing the length scale controlling spatial autocorrelation (equation 4.3) and the power exponent regulating point concentration (equation 4.4). Their optimal values are determined by minimizing the Wasserstein distance between synthetic and observed fragment size distributions (equation 4.5). We measure the Wasserstein distance W_1 by ranking fragments by descending size, constructing the cumulative distribution function of fragment size in the actual case A and the synthetic case S , and summing the difference between the two over every rank r .

$$W_1(A, S) = \sum_{r \in R} |CDF_A(r) - CDF_S(r)| \quad (4.5)$$

Second, once optimal parameters are identified, we generate 500 synthetic partitions using these optimized parameters and select the top 200 partitions with the smallest Wasserstein distances to the actual fragmentation partition. This approach creates multiple statistically equivalent alternatives that preserve the size distribution of the actual urban fragmentation partition.

Statistical significance of alignment between residential segregation and urban fragmentation

We test the statistical significance of the alignment between the demographic partition and the fragmentation partition by comparing the observed purity score against the distribution of scores from 200 synthetic fragmentation partitions. For each city, we calculate the quantile rank of the actual purity score within the distribution of synthetic purity scores, which provides a standardized measure of statistical significance at the city level.

The quantile rank q for city j is computed as:

$$q_j = \frac{|s \in S_j : \eta_s < \eta_{observed,j}|}{|S_j|} \quad (4.6)$$

where $\pi_{observed,c}$ represents the observed purity score for city j , S_j denotes the set of 200 synthetic purity scores for that city, and $|\cdot|$ indicates set cardinality. This quantile rank ranges from 0 to 1, with values approaching 1 indicating that the observed alignment is exceptionally high compared to what would be expected under random spatial arrangements of urban fragments.

A quantile rank above 0.95 suggests that the observed alignment between residential segregation and urban fragmentation occurs in fewer than 5 % of random scenarios, indicating statistically significant alignment at the $\alpha = 0.05$ level. Conversely, quantile ranks below 0.05 indicate significantly less alignment than expected by chance, suggesting that urban barriers actively disagree with social frontiers.

This quantile-based approach provides several methodological advantages: it naturally accounts for the varying number and size distribution of urban fragments across cities, it makes no distributional assumptions about the purity scores, and it provides an intuitive interpretation where the quantile rank directly corresponds to the probability that a randomly generated urban fragmentation would yield lower alignment than the observed pattern.

4.4.3 Post-hoc statistical testing

After conducting our Monte Carlo analysis to assess alignment between residential segregation and urban fragmentation in individual cities, we perform two additional statistical procedures to analyze patterns across our entire sample. First, we test whether the distribution of quantile ranks across all cities deviates significantly from what would be expected under the null hypothesis of no systematic relationship (subsection 4.4.3). Second, we examine country-specific effects on alignment patterns while controlling for city-level characteristics that might influence this relationship (subsection 4.4.3).

Benjamini-Yekutieli test

In section 4.2.1, we examine which bins deviate from what we would expect under the null hypothesis of uniformity. A key challenge in this analysis is the dependence between bins: if one bin has significantly more observations than expected, other bins have, by construction, fewer observations. The **Benjamini and Yekutieli, 2001** procedure handles such dependencies between tests. This procedure controls the false discovery rate (*FDR*)—the expected proportion of false positives among all rejected null hypotheses—when performing multiple hypothesis tests, which we set to $FDR = 0.05$ in this study. This approach provides a statistical framework that accounts for the inherent dependency in our quantile rank distribution analysis. The implementation involves five steps:

1. Conducting Chi-Square tests comparing observed versus expected counts for each bin b , and computing the p-values per bin p_b .
2. Ordering the resulting p-values from smallest to largest: $p^{(1)} \leq p^{(2)} \leq \dots \leq p^{(20)}$, where κ denotes the rank
3. Calculating Benjamini-Yekutieli critical values τ_κ (see equation 4.7)
4. Finding the largest $\hat{\kappa}$ where $p^{(\hat{\kappa})} \leq \tau_{\hat{\kappa}}$
5. Rejecting the null hypothesis for all bins with p-values $\leq p^{(\hat{\kappa})}$

$$\tau_\kappa = \frac{\kappa}{20} \times \frac{FDR}{c} \quad \text{where} \quad c = \sum_{\kappa'=1}^{20} \frac{1}{\kappa'} \quad (4.7)$$

Beta regression

To measure the absolute effect of each country on the alignment between urban barriers and social frontiers, we employ beta regression analysis while controlling for city-level characteristics that might influence this relationship. Beta regression is particularly appropriate for our analysis because it is specifically designed for continuous response variables bounded between 0 and 1, exactly matching the properties of our quantile ranks (**Ferrari & Cribari-Neto, 2004**). This regression framework allows us to test whether country-specific effects remain statistically significant after accounting for segregation levels and decentralization patterns. Equation 4.9 provides the regression specifications, where Y_{jc} represents the response variable (quantile rank) for city j in country c . The variable follows a beta distribution with mean μ_j and precision parameter ϕ . We use logit as the link function $g(\cdot)$ connecting the linear predictor $\mathbf{x}_j^T \boldsymbol{\beta}$ to the mean μ_j , which ensures predicted values remain within the (0,1) interval while maintaining linearity on the transformed scale.

$$y_{jc} \sim \text{Beta}(\mu_{jc}, \phi) \quad (4.8)$$

$$\text{logit}(\mu_{jc}) = \alpha_c + \beta_D \cdot D_{jc} + \beta_\delta \cdot \delta_{jc} + \varepsilon_{jc} \quad (4.9)$$

We control for segregation intensity and decentralization, denoted respectively by D and δ in equation 4.9. Both control variables are centered around their respective means to enable meaningful coefficient interpretation, since baseline scenarios with zero decentralization or segregation are neither realistic nor meaningful in urban contexts.

We measure the intensity of residential segregation in an urban area using the dissimilarity index (**Duncan & Duncan, 1955**). This index noted D is measured using equation 4.10 where u_k and v_k represent respectively the migrant and the non-migrant population in unit k , while the total migrant and non-migrant population in the urban area are noted U and V respectively. This index ranges from 0 (no segregation) to 1 (complete segregation), providing a standardized measure of spatial separation. We use this established measure to maintain consistency with prior studies employing the same dataset (**Benassi et al., 2020, 2023; Marcińczak et al., 2023**).

$$D = \frac{1}{2} \sum_k \left| \frac{u_k}{U} - \frac{v_k}{V} \right| \quad (4.10)$$

We also control for decentralization, which captures the fundamental spatial organization of urban development. Cities can be organized along a spectrum from monocentric structures—where development radiates outward from a single dominant center—to multicentric structures—where development is distributed across multiple distinct centers or nodes. The alignment between segregation and urban fragmentation could be partly due to decentralization. In multicentric cities, subcenters may be spatially separated by uninhabited areas that fragment space, and different demographic

groups may be attracted to different subcenters based on factors like real estate prices or local amenities. The decentralization is computed using the average distance from any cell k located at position (x_k, y_k) to the city's center of gravity (x_g, y_g) weighted by the cell's population pop_k (see equation 4.11). The city's center of gravity is also population-weighted. This indicator is linearly correlated to the city size. To enable cross-city comparisons, we normalize this measure by dividing it by the theoretical average distance from the city center ADC_{circle} of a perfectly circular city with the same area S (equation 4.12). The normalization in equation 4.13 yields the decentralization indicator δ , a dimensionless metric that allows for meaningful comparison of decentralization patterns across urban areas of different sizes and shapes.

4

$$ADC = \frac{\sum_k \sqrt{(x_k - x_g)^2 + (y_k - y_g)^2} \cdot pop_k}{\sum_k pop_k} \quad (4.11)$$

$$ADC_{circle} = \frac{2}{3} \cdot \sqrt{\frac{S}{\pi}} \quad (4.12)$$

$$\delta = \frac{ADC}{ADC_{circle}} \quad (4.13)$$

The country fixed effects are denoted by α_c . Our specification excludes an intercept term and retains all country fixed effects, allowing us to assess whether each country exhibits positive or negative effects on the alignment phenomenon rather than comparing countries against a single reference.

Chapter 5

Studying riots through the lens of social media

Abstract

This chapter examines riots. In contrast to residential segregation's latent, chronic patterns studied in previous chapters, riots represent explicit, acute events that erupt into the open. Despite their importance, our empirical understanding of how riots actually unfold remains limited by coarse spatiotemporal analysis (municipality-level, day-by-day). This prevents researchers from characterizing the fine-scale spatiotemporal dynamics of riot episodes—how incidents are distributed across space and time within urban areas. This empirical gap persists because traditional data sources lack fine-grained resolution, and researchers have been unable to systematically process unstructured visual content from live-reporting social media platforms. In this chapter, we address both challenges by developing a computer vision approach to identify riot-related content from geolocated Snapchat posts. Applying this method to map riot incidents at unprecedented resolution reveals that riots are characterized by highly localized, short-lived events that unfold at the neighborhood scale over hours—demonstrating the bursty, concentrated spatiotemporal nature of these phenomena and opening new avenues for understanding riot dynamics.

The chapter is organized as follows. Section 5.1 introduces the issue and review methods used to study riots. Section 5.2 presents the 2023 Nahel Merzouk riots case study and describes the datasets: geolocated Snapchat posts and press reports. Section 5.3 develops the methodology, including the computer vision model for classifying riot-related visual content (architecture, training, threshold optimization) and spatiotemporal analytical tools (Ripley's K function for burst analysis and density-based clustering for event detection). Section 5.4 validates the computer vision approach through quantitative metrics, qualitative assessment, and cross-validation against press data, then demonstrates fine-grained spatiotemporal mapping of riot evolution across seven French cities. Section 5.5 discusses implications for understanding riot dynamics, acknowledges data limitations, and outlines directions for future research. A data story summarizing the findings of this research is available [here](#) link.

This chapter is based on the journal article: Spierenburg, L., van Cranenburgh, S., Cats, O. (2025). Studying riots through the lens of social media. *Journal of Big Data*, 12(1), 1-20.

5.1 Introduction

Riots are instances of agitated social unrest, typically spontaneous and violent, characterized by vandalism, looting, and clashes with law enforcement (**Lachman, 1996; Wilkinson, 2009**). Often reflecting latent societal tension, riots result in substantial economic, human, and social costs such as injuries, damage, and mass arrests (**Cicchelli et al., 2007; Home Office, 2011; Sénat, 2024**). Extensive research has been devoted to understanding riots, with two primary objectives: first, identifying the underlying causes to address social tensions and prevent future incidents; and second, analyzing riot dynamics to develop effective crisis management strategies that limit social and material consequences (**DiPasquale & Glaeser, 1998; Gurr, 2015; Holdo & Bengtsson, 2020; Moran & Waddington, 2016; Newburn, 2021; Wilkinson, 2009**). Such understanding necessitates reliable data that capture the spatiotemporal distribution of such events, highlighting where and when unrest started and how it evolves over time. Data sources commonly used in the literature, such as police records or press reports, suffer from poor spatiotemporal resolution (**Baudains, Braithwaite & Johnson, 2013; Baudains, Johnson & Braithwaite, 2013; Bonnasse-Gahot et al., 2018; Davies et al., 2013**). For instance, the police records used in **Bonnasse-Gahot et al., 2018** provides the daily number of rioting events at the municipal level which does not allow to track how a single event escalated over time or spread in space.

The emergence of social media opens new opportunities for mapping riots, as they are massively used by urban residents to report incidents and provide data with a much finer spatiotemporal resolution than traditional sources (**Ahmed et al., 2019; Nicoletti et al., 2021; Panagiotopoulos et al., 2016; Prabhu et al., 2021; Procter et al., 2013; Tonkin et al., 2012; Vis, 2013**). Previous research demonstrates the potential of social media data to complement traditional sources in mapping social unrest (**Alsaedi et al., 2017; Nicoletti et al., 2021; S. L. Wilson, 2017**). For instance, **Alsaedi et al., 2017** could trace the London 2011 riots from hashtags¹ of geolocated tweets. These studies showcase the ability of social media data to trace events with substantial agreement with ground-truth datasets (**Alsaedi et al., 2017; Nicoletti et al., 2021; S. L. Wilson, 2017**).

However, research that leverages social media to map riots has been constrained by a significant methodological limitation: the reliance on textual content and structured metadata rather than visual content. Previous studies have focused primarily on analyzing tweets through hashtags, keywords, or post metadata (**Alsaedi et al., 2017; George et al., 2021; Nicoletti et al., 2021; S. L. Wilson, 2017**). This approach overlooks rich visual information contained in footage that users extensively share during riot events—content that often provides more direct and objective evidence of unrest than textual descriptions.

Addressing this methodological gap, we develop a computer vision approach capable of identifying riot-related posts from massive unstructured visual datasets. This primary contribution represents a significant departure from prior work by directly analyzing the visual content of social media posts rather than relying on textual or metadata

¹short textual tags attached to posts

cues. We train a computer vision model to classify photos and videos as riot-related using geolocated footage collected during the 2023 'Nahel Merzouk' riots in France, demonstrating the method's effectiveness through validation against press reports.

As a secondary contribution, our approach enables fine-grained spatiotemporal analysis that demonstrates the analytical potential unlocked by visual social media content. We showcase this capability by animating the evolution of riots across 7 French cities at 300×300 meter resolution and hourly intervals, revealing riot dynamics—such as their manifestation as discrete, localized bursts—that traditional data sources and text-based social media approaches cannot capture.

We organize the remainder of the paper as follows. Section 5.2 introduces the case study and the dataset employed, section 5.3 outlines the implemented methodology, section 5.4 presents the findings, and section 5.5 concludes the analysis.

5.2 Case study and datasets

We study the 'Nahel Merzouk' riots in France which took place in the summer of 2023 (see subsection 5.2.1). After collecting geolocated visual data from the social media platform Snapchat during this period, we annotate and organize the dataset to train and evaluate our computer vision model (see subsection 5.2.2). We use an external dataset documenting the spatiotemporal distribution of these riots from press reports to test the validity of model results (see subsection 5.2.3).

5.2.1 Case study

Our analysis focuses on the 'Nahel Merzouk' riots, which took place across France between June 27th and July 5th, 2023, (**Sénat, 2024**). These riots were sparked by the fatal police shooting of Nahel Merzouk, a 17-year-old teenager of Moroccan and Algerian descent, in the Paris suburb of Nanterre after attempting to drive away from police officers, who stopped him after he was caught speeding along a bus lane. The incident, viewed as emblematic of latent discrimination in disadvantaged neighborhoods, quickly ignited widespread outrage (**Beaman, 2023**). The riots initially broke out in several suburbs of Paris and quickly spread to cities and towns throughout the country. The unrest was marked by intense confrontations with the police, the burning of vehicles and buildings, and widespread acts of vandalism. Public infrastructure, including town halls, schools, and public transport facilities, as well as private properties, were severely damaged. Looting also occurred in many areas, further exacerbating the social and economic impact of the unrest (**Sénat, 2024**). In response, French authorities deployed tens of thousands of police officers across affected areas and implemented emergency measures, including curfews and restrictions on public gatherings. Social media played a pivotal role in documenting the riots, with users sharing footage about how the events unfolded. Our study demonstrates how these data can be used to draw the spatiotemporal distribution of events that are often difficult to capture with traditional data sources.

5.2.2 Social media data

We collected publicly available footage from the social media platform Snapchat taken in 7 French urban areas; namely Paris, Marseille, Lyon, Toulouse, Lille, Bordeaux, and Grenoble between June 25th and July 16th. We use the OECD definition of the urban core to delineate the study perimeter (**Dijkstra et al., 2019**). Map A in figure 5.1 shows the Paris urban core as delineated by the OECD. A critical constraint in our data collection was Snapchat's ephemeral nature: posts on the platform automatically expire and become unavailable after 1 day to 1 week, depending on user settings and content type. We began systematic data collection on June 29th. Consequently, our dataset provides comprehensive coverage from June 29th onward, while data from earlier dates (June 25th-28th) may be incomplete due to posts that had already expired before our collection began. The data consist of around 107 thousand short videos (approximately 81 %) and images (19 %) posted publicly by users. A 300×300 m sampling grid was used for data collection. Map B in figure 5.1 shows the spatial distribution of the posts collected within the time period.

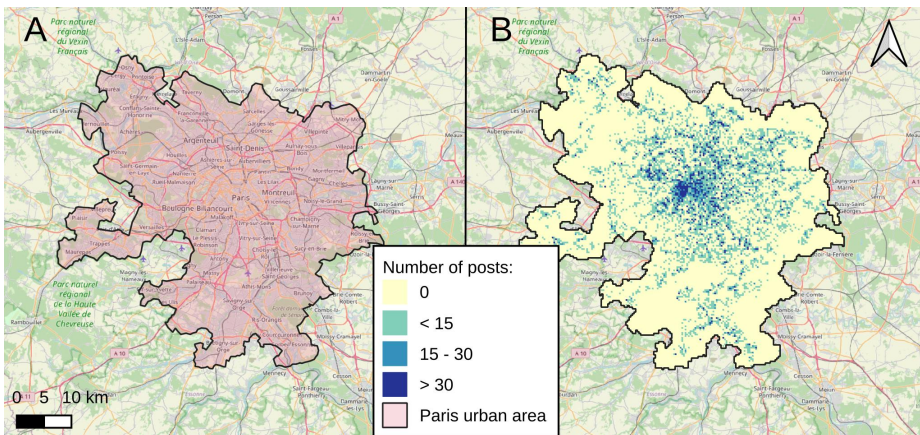


Figure 5.1 A: Paris urban core, as defined by **Dijkstra et al., 2019**. B: Number of posts per 300×300 m grid cells.

We label by hand 3,917 segments taken from 3,234 short videos. A segment consists of a set of successive frames that either (1) represents a rioting event (positive label) or (2) does not represent such an event (negative label). In our annotation process, we specifically focus on violent riot activities including arson, looting, damage to public infrastructure, and confrontations with law enforcement, explicitly excluding peaceful gatherings, demonstrations, or protests without violent elements. We annotate the data conservatively, meaning that ambiguous posts for a human annotator were classified as negative. A video can be composed of several segments with different labels. For instance, if the video scene switches from a car set on fire to the sidewalk, we manually separate the video into two segments, respectively labeled as positive and negative. A

total of 257 videos (8 %) have at least one video segment relating to a riot in this annotated dataset.

Next, the data are divided into a training set, a validation set, and a test set containing, respectively, 40%, 30%, and 30% of the labeled videos. We ensure that each video is exclusively present in one of the three sets. We tune the weights of the model using the training set. The validation set serves to keep track of the model's performance while training. The test set is a third independent dataset that is used to fairly assess the performance of the model after training on unseen data (see subsection 5.4.1).

The implementation of the proposed approach involves major privacy considerations as we collect data from a social media platform. The data collection is non-discriminative, i.e. we cannot filter irrelevant posts prior to collecting data. The data may also contain identifiable information in some cases as the identity of individuals can be revealed if their faces are included in the image or video. Sensitive personal data may therefore be exposed in case of a data breach. We reduce this risk in two stages. We first blur faces upon collection and destroy the raw data to limit identifiable details in the collected data. Then, we destroy the blurred footage after publication and publish publicly the metadata which contains the time, the location of posts, as well as the label predicted by the computer vision model (Spierenburg, Cats & van Cranenburgh, 2024).

5.2.3 Press data

We validate the results of our approach by comparing them to a dataset listing all events reported by the local press (Aubert et al., 2023). This dataset was created by a group of researchers that listed all rioting events reported in the press, day by day, in every municipality. We merge these data with the geographic borders of municipalities, allowing us to map the spatial distribution of riots on a daily basis. For instance, map A in figure 5.2 displays the rioting events reported for the night of June 29th in the city of Paris and neighboring municipalities. The dataset relates 576 events taking place in 408 French municipalities between June 27th and July 3rd. In this dataset, we observe a surge in rioting events between June 28th and June 30th (76 % of all registered events) followed a progressive return to normal between July 1st and July 3rd. We test the validity of our approach against this local press dataset by overlapping the spatiotemporal distribution of social media posts representing a riot with a map like map A in figure 5.2. Map B in figure 5.2 shows the typical spatial resolution of the press dataset, using the spatial extent of Nanterre (12 km²) as an example, which is much coarser than our 300 × 300 m resolution.

5.3 Method

This section describes the training of the computer vision model that detects rioting events from visual social media content (subsection 5.3.1) and introduces the tools used for the spatiotemporal analysis of the detected riot-related posts (subsection 5.3.2).

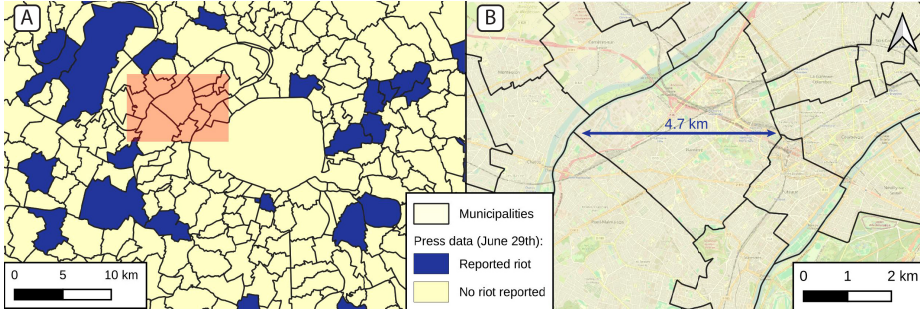


Figure 5.2 A: Rioting events on the night of June 29th were reported in the press dataset. B: Spatial resolution of the dataset.

5.3.1 Computer vision model

This study proposes a model that detects rioting events from geolocated visual content. Due to the unstructured nature of footage data, the model needs to learn what in images pertains to the concept of riots based on a sample of hand-annotated pictures. We address this challenge by fine-tuning a computer vision model that extracts visual patterns relevant to our task and classifies images based on them. Subsection 5.3.1 elaborates on the image classification task and on the model used for that purpose. Subsection 5.3.1 discusses the training process. Subsection 5.3.1 presents the decision threshold tuning for optimal classification performance.

Image classification

The task for the computer vision model is to predict the label of an image, a binary variable with a value of 1 if the image contains a riot and 0 otherwise. Training a deep learning model from scratch requires an extensive labeled dataset. In this work, we reduce drastically the amount of data needed using transfer learning (Weiss et al., 2016). This approach involves using both the architecture and the weights of an existing model trained for a similar image classification task and retraining part of its layers to specialize it to another use case.

We choose an Efficientnet architecture as the base model for transfer learning that is both more efficient and performs better on benchmark datasets than most image classification models (Tan, 2019). Vision transformers (ViT) demonstrate higher accuracy but do not justify the increased computational requirements (Dosovitskiy, 2020). We implement the smallest architecture 'Efficientnet b0', as our training set is limited and larger architectures did not show significantly better performance, using the weights trained on the ImageNet 2 dataset, publicly available on the Pytorch library (PyTorch Developers, 2024; Recht et al., 2019).

The base model efficientnet b0 is a deep neural network structured as follows: a classifier suggests the class of the image based on the visual features identified by the

feature extractor. We replace the base classifier by a custom multilayer perceptron, and optimize its architecture (depth and width) in a hyperparameter tuning step.

The Efficientnet b0 architecture constrains the data resolution. Our 480x852 images are resized to fit the 224x224 image resolution expected by the model (Tan, 2019). Other alternatives to resizing are cropping and padding, but they resulted in lower performances on the validation set². Finally, we normalize the pixel colors to the same scale as the ImageNet dataset, as it improves the performance.

Training procedure

Our training procedure includes data balancing, hard-example mining, and gradual unfreezing to improve the model's performance. This subsection discusses these three elements. First, the social media dataset is highly imbalanced towards the negative class, which biases the model towards predicting this class more. We address this issue with data balancing. We sample a number of video segments from the negative class equal to the number of video segments from the positive class at each epoch. Second, some scenes from the negative class are harder to differentiate from a riot than others (e.g. a night concert). We implement hard-example mining in the sampling process, where misclassified video segments are more likely to be sampled. Each video segment is associated with difficulty score based on cross-entropy loss, used as a weight in the sampling process. Third, transfer learning is subject to 'catastrophic forgetting', where the model discards patterns observed in the general dataset used for pretraining to capture patterns observed in the training data. This is problematic when the model starts overfitting certain visual details from the training data while discarding more general patterns observed in the pretraining data. Gradual unfreezing addresses this issue by focusing the training on the classifier part of the model (that is unrelated to the pretrained model), while preserving as much as possible the weights from the feature extractor (Davila et al., 2024). In an initial phase, we train only the model's classifier and later unfreeze layers from the feature extractor when the performance reaches a plateau on the validation set.

Decision threshold tuning

Once trained, the model produces the probability for an image to represent a riot. For videos in the dataset—composed of a series of still images (frames)—the probability of a video representing a riot is determined by aggregating probabilities across all its frames. We classify a post as representing a riot if the probability provided by the model is greater than a decision threshold. Given the dataset's significant imbalance toward the negative class, the threshold is optimized by maximizing the f1-score rather than accuracy. We then evaluate how the decision threshold influences the f1-score (plot A in figure 5.3). The f1-score reaches its maximum (0.77) for a decision threshold of 0.42,

²The reader should note that this result is not generalizable to other case studies as all our images had the same initial dimensions and were therefore distorted in the same way. The model could have learned to compensate for the constant distortion.

plateaus, and decreases when the threshold exceeds 0.6. The decision threshold yields the best results on the validation set within the range of 0.35 to 0.6.

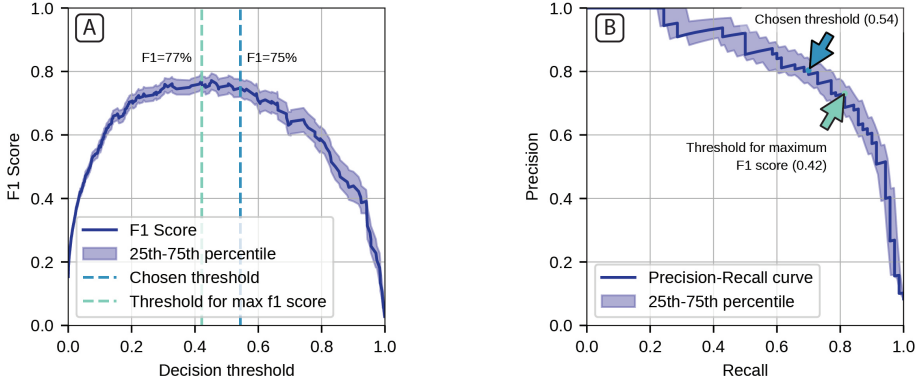


Figure 5.3 A: Effect of the decision threshold on the f1-score. B: Trade-off between precision and recall in the validation set. Percentiles calculated using bootstrap resampling with 1000 iterations.

We conduct a precision-recall analysis to set the threshold to a value within the range 0.35 to 0.6 (plot B in figure 5.3). Within the range of optimal F1-scores, precision is prioritized over recall. In other words, we prioritize being confident in labeling posts as positive at the cost of missing some, rather than aiming to avoid misses at the risk of overshooting. We anticipate significant redundancy in the data, as multiple posts may pertain to the same rioting event. This redundancy reduces the risk of missing riots when mapping their spatiotemporal distribution. Conversely, misclassifying a negative post as representing a riot leads to the erroneous identification of a non-existent event in the spatiotemporal distribution of riots. Based on these considerations, the decision threshold is set to 0.54, which maximizes the precision while maintaining a good f1-score on the validation set (0.75).

5.3.2 Spatiotemporal analysis

Beyond developing the computer vision model, this paper showcases how the method can be applied to reveal riot dynamics at high spatiotemporal resolution. This subsection introduces the analytical tools for this demonstration. We employ two approaches: quantifying clustering behavior using Ripley's spatiotemporal K function to measure burstiness, and applying density-based clustering to group posts into discrete rioting events. Subsubsection 5.3.2 describes the burst analysis methodology. Subsubsection 5.3.2 outlines the clustering approach for riot event detection.

Burst analysis

We quantify the burstiness of riot-related posts using Ripley's spatiotemporal K function, often used in crime epidemiology (Anselin et al., 2009; Gabriel, 2014; He et al., 2020;

Ripley, 1976). This function, expressed in equation 5.1, measures the average number of neighbors each post has within a spatiotemporal neighborhood delimited by spatial neighborhood radius d and temporal neighborhood radius t , quantifying the density of neighboring posts within these radii. S is the spatial extent of the study region, T is the temporal extent, n is the total number of posts, d_{ij} and t_{ij} represent the spatial and temporal distances between posts i and j respectively, and $\mathbf{1}_{condition}$ is the indicator function. The factor w_{ij} corrects for edge effects that arise when spatiotemporal neighborhoods extend beyond the boundaries of the study region, see **Gabriel, 2014** for more details.

$$K(d, t) = \frac{S \cdot T}{n^2} \sum_i^n \sum_{j, j \neq i}^n \frac{1}{w_{ij}} \cdot \mathbf{1}_{d_{ij} \leq d, t_{ij} \leq t} \quad (5.1)$$

The K function is commonly used to qualify a spatiotemporal process as bursty by comparing its value to a theoretical K function assuming random distribution across time and space. Values of $K(d, t)$ that exceed the expected under complete spatiotemporal randomness indicate clustering, while lower values suggest the presence of a regularity pattern. Social media posts are naturally highly clustered in space and time (users may upload several posts in the same location within short time intervals) and a random spatiotemporal process is therefore not a good baseline. We measure the burstiness of riot-related posts by comparing the K function to the K function of a baseline social media activity (non-riot-related posts).

We conduct separate analyses for each of the 7 cities included in our study, where the spatial extent S is defined by the geographical boundaries encompassing all posts within each urban area. The analysis is partitioned temporally on a day-by-day basis to capture daily patterns in riot activity. We measure the K function for riot-related posts during nighttime hours (6 PM to 6 AM) and non-riot-related posts during daytime hours (6 AM to 6 PM), setting the temporal extent T to 12 hours. We then compute the K function across a range of spatial radii d and temporal radii t to identify the characteristic scales at which riot events exhibit the strongest clustering. We then average the K function values across all cities and all days and calculate the ratio between the K function values for every pair of spatial and temporal radii.

Detection of riots

Rioting events cannot be readily derived from positive posts as (1) some posts may be misclassified and (2) several posts may represent the same event. We address these two issues using DBSCAN, a density-based cluster analysis, to group positive posts into rioting events (**Ester et al., 1996**). This approach assigns positive posts that are close in space and time to rioting events while filtering out isolated posts. Several parameters govern the cluster analysis: the distance function, the distance threshold d_{th} under which two posts are considered neighbors, and the minimum number of neighboring posts

n_{core} to form a cluster. The distance between two posts is calculated as the straight-line (Euclidean) distance between their spatial locations, plus a weighted measure of their temporal separation (see equation 5.2). In this equation, x_i and y_i represent the spatial coordinates of post i in meters and t_i is the instant at which the post is taken, in seconds. The parameter α relates temporal and spatial proximity. The parameters d_{th} , α , and n_{core} are set based on the results of the burst analysis in section 5.4.2.

$$d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} + \alpha|t_j - t_i| \quad (5.2)$$

5.4 Results

This section validates our computer vision approach for detecting riot-related content in social media footage and demonstrates its application for mapping urban unrest. We first assess the performance of our computer vision model through quantitative analysis on a held-out test set, qualitative examination of misclassified cases, and cross-validation against an external press dataset (subsection 5.4.1). We then showcase the analytical potential of our validated method by conducting fine-grained spatiotemporal analysis of riot evolution (subsection 5.4.2), demonstrating capabilities that traditional data sources cannot provide.

5.4.1 Validation of the computer vision model

In the following, we demonstrate the performance of our method for identifying and mapping riots from user-generated footage. We analyze the performance of our computer vision model in identifying riot-related posts quantitatively in subsection 5.4.1 and assess misclassified posts qualitatively in subsection 5.4.1. We then validate our approach by comparing the derived spatiotemporal distribution of riots with an external dataset (subsection 5.4.1).

Quantitative analysis of model performance

We assess the performance of the computer vision model on the test set, held out of the training procedure (see subsection 5.2.2 for more details). Our assessment combines quantitative analysis using key performance indicators (accuracy, precision, recall) with qualitative analysis of misclassified footage to identify edge cases and understand classification errors. The test set contains 1176 pieces of footage, with 96 riot and 1080 non-riot instances. Given this imbalanced dataset, we focus on precision and recall metrics rather than accuracy, as a naive classifier would achieve 92 % accuracy by simply predicting the majority class. We separate classified images into true negatives, true positives, false negatives, and false positives (see confusion matrix in figure 5.4). As desired, our model achieves high precision in riot classification, measured at 77% (95% CI: 69-86%³) on the test set. The recall of 68 % (95% CI: 58-77%) indicates that while

³Confidence intervals calculated using bootstrap resampling with 1000 iterations.

most riot classifications are correct, the model misses approximately one-third of actual riot footage.

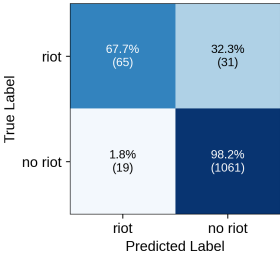


Figure 5.4 Confusion matrix for the videos in the test set, normalized by true class.

Qualitative analysis of misclassified footage

We complement our quantitative analysis of the model’s performance with a qualitative analysis of misclassified images. The objective of the qualitative analysis is to examine misclassified images and assess potential impacts on the final results, particularly concerning systematic biases that could affect riot distribution mapping. For instance, if the model consistently identifies nightclub scenes as riots, nightclubs may (incorrectly) appear as hot spots when drawing the spatiotemporal distribution of riots.

Of the 31 false negatives, the majority consist of looting scenes (14 instances), destruction events (13 instances), and fire-related incidents (4 instances). These misclassifications can be attributed to two primary factors. First, many videos contain only a few frames clearly depicting rioting activity, and when probabilities are averaged across all frames, the riot-related signal becomes diluted and falls below the classification threshold. Second, destruction and looting scenes are conceptually similar to non-riot activities from a visual perspective—people gathering or carrying objects—making them particularly challenging for the model to distinguish at the individual frame level, especially with limited training examples. The top row of figure 5.5 shows three of such examples.

Similarly, of the 19 false positives, 9 are clearly negative, meaning they would have been correctly labeled by a human annotator. These clearly misclassified posts include four point-of-view driving scenes at night featuring bright halos from streetlights, two television screens displaying riot footage, two general nighttime scenes, and one daytime scene. The remaining 10 posts are ambiguous, as additional context would be needed to confidently classify these scenes as negative. We show three of these ambiguous false positives in the bottom row of figure 5.5.

The analysis reveals no concerning patterns of misclassification that would significantly affect riot distribution mapping. Arguably, misclassified posts from the test set do not justify improving the model further, as they do not substantially affect the

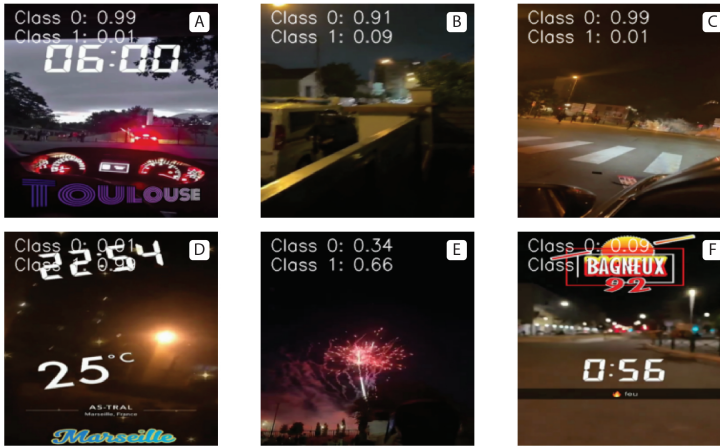


Figure 5.5 Examples of ambiguous misclassified posts. A, B, and C are false negatives, while D, E, and F are false positives. The top left corner indicates the class probability for a riot (class 1) and "no riot" (class 0). Images have been transformed as they are when supplied to the model (vertical squeeze).

5

spatiotemporal distribution of riots. Since we do not observe any systematic bias in the test data, we have no reason to expect false positives to be non-randomly distributed across space; they would add white noise to the distribution. False negatives are less problematic as intense rioting events are likely to be reported by multiple posts.

Cross-validation with press data

After assessing the performance on the test set, we infer the classes of all posts in the full dataset. This section validates the spatial distribution of riots constructed from positive posts by comparing it with the press dataset in the Paris agglomeration (see subsection 5.2.3). We investigate the extent to which both datasets overlap spatially between June 27th and July 2nd in figure 5.6 below. For a given day, municipalities where riots were reported in the press dataset are highlighted in orange. Then, for each spatial unit, we measure the share of posts classified as riot-related among the unit's total number of posts during that given day. This share is shown in blue in figure 5.6. To provide a quantitative assessment of the correspondence between our approach and the press data, we calculate the Area Under the Receiver Operating Characteristic curve (AUROC) for each day. For each municipality and day, we compute the proportion of riot-related posts among all posts in that municipality as our predictor variable, with the binary outcome indicating whether the press dataset reported a riot event in that municipality on the given day (see figure 5.6).

The quantitative assessment using AUROC reveals variable correspondence between the datasets across different days. The correspondence is moderate to good throughout the full period (AUROC ranging from 0.70 to 0.94) except from June 29th, demonstrating

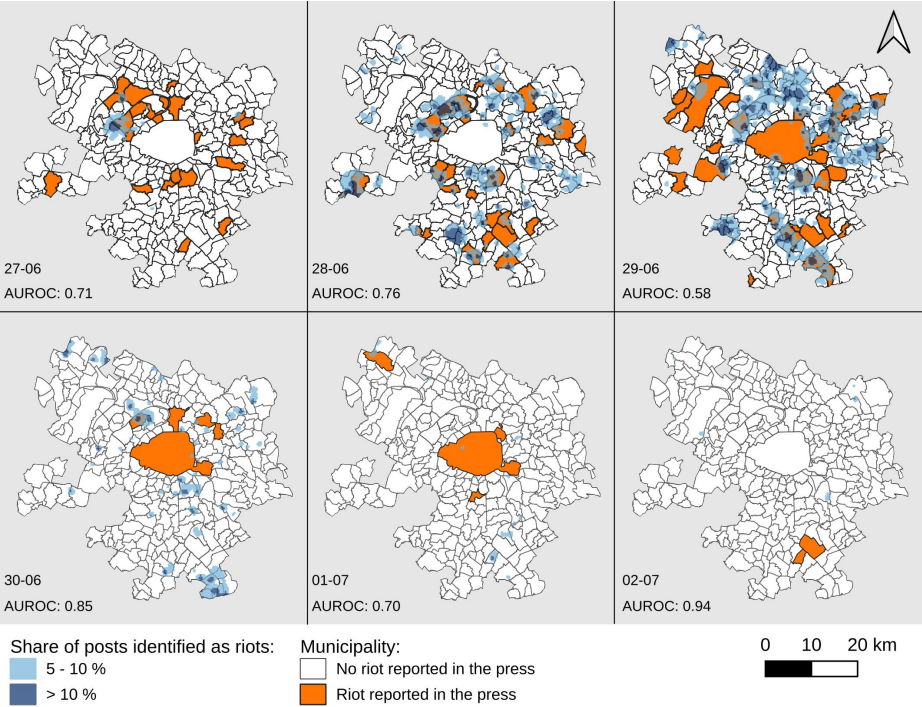


Figure 5.6 Spatial overlap between rioting events reported in the press and rioting events reported on social media night after night.

that our approach can effectively discriminate between municipalities with and without press-reported riots on several days. However, correspondence is weaker on June 29th (AUROC = 0.58).

Several factors may explain the observed misalignment across different days. Our social media data collection faced incomplete coverage during early days (June 27th-28th), which may have affected detection accuracy (see subsection 5.2.2). Additionally, the press dataset involves two distinct interpretation layers that can introduce discrepancies compared to real-time social media documentation: (1) journalist coverage decisions, where reporters make editorial choices about which events to cover based on perceived newsworthiness and (2) subsequent analyst data compilation, where researchers interpret and categorize press reports, potentially introducing additional filtering or classification biases. For example, our approach captures smaller incidents that press outlets might not consider newsworthy enough to report, potentially leading to spatial mismatches. These editorial decisions are further complicated by the temporal evolution of newsworthiness throughout the crisis. The press often prioritizes covering unexpected events during their initial phases, when public interest and the

newsworthiness of the situation are at their peak. As a result, even relatively minor incidents may have been reported more extensively on June 27th, contributing to the higher coverage compared to subsequent days. From June 29th, the press reported fewer events than observed using our approach. This might indicate coverage fatigue by press outlets as the crisis progresses into its third day. Finally, temporal discrepancies also occur due to different conventions for dating events that happen after midnight. This temporal discrepancy is evident in the southeastern area of our study region, where municipalities appear as orange (indicating riot activity) on July 2nd in the press dataset, while we report these same events as occurring in the night of July 1st.

Our results demonstrate the ability to identify hot spots in the spatial distribution of riot-related posts that align with press data, indicating that the false negative rate does not hinder the clear identification of rioting events. Moreover, false positives do not appear to generate spurious hot spots. When rioting events are, in fact, more scarce and less intense, towards the end of the crisis, we observe little hot spots in the constructed spatial distribution of riots (bottom right map of figure 5.6). False positives are not concentrated enough to appear as rioting events. We can conclude that the false negative (32 %) and the false positive (23 %) rates observed on the test set in subsection 5.4.1 do not impact substantially the subsequent spatial analysis.

5

5.4.2 Spatiotemporal analysis of riots

With our computer vision approach validated, we demonstrate its analytical potential by mapping riot dynamics at 300×300 meter spatial resolution and hourly temporal intervals. We first conduct descriptive analysis of riot evolution during a high-intensity day (subsection 5.4.2), then characterize the burstiness and spatiotemporal scales of riot activity (subsection 5.4.2), and finally regroup posts into discrete rioting events through clustering analysis (subsection 5.4.2).

Descriptive analysis

To demonstrate our approach's effectiveness in mapping riot evolution, we examine the Paris agglomeration during the night of June 29-30, the period with the highest number of recorded rioting events according to public authorities (**Sénat, 2024**). The histogram A on figure 5.7 shows the hourly distribution of posts between June 29th midday and June 30th midday for the entire Paris agglomeration. The light green bars indicate the number of negative posts, while the blue bars reflect the number of riot-related posts. The number of negative posts increases in the morning from 5 am to 10 am, reaches a peak between 10 am and 7 pm and starts decreasing in the evening at 8 pm, following typical daily activity patterns. In contrast, posts classified as riot-related exhibit a completely different temporal distribution. They occur at night, mostly between 8 pm and 3 am, peaking at 10 pm, which is consistent with the rioting activities reported in the press (**Le Parisien, 2023; L'obs, 2023**). We also plot the spatial distribution of riot-related posts on the right maps of figure 5.7. With our data, we are able to draw the spatial extent and the intensity of riots across the night. This demonstrates one of the main advantages of our analysis

compared to those relying on other data sources. It enables us to describe the evolution of riots hour by hour, at a 300×300 m resolution. Such data are instrumental in understanding the development and dynamics of riots, as they allow us to model the speed and direction in which events spread or compress.

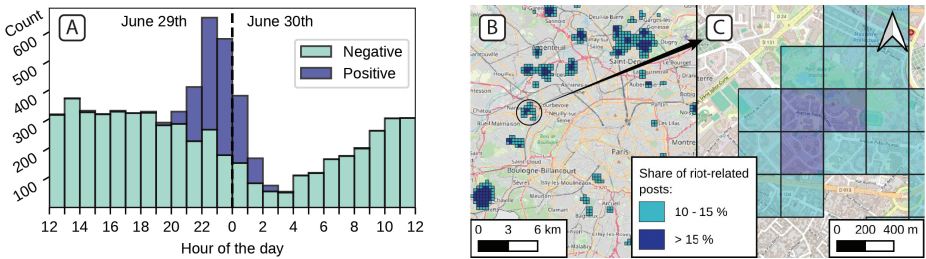


Figure 5.7 A: Temporal distribution of riot-related and non riot-related posts in the Paris agglomeration between June 29th and June 30th. B: Spatial distribution of riot-related posts in the Paris agglomeration between June 29th and June 30th. C: Zoom on a neighborhood in Rueil-Malmaison.

Burst analysis

Throughout the period June 27th-July 2nd and across all 7 cities studied, riot-related posts reported on social media occur in short and localized bursts rather than in a smooth and progressive spatiotemporal distribution. We apply the spatiotemporal K function described in subsection 5.3.2 to characterize the clustering behavior of riot-related posts across different temporal and spatial scales. Figure 5.8 shows the ratio noted f between the K function for riot-related content compared to the baseline social media activity K_0 . For all combinations of spatial and temporal radii (d , t), the K function for riot-related posts is consistently larger than the baseline's, implying that riot-related posts are spatially and temporally more concentrated than the benchmark.

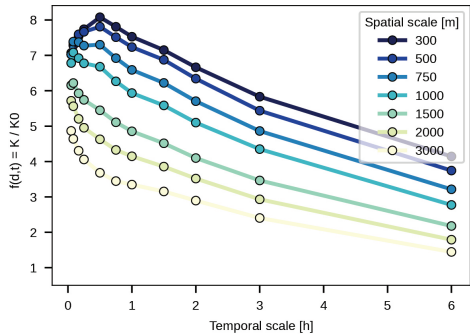


Figure 5.8 Ratio f between the K function for riot-related posts and the K function for baseline social media activity (annotated K_0) for several temporal and spatial scale.

Clustering analysis

We cluster social media posts into rioting events using the DBSCAN approach described in subsection 5.3.2. We set the α parameter relating the temporal dimension to the spatial dimension in the distance metric to 0.25 based on a sensitivity analysis of the spatiotemporal K function. The maximum of $f = K/K_0$ in figure 5.8 is reached for $t_{max}=0.5$ h, and $d_{max}=300$ m., the partial derivative analysis shows that burstiness decays 4 times stronger with distance than with time⁴ ($\frac{\partial f/\partial d}{\partial f/\partial t} = 4$). We want our clustering algorithm to reflect this asymmetry by being more restrictive spatially than temporally. Therefore, we set α to 0.25 (see equation 5.4) to reduce the weight of the temporal dimension in our distance calculation. This allows posts separated by longer time intervals to be grouped into the same riot event while maintaining strict spatial proximity requirements. The distance threshold d_{th} is set to 750, to match the function's maximum (see equation 5.5). The minimum number of neighbors n_{core} to form a cluster is set to 3, which minimizes the number of isolated points (points belonging to no clusters).

5

$$f(d, t) = K(d, t)/K_0(d, t) \quad (5.3)$$

$$\alpha = \frac{\partial f(d = d_{max}, t = t_{max})/\partial t}{\partial f(d = d_{max}, t = t_{max})/\partial d} = 1/4 \quad (5.4)$$

$$d_{th} = d_{max} + \alpha \cdot t_{max} \quad (5.5)$$

Table 5.1 summarizes the results of the clustering analysis for the cities and days with more than 40 riot-related posts. The average spatial and temporal extent of the clusters confirms the short and localized burst trend observed in the analysis of the spatiotemporal K function. The vast majority (80%) of identified riot clusters last no more than 50 minutes and remain confined to areas under 0.3 km².

One possible explanation for this pattern is the decentralized organization of rioting behavior, where rioters operate in small groups with separate objectives, resulting in localized and short-lived events. Another contributing factor could be police interventions; rapid and targeted actions may disperse a riot, prompting it to re-form at another location. The municipality of Évry-Courcouronnes on the night of June 29th illustrates the typical burst pattern observed throughout our dataset. We identify three intense bursts separated by spatiotemporal gaps of around 800 meters and 1 hour (see histogram 1.A and map 1.B in figure 5.9). Yet, we do observe several exceptions such as a riot occurring in the neighborhood of the Old Port of Marseille on July 1st (see histogram 2.A and map 2.B in figure 5.9 and table 5.1). The distribution of positive posts is much more homogeneous in space and in time than in the other riots observed in our dataset. Here, rioters may all have shared the same target neighborhood, the Old Port, being central and commercial, overwhelming public authorities.

⁴when time is expressed in seconds and space in meters

City	Date [dd-mm]	Riot- related posts	Events detected	Median posts per event	Median surface of events [km ²]	Median duration of events [hh:mm:ss]
Paris	27-06	170	19	5	0.141	00:25:07
Paris	28-06	1020	103	5	0.141	00:43:38
Paris	29-06	1370	146	5	0.114	00:38:24
Paris	30-06	491	65	4	0.141	00:16:36
Paris	01-07	203	16	4	0.212	00:42:02
Paris	02-07	52	6	4	0.071	00:05:13
Marseille	30-06	138	10	6	0.212	00:48:07
Marseille	01-07	122	4	6	0.706	02:09:27
Lyon	29-06	123	16	4	0.071	00:41:07
Lyon	30-06	116	17	4	0.071	00:30:21
Lyon	01-07	54	5	4	0.071	00:23:31
Lille	28-06	84	9	4	0.071	01:07:11
Lille	29-06	221	17	5	0.141	01:33:33
Lille	30-06	43	6	6	0.071	00:31:03
Toulouse	28-06	51	7	5	0.212	00:42:28

Table 5.1 Characteristic spatial and temporal scale of rioting clusters across cities and days with more than 40 riot-related posts.

5.5 Discussion and conclusion

This study highlights the potential of social media data for mapping riot-related events at a high spatiotemporal resolution. By employing a computer vision approach, we analyze geolocated visual content at scale, moving beyond the textual data and structured metadata used in prior research. The results demonstrate strong alignment with press-reported data, supporting the validity of our method and reinforcing the value of social media data in studying urban unrest.

In addition to demonstrating significant potential for understanding riot dynamics, our approach also gives rise to concerning possibilities for surveillance overreach and misuse beyond research contexts (Owen, 2017). Real-time deployment could enable authorities to preemptively disperse peaceful protests before they escalate, potentially infringing on democratic rights to assembly. Beyond immediate deployment concerns, the potential identification of individuals through visual content becomes especially problematic if such frameworks target legitimate protests rather than violent unrest (Smith & Miller, 2022). Additionally, historical riot mapping could justify disproportionate surveillance in specific neighborhoods, perpetuating existing policing inequities through location-based profiling (Alikhademi et al., 2022). These dual-use implications highlight the critical need for responsible implementations of such research developments.

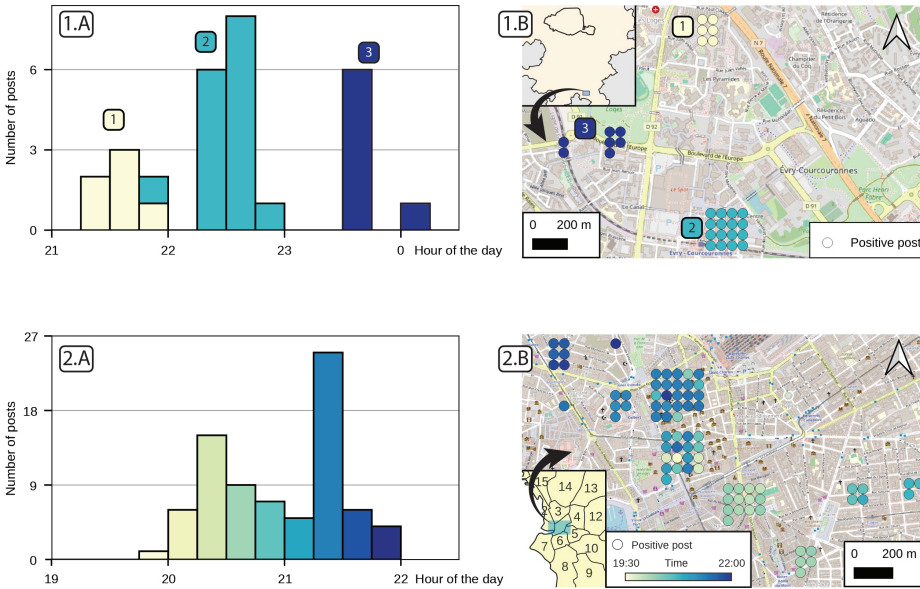


Figure 5.9 1. Three bursts identified in the temporal (1.A) and spatial (1.B) distribution of riot-related posts in Évry-Courcouronnes in the night of June 29th. Bursts are labeled by a color and a number. 2. Temporal (2.A) and spatial (2.B) distribution of riot-related posts observed in the Old Port of Marseille in the night of July 1st. The number in map 2.B annotate the districts of Marseille.

Our work presents several limitations due to the use of social media data. First, the data collection is highly dependent on social media platforms and public authorities. Platforms can limit access to their data, provide biased content by blocking or prioritizing certain posts, or even spread fake news (Timoneda, 2018; Vosoughi et al., 2018; Zimmer et al., 2019). Similarly, public authorities can also temporarily ban or restrict the usage of the platform (Baker, 2012). In the case study, we successfully collected geolocated social media posts and cross-validated them with press data, though the applicability of this method in future studies will depend on continued data access from social media platforms. Second, user-generated data introduces subjectivity (Olteanu et al., 2019). Even visual posts, though generally more objective than text, may provide an incomplete representation of the scene. Third, platform usage varies geographically and demographically across populations, potentially limiting result generalizability (Bazzaz Abkenar et al., 2021). Fourth, due to the sensitivity of the data used, the raw data cannot be shared openly. This hampers the replicability of the results (Stodden, 2014). Yet, we do share publicly the spatiotemporal distribution of the annotated posts (Spierenburg, Cats & van Cranenburgh, 2024).

While presenting notable challenges, our approach opens two promising avenues for further research: one for understanding the dynamics and another for investigating

the causes of riots. First, the high-resolution data produced by our approach are instrumental for studies modeling the dynamics of riots (**Baudains, Johnson & Braithwaite, 2013; Bonnasse-Gahot et al., 2018; Davies et al., 2013**). For example, our data can help researchers understand how riots spread, how quickly they escalate, and how localized riots transform into large-scale uprisings, which is instrumental to designing effective crisis management strategies. In particular, agent-based modeling approaches could prove especially valuable in unraveling the microscopic behavioral rules—whether at individual or small-group levels—that aggregate to produce the fragmented spatiotemporal patterns we observe. Second, the data can support studies investigating what motivates participation in riots. Spatial demographic analysis could examine whether rioting intensity correlates with indicators of inequality or urban deprivation at the neighborhood or city level, leveraging our high-resolution data to align with granular socioeconomic datasets (**Gurr, 2015; Holdo & Bengtsson, 2020; Newburn, 2021**). Additionally, investigating the relationship between online mobilization and physical manifestations of unrest could reveal how digital engagement translates into real-world action. This could involve analyzing textual social media data from platforms like X to measure online support for causes and examining how such digital mobilization correlates with the intensity and geographic distribution of actual rioting events captured through our computer vision approach. Such insights are not only valuable for academic understanding but could ultimately inform more effective approaches to addressing and preventing social tensions in urban environments.

Chapter 6

Conclusion

6.1 Main conclusions

The overarching objective of this dissertation is to reveal the spatial structure of residential segregation and riots — two manifestations of social tensions in cities (Alesina & La Ferrara, 2002; Putnam, 2000; W. Wilson, 1987). To achieve this, the research develops analytical methods capable of capturing these phenomena at the scales where they actually unfold, and applies these methods in large-scale empirical studies. These contributions are organized across four studies that form a coherent series, examining residential segregation through three complementary analytical perspectives — characterizing spatial structure across Dutch municipalities (RQ1), tracking temporal evolution of segregation patterns (RQ2), and assessing alignment between segregation and urban barriers across European cities (RQ3) — alongside analysis of riots as acute manifestations of social tension using computer vision methods (RQ4).

6.1.1 What types of spatial structure does residential segregation exhibit across urban contexts? (chapter 2)

I characterize the spatial structure of residential segregation along three dimensions: intensity (the degree to which groups are over-represented in certain regions), separation (the proportion of people living in regions in which their group is over-represented), and scale (the spatial extent of these regions). When applied to a Dutch case study to investigate segregation of the population with non-Western migration background, three main results emerge. First, the method reveals diversity across urban contexts: Dutch municipalities exhibit six of the eight theoretically possible combinations of high/low values across the three segregation dimensions (intensity, separation, and scale), with municipalities spanning nearly the full conceptual space rather than clustering around a single typical pattern. Second, my analysis reveals that intensity and separation operate independently ($\rho = -0.09$), undermining composite indices like spatial entropy, which conflate these dimensions and thus obscure important differences between cities when informing policy decisions. For instance, a city with high intensity but low separation features highly homogeneous enclaves where the group of interest is strongly overrepresented, yet most people from this group live elsewhere in mixed neighborhoods. Conversely, a city with low intensity but high separation shows moderately segregated areas — where the group is overrepresented but not to an extreme degree — that

nevertheless contain most of the group's population. These distinct configurations may require different policy approaches: the first case suggests concentrated interventions in a few highly segregated locations, while the second indicates the need for broader, city-wide measures affecting multiple moderately segregated areas — yet composite indices like spatial entropy (Theil & Finizza, 1971) would assign them similar values. Third, I also uncover a strong association between the scale of segregation and total city population: absolute segregation scale is nearly linear with city population ($\rho = 0.93$), meaning larger cities have proportionally larger segregated regions rather than more numerous smaller regions. This reveals that larger cities face potentially more important isolation issues between social groups: larger segregated regions imply greater physical distances between groups, potentially reducing opportunities for inter-group contact.

6.1.2 How does the spatial structure of residential segregation patterns evolve over time and relate to demographic change at the municipal scale? (chapter 3)

Building upon the cross-sectional analysis from the previous chapter, I track the temporal evolution of segregation patterns across 82 Dutch municipalities between 2015 and 2020, revealing two main findings. First, segregation patterns converged: municipalities with high initial intensity and separation decreased along these dimensions, while those with low initial values increased. This convergence followed clear geographic lines, with the four largest cities in the Randstad decreasing while secondary towns and rural municipalities increased. Second, the analysis reveals a positive association between demographic change and specific dimensions of segregation. The population with a non-Western migration background increased across all municipalities (mean +1.9 pp), but varied substantially by location — secondary Randstad towns experienced the highest increases while major cities and rural areas saw more modest growth. Municipalities with larger increases in non-Western migration background population experienced corresponding increases in separation and scale, while intensity remained stable. This pattern suggests that demographic growth manifested through the spatial extension of existing segregated regions and the creation of new ones, rather than through increased over-representation within existing segregated regions.

One mechanism that could underlie the decrease in separation in major cities from the Randstad while increasing in surrounding suburbs is the relocation of lower-income residents from gentrifying urban cores toward surrounding suburban municipalities (Booi, 2024; Janssen et al., 2024). As neighborhoods historically overrepresenting populations with a non-Western migration background such as Amsterdam Oost have become more mixed through in-migration of higher-income residents attracted by relative affordability, lower-income households have relocated toward suburban towns (e.g. Zaandam, Haarlemmermeer) with lower access to amenities and fewer opportunities for social integration (Hochstenbach & Musterd, 2018). Since income and migration background are correlated in the Netherlands, such income-driven displacement would manifest in the data as changing segregation patterns along migration background lines,

even though the underlying mechanism operates through housing costs rather than ethnicity per se (Jongen et al., 2024; San Millán et al., 2025). However, since this analysis tracks patterns at the municipal level and focuses on the migration background rather than individual residential trajectories, this interpretation remains speculative.

6.1.3 To what extent the spatial structure of residential segregation follows urban barriers? (chapter 4)

Research has documented alignment between residential segregation and urban barriers in the US and South African contexts due to the legacy of past segregation policies (Ananat, 2011; Maharaj, 2020). To investigate whether similar alignment occurs in contexts without such policies, I examined the segregation of populations with a non-EU migration background across 520 cities in eight European countries. In European contexts without explicit segregation policies, I do not observe general alignment between residential segregation and urban fragmentation. In fact, many cities exhibit significantly less alignment than would be expected by chance. This pattern may reflect that urban barriers create similar conditions on both sides — whether positive conditions like scenic views along rivers or negative conditions like noise and air pollution near motorways — and these attraction or repulsion effects may exert stronger influence on residential choice than any preference to use barriers as separators from dissimilar groups. Despite this overall pattern, I identify substantial regional variations: cities in Germany (15% showing statistically significant alignment) and the Netherlands (17%) demonstrate notably stronger alignment than random. In these cities where alignment does occur, barriers may function as logical separators between distinct spatial units — such as delineating social housing developments from private housing — with segregation patterns reflecting differences in housing stock rather than the barriers themselves creating social separation. The geographic clustering of these cities in Germany and the Netherlands suggests this pattern may reflect country-specific urban planning tendencies in how infrastructure relates to spatial organization of development.

6.1.4 How to identify riot-related content from visual social media data, and what does this reveal about the spatiotemporal nature of riots?(chapter 5)

I show that visual social media data can map the spatiotemporal distribution of riots at fine resolution, demonstrated through a case study of the 2023 Nohel Merzouk riots in France. This visual social media analysis revealed riot dynamics at 300m spatial resolution — granularity unattainable with traditional data sources. The computer vision model identified riot-related content with 72% f1 score (77% precision, 68% recall), and the resulting spatiotemporal distribution showed good correspondence with municipal-level, day-by-day press data. Most strikingly, the fine-grained analysis revealed riots as highly localized, bursty phenomena: 80% of events lasted less than 50 minutes and remained confined to areas under 0.3 km².

The bursty, localized nature of rioting events revealed by this analysis invites a reframing of how riots are theoretically understood. The dominant view in the literature models riots as collective phenomena driven by crowd contagion, where unrest spreads as individuals observe and join others (**Bonnasse-Gahot et al., 2018; Nicoletti et al., 2021**). However, highly localized events, scattered simultaneously across the urban area, are more consistent with distributed, parallel mobilization by small independent groups responding to a shared grievance, here Nahel Merzouk's death, than with a single spreading wave. This distinction carries both theoretical and practical implications: contagion-driven riots would call for containment strategies that break transmission chains between locations, while distributed independent mobilization suggests that containment is less effective and that policy should instead focus on addressing the underlying conditions that simultaneously activate multiple groups across a city.

6

To sum up, the findings of this thesis reveal that both segregation and riots are not social issues that occur uniformly within cities, but spatial phenomena whose organization reflects underlying social processes. Research on residential segregation has increasingly recognized this, developing spatially-sensitive indices that account for the scale at which segregation manifests by incorporating the local context of spatial units in their computation (**Clark et al., 2015; Olteanu et al., 2019; Petrović et al., 2018; Reardon & O'Sullivan, 2004**). This thesis continues and extends this line of work: where these approaches capture the scale at which segregation operates, I argue for also examining its spatial structure, that is, the geographic arrangement of segregated areas relative to each other and to urban features. Conventional riot analyses, by contrast, remain at an earlier stage, counting events at the municipal level (**Bonnasse-Gahot et al., 2018; Nicoletti et al., 2021**) and collapsing fine-grained dynamics into aggregate tallies that obscure where, when, and how unrest unfolds. In both cases, moving beyond metrics that discard spatial arrangement, toward approaches that explicitly interrogate how these phenomena are organized across urban space, is essential for understanding the processes that generate them. This methodological argument unfolds progressively across the four studies. Chapters 2 and 3 develop indicators sensitive to the spatial structure, derived from explicit spatial regionalization. Chapter 4 takes the next step relating the spatial arrangement of segregation directly to urban form, asking not just how segregation is structured, but whether that structure aligns with urban barriers in the city. Chapter 5 makes a parallel move: rather than aggregating activity to municipal daily counts as prior work has done, it maps discrete events at 300m resolution and hourly intervals, revealing the bursty, localized structure that coarser approaches could not detect. Together, these studies suggest that social tensions, whether chronic patterns like segregation or acute events like riots, can only be properly understood through approaches that preserve and interrogate their spatial organization, and that methods which discard spatial arrangement risk mischaracterizing the phenomena they seek to explain.

6.2 Limitations and directions for future work

While the preceding chapters provide answers to the four research questions, each answer opens avenues for deeper investigation. I have documented what segregation patterns exist across Dutch municipalities, how they evolve over time, and where they align with urban barriers in European cities. I have also mapped the fine-grained spatiotemporal distribution of riots. This section identifies limitations in my research and proposes research directions that would move from documenting patterns toward explaining the processes that generate them. Pursuing these directions would advance both theoretical understanding and enable more targeted policy interventions.

6.2.1 From "what patterns exist" to "how patterns emerge"

My analysis in chapter 2 and 3 examines residential segregation at the spatial pattern level, while segregation results from the accumulation of individual residential choices and constraints. My answers to RQ1 — What types of spatial structure does residential segregation exhibit? — and RQ2 — How does this structure evolve over time? — characterize segregation patterns and track their changes, but do not look into underlying housing mobility processes — the moves into, out of, and within neighborhoods — that drive these pattern formations and transformations. "Investigating housing dynamics would extend my answers from "what patterns exist" to explaining "how and why these patterns emerge", building theory about how individual mobility decisions generate observed segregation structures. Such research would enable interventions that address behavioral drivers of segregation rather than responding to spatial outcomes.

This line of research would trace how individual housing choices, shaped by preferences, constraints, and life course events, accumulate into city-level segregation patterns. Importantly, it would allow distinguishing between self-selection (sorting based on amenities, affordability) and homophily (preferences for same-group proximity) as drivers of segregation (**Banzhaf & Walsh, 2008; Brueckner et al., 1999; Clark, 1991; Fossett, 2006; Massey & Denton, 1988; Schelling, 1971; Tiebout, 1956**). This could be pursued by combining longitudinal individual tracking — following people's residential moves over time — with residential choice modeling that statistically identifies which individual attributes, housing characteristics, and neighborhood features drive relocation decisions and destination choices. Individual-level microdata enabling such analysis is already accessible in countries like the Netherlands (CBS microdata) and Sweden (**Centraal Bureau voor de Statistiek, 2021**). Agent-based modeling could complement this approach by testing whether the behavioral mechanisms identified through choice models, when simulated at the city scale as in **Anderson et al., 2021; Schelling, 1971; Zuccotti et al., 2024**, generate segregation patterns consistent with observed reality.

6.2.2 Understanding the mechanisms underlying segregation-barrier alignment

In answering RQ3 ("To what extent does the spatial structure of residential segregation follow railways, roads, and rivers?") in chapter 4, I identify alignment patterns across 520 European cities but do not explain *why* this alignment emerges in specific countries — particularly Germany and the Netherlands — while remaining absent elsewhere. Identifying whether specific urban development strategies in Germany and the Netherlands generate this alignment would clarify the mechanisms linking urban barriers to demographic patterns. For instance, if barriers initially marked the edge of urban development at a particular period — with one side containing housing from that building epoch — and the city later expanded beyond these barriers with different housing types (such as social housing or newer residential developments), then contemporary segregation patterns would reflect this sequential development process rather than the barrier itself creating demographic separation. Understanding these mechanisms could also guide future urban planning helping policymakers to anticipate and potentially mitigate segregation effects.

Future research could pursue this explanatory goal by distinguishing whether alignment reflects barriers functioning as active demographic separators (homophily-driven) (Clark, 1991; Fossett, 2006; Massey & Denton, 1988; Schelling, 1971) versus barriers delineating different urban development zones (housing stock-driven self-selection) (Banzhaf & Walsh, 2008; Brueckner et al., 1999; Tiebout, 1956). One approach would leverage natural experiments where barriers have been removed — such as the underground relocation of railways in Delft — using methods that analyze household mobility patterns across demographic boundaries to test whether segregation patterns persist or dissolve over time. Such an analysis could build on established causal inference frameworks for natural experiments (Imbens & Angrist, 1994), drawing methodological inspiration from studies examining mobility patterns across physical divides (Olnier et al., 2024) and leveraging infrastructure changes as quasi-experimental settings to identify local treatment effects (Foley et al., 2017). If segregation diminishes following barrier removal, this would suggest barriers actively facilitated segregation through homophily-driven preferences to maintain distance from dissimilar groups.

6.2.3 Modeling spatiotemporal dependencies in riot propagation

In chapter 5, I answer RQ4 ("What can visual social media reveal about the spatiotemporal distribution of riots?") by mapping rioting events at fine spatiotemporal resolution and characterizing them as short-lived, localized, and bursty. However, I did not examine spatiotemporal dependencies between events, such as ripple or cascading effects. Examining these spatiotemporal dependencies would advance theoretical understanding of riot dynamics — revealing whether riots spread through contagion-like processes, operate as independent outbreaks, or follow predictable spatiotemporal patterns. Such understanding would inform containment strategies by revealing whether interventions should prioritize rapid response to emerging hotspots (if riots spread

locally), monitoring of transportation corridors (if activity jumps discontinuously), or prevention in vulnerable neighborhoods (if events remain spatially independent).

This points toward a promising research direction: developing methods to model spatiotemporal dependencies between rioting events. One possibility could be investigating whether riots exhibit self-reinforcing dynamics where events trigger subsequent nearby events. This encompasses point process approaches such as Hawkes processes (**Hawkes, 1971**), which have been applied theoretically to riot modeling (**Berestycki et al., 2015**); epidemic-style models (SIR/SEIR adapted for urban unrest) (**Haggett, 2000**), which have been calibrated with real riot data at coarse spatial resolutions (**Bonnasse-Gahot et al., 2018**); and spatial diffusion models (**Cliff & Ord, 1981**; **Myers, 2000**). Such models would reveal whether riots spread through endogenous escalation mechanisms or remain as independent, isolated incidents — informing whether containment should focus on preventing initial sparks versus breaking transmission chains. A second approach would characterize how riot activity moves through urban space by analyzing directional spread patterns, distance-decay relationships, and the role of urban features as propagation corridors or barriers (**Baudains, Johnson & Braithwaite, 2013**). Determining whether riots spread radially from initial locations, jump discontinuously across space, or follow specific urban pathways (transportation networks, neighborhood boundaries) would distinguish between underlying mechanisms — spontaneous local ignition, coordinated movement, social contagion, or strategic targeting.

6.2.4 Toward an integrated spatial analysis of segregation and riots

The four studies in this dissertation treat residential segregation and riots as distinct phenomena, each examined through its own analytical lens. Yet both are framed from the outset as manifestations underlying social tensions in cities. This raises a question that this dissertation does not address but that future research should: are these two phenomena spatially connected?

There are reasons to expect they might be. Studies on residential segregation and on riots have documented common structural conditions. More specifically, income inequality and discrimination have been identified as antecedents of both (**Body-Gendrot, 2013**; **Reardon & Bischoff, 2011**; **San Millán et al., 2025**). Beyond these shared roots, a body of qualitative research suggests a more direct link: residential segregation concentrates disadvantaged populations in areas with fewer amenities, weaker institutional presence, and limited economic opportunity, generating the accumulated grievances that can erupt into collective violence (**Holdo & Bengtsson, 2020**; **Kawalerowicz & Biggs, 2015**; **Lapeyronnie, 2006**). This mechanism has been documented qualitatively across several European riot episodes, including the French banlieue riots of 2005, the London riots of 2011, and the Stockholm riots of 2013, where the most affected neighborhoods were typically those combining overrepresentation of population with a migration background and socioeconomic marginalization.

Despite this qualitative evidence, systematic quantitative research testing whether areas with stronger segregation experience more intense riot activity remains scarce. Such a study would require data on both segregation structure and riot intensity across multiple cities and multiple riot episodes. A panel approach tracking segregation profiles and riot intensity across successive unrest episodes within the same cities, for instance comparing the 2005 and 2023 French riots across neighborhoods, would allow testing whether the association between segregation and riot propensity is consistent across time and context. Establishing this association empirically would strengthen the policy case for reducing residential segregation not only as an equity goal in itself, but as a means of addressing the conditions that make urban unrest more likely.

6.3 Implications for policymakers

The empirical findings from this dissertation yield concrete implications for policymakers across multiple domains — from urban planning and housing policy to data governance and the ethical use of emerging technologies. I address policymakers at three governance levels: municipal authorities managing housing and urban planning, national governments responsible for resource allocation and law enforcement oversight, and EU institutions setting data policy frameworks. These implications serve a dual purpose: identifying actionable strategies to address residential segregation while cautioning against potentially harmful applications of analytical capabilities. Some recommendations are specific to Dutch or European contexts, reflecting the geographic scope of my analyses, yet the underlying principles may inform policy decisions in other settings.

Since 2015, residential segregation along migration background in the Netherlands persists and has worsened in many municipalities where it was historically less severe (RQ2). Specifically, in suburban towns within the Randstad region and in rural municipalities — where the proportion of migrants living in segregated areas (separation) and the spatial extent of these areas (scale) were historically lower than in the four largest cities — these indicators have increased and are converging toward the levels observed in major cities. This pattern suggests that national governments — particularly ministries responsible for housing and integration policy — should direct greater attention and resources to secondary towns and rural municipalities through policies that promote mixed-tenure developments (integrating social and private housing) rather than concentrated social housing estates, as preventing the emergence or early-stage growth of segregation may be more feasible than reversing entrenched patterns in cities where segregation is already severe. Furthermore, municipalities that experienced larger increases in their population with a non-Western migration background also saw corresponding increases in separation and scale, suggesting that demographic growth directly contributed to the spatial expansion of segregated areas. These findings are directly relevant to current Dutch policy debates surrounding the *Spreadingswet* (Dispersal Act), which mandates the distribution of asylum seekers across municipalities (**der Nederlanden, 2024**). While this law aims to share the responsibility

for sheltering refugees, my findings underscore that dispersal alone is insufficient — receiving municipalities must simultaneously implement integration-oriented planning strategies. This underscores the need for municipal planning departments and housing authorities to tailor urban planning strategies to accommodate demographic changes resulting from immigration policies, ensuring that housing allocation and neighborhood development facilitate residential integration rather than reinforcing existing patterns of spatial segregation.

Chapter 4 examines whether urban barriers — railways, motorways, and waterways — function as boundaries between social groups in European cities, a pattern commonly observed in the United States, where the expression ‘wrong side of the tracks’ captures how infrastructure divides communities along socioeconomic lines. The analysis of 520 European cities reveals that the segregation of populations with a non-EU migration background does not systematically align with these urban barriers. In other words, residents do not use physical barriers to separate themselves from dissimilar groups. If such preferences exist, they are weaker than other residential choice factors, such as housing affordability, proximity to amenities, or the environmental effects (noise, pollution, views) that barriers themselves create on both sides. These findings suggest that municipal planning departments and national infrastructure ministries should recognize that while costly infrastructure interventions — such as putting railways or highways underground — may be justified on other grounds (noise reduction, urban aesthetics, walkability), they should not be pursued as a strategy for reducing segregation, as such investments are unlikely to deliver meaningful social integration benefits in European contexts.

The availability of fine-resolution demographic data — at the scale of Dutch postcodes containing approximately 50 people — enables the use of methods that capture the spatial structure of social phenomena at scales where it actually manifests. National statistical agencies (such as CBS in the Netherlands) should continue to make fine-resolution demographic data openly accessible to researchers, recognizing that such data is essential for understanding and addressing urban social phenomena. This requires balancing data resolution against privacy protection through anonymization techniques such as cell suppression and perturbation, as currently employed by the Dutch Statistics Office. While some countries like the Netherlands have made such data available, the absence of consistent standards across Europe limits comparative research. EU institutions responsible for data policy — particularly the European Commission (through Eurostat) and the European Parliament — should establish EU-wide frameworks for fine-resolution demographic data — defining both spatial units (finer than existing NUTS-3 regions) and harmonized demographic categories — would enable systematic cross-country comparisons and help identify generalizable patterns across different European urban contexts.

Chapter 5 from this dissertation shows that analyzing visual social media data from platforms like Snapchat can reveal fine-grained riot dynamics, capturing how these events emerge, spread, and evolve across urban spaces. However, while this

capability can serve analytical purposes, it also raises ethical concerns when applied for participant identification or predictive policing. First, using social media data to identify riot participants represents a serious erosion of civil liberties — particularly given the risk that such methods could extend from violent riots to peaceful protests. National legislators and EU institutions responsible for digital rights frameworks — particularly the European Parliament and national parliaments — should prohibit identification approaches and instead limit applications to understanding riot dynamics and addressing underlying causes. Second, using social media data to predict riots and deploy police preemptively risks creating a counterproductive cycle: targeted neighborhoods become stigmatized as dangerous, deepening potential mistrust between residents and authorities and reinforcing the social tensions that contribute to unrest. Police oversight bodies, interior ministries, and municipal authorities responsible for public order should therefore refrain from developing or deploying predictive policing approaches based on social media data.

6

This dissertation set out to reveal the spatial structure of residential segregation and riots in cities. By developing methods to capture these phenomena at the scales where they actually unfold, this research demonstrates that both segregation and riots are not merely social issues that happen to occur in urban spaces, but fundamentally spatial phenomena whose organization reflects and shapes urban social dynamics. Segregation operates as a chronic, large-scale embedding of social divisions in residential patterns, unfolding over years and decades across broad urban areas, while riots emerge as acute, micro-scale eruptions concentrated in specific locations over minutes and hours. Understanding this spatial dimension — how these phenomena organize across different spatiotemporal registers — is essential for developing effective interventions, whether addressing the slow-moving residential patterns of segregation or the rapid emergence of riots.

Bibliography

- Ahmed, S., Cho, J., & Jaidka, K. (2019). Framing social conflicts in news coverage and social media: A multicountry comparative study. *International Communication Gazette*, 81(4), 346–371.
- Aiello, L. M., Vybornova, A., Juhász, S., Szell, M., & Bokányi, E. (2025). Urban highways are barriers to social ties. *Proceedings of the National Academy of Sciences*, 122(10), e2408937122.
- Albada, K., Hansen, N., & Otten, S. (2021). Polarization in attitudes towards refugees and migrants in the Netherlands. *European Journal of Social Psychology*, 51, 627–643. <https://doi.org/10.1002/ejsp.2766>
- Alesina, A., & La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics*, 85(2), 207–234. [https://doi.org/10.1016/S0047-2727\(01\)00084-6](https://doi.org/10.1016/S0047-2727(01)00084-6)
- Alessandrini, A., Natale, F., Sermi, F., & Vespe, M. (2017). *High resolution map of migrants in the EU* (tech. rep.). Joint Research Centre. Luxembourg (Luxembourg), Publications Office of the European Union.
- Alhabash, S., & Ma, M. (2017). A tale of four platforms: Motivations and uses of facebook, twitter, instagram, and snapchat among college students? *Social media+ society*, 3(1), 2056305117691544.
- Alikhademi, K., Drobina, E., Prioleau, D., Richardson, B., Purves, D., & Gilbert, J. E. (2022). A review of predictive policing from the perspective of fairness. *Artificial Intelligence and Law*, 1–17.
- Alsaedi, N., Burnap, P., & Rana, O. (2017). Can we predict a riot? disruptive event detection using Twitter. *ACM Transactions on Internet Technology (TOIT)*, 17(2), 1–26.
- Ananat, E. O. (2011). The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *American Economic Journal: Applied Economics*, 3, 34–66. <https://doi.org/10.1257/app.3.2.34>
- Anciaes, P. R., Jones, P., & Mindell, J. S. (2016). Community severance: Where is it found and at what cost? *Transport Reviews*, 36(3), 293–317.
- Anderson, T., Leung, A., Dragicevic, S., & Perez, L. (2021). Modeling the geospatial dynamics of residential segregation in three canadian cities: An agent-based approach. *Transactions in GIS*, 25(2), 948–967.
- Andersson, E. K., & Malmberg, B. (2015). Contextual effects on educational attainment in individualised, scalable neighbourhoods: Differences across gender and social class. *Urban Studies*, 52, 2117–2133. <https://doi.org/10.1177/0042098014542487>
- Andersson, E. K., Malmberg, B., Costa, R., Sleutjes, B., Marcin, Stonawski, J., & Valk, H. A. G. D. (2018). A comparative study of segregation patterns in Belgium,

- Denmark, the Netherlands and Sweden: Neighbourhood concentration and representation of non-European migrants. *European Journal of Population*, 34, 251–275. <https://doi.org/10.1007/s10680>
- Anselin, L., Griffiths, E., & Tita, G. (2009). Crime mapping and hot spot analysis. In *Environmental criminology and crime analysis* (pp. 119–138). Willan.
- Arbaci, S. (2007). Ethnic segregation, housing systems and welfare regimes in europe. *European Journal of Housing Policy*, 7(4), 401–433.
- Atkinson, R., Wulff, M., Reynolds, M., & Spinney, A. (2011). *Gentrification and displacement: The household impacts of neighbourhood change* (tech. rep.). Australian Housing and Urban Research Institute.
- Aubert, R., Dupont, S., Laurent, S., & Pierre, S. (2023). La cartographie d'une semaine d'émeutes en France [Last accessed on 2024-09-06 from https://www.lemonde.fr/societe/article/2023/07/07/la-cartographie-d-une-semaine-d-emeutes-en-france_6180894_3224.html]. *Le Monde*. https://www.lemonde.fr/societe/article/2023/07/07/la-cartographie-d-une-semaine-d-emeutes-en-france_6180894_3224.html
- Baker, S. A. (2012). The english riots of 2011: A summer of discontent. Waterside Press.
- Baldwin Hess, D., Tammaru, T., & Van Ham, M. (2018). *Housing estates in Europe: Poverty, ethnic segregation and policy challenges*. Springer Nature.
- Ballard, R., & Hamann, C. (2021). Urban socio-economic segregation and income inequality: A global perspective. Springer International Publishing.
- Banzhaf, H. S., & Walsh, R. P. (2008). Do people vote with their feet? An empirical test of tiebout's mechanism. *American economic review*, 98(3), 843–863.
- Banzhaf, H. S., & Walsh, R. P. (2013). Segregation and tiebout sorting: The link between place-based investments and neighborhood tipping. *Journal of Urban Economics*, 74, 83–98.
- Barrie, C., & Frey, A. (2021). Faces in the crowd: Twitter as alternative to protest surveys. *PloS one*, 16(11), e0259972.
- Baudains, P., Braithwaite, A., & Johnson, S. D. (2013). Target choice during extreme events: A discrete spatial choice model of the 2011 London riots. *Criminology*, 51(2), 251–285.
- Baudains, P., Johnson, S. D., & Braithwaite, A. M. (2013). Geographic patterns of diffusion in the 2011 London riots. *Applied Geography*, 45, 211–219.
- Bazzaz Abkenar, S., Haghi Kashani, M., Mahdipour, E., & Jameii, S. M. (2021). Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. *Telematics and Informatics*, 57, 101517. <https://doi.org/https://doi.org/10.1016/j.tele.2020.101517>
- Beaman, J. (2023). George Floyd across Europe, or Europe's George Floyds. *European Journal of Cultural and Political Sociology*, 10(4), 678–685.
- Bellman, B., Spielman, S. E., & Franklin, R. S. (2018). Local population change and variations in racial integration in the United States, 2000–2010. *International Regional Science Review*, 41, 233–255. <https://doi.org/10.1177/0160017616665669>

- Benassi, F., Bonifazi, C., Heins, F., Lipizzi, F., & Strozza, S. (2020). Comparing residential segregation of migrant populations in selected European urban and metropolitan areas. *Spatial Demography*, 8, 269–290.
- Benassi, F., Naccarato, A., Iglesias-Pascual, R., Salvati, L., & Strozza, S. (2023). Measuring residential segregation in multi-ethnic and unequal European cities. *International Migration*, 61(2), 341–361.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165–1188.
- Berestycki, H., Nadal, J.-P., & Rodriguez, N. (2015). A model of riots dynamics: Shocks, diffusion and thresholds. <https://arxiv.org/abs/1502.04725>
- Bickford, A., & Massey, D. S. (1991). Segregation in the second ghetto: Racial and ethnic segregation in American public housing, 1977. *Social Forces*, 69(4), 1011–1036.
- Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The quarterly journal of economics*, 114(2), 577–599.
- Body-Gendrot, S. (2013). Urban violence in France and England: Comparing Paris (2005) and London (2011). *Policing and society*, 23(1), 6–25.
- Body-Gendrot, S. (2016). Making sense of French urban disorders in 2005. *European Journal of Criminology*, 13(5), 556–572. <https://doi.org/10.1177/1477370816640819>
- Boeing, G. (2017). Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65, 126–139.
- Bohstedt, J. (2013). The dynamics of riots: Escalation and diffusion/contagion. In *The dynamics of aggression* (pp. 257–306). Psychology Press.
- Bonnasse-Gahot, L., Berestycki, H., Depuiset, M.-A., Gordon, M. B., Roché, S., Rodriguez, N., & Nadal, J.-P. (2018). Epidemiological modelling of the 2005 French riots: A spreading wave and the role of contagion. *Scientific Reports*, 8(1), 107.
- Booi, H. (2024). Spillover of urban gentrification and changing suburban poverty in the Amsterdam metropolis. *Urban Studies*, 61(3), 495–512.
- Boschman, S., & van Ham, M. (2015). Neighbourhood selection of non-western ethnic minorities: Testing the own-group effects hypothesis using a conditional logit model. *Environment and Planning A*, 47, 1155–1174.
- Boterman, W. R., Musterd, S., & Manting, D. (2021). Multiple dimensions of residential segregation. the case of the metropolitan area of Amsterdam. *Urban Geography*, 42(4), 481–506.
- Briggs, D. (2012). *The english riots of 2011: A summer of discontent*. Waterside Press.
- Brown, L. A., & Chung, S.-Y. (2006). Spatial segregation, segregation indices and the geographical perspective. *Population, space and place*, 12(2), 125–143.
- Brueckner, J. K., Thisse, J.-F., & Zenou, Y. (1999). Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European economic review*, 43(1), 91–107.

- Castells, M. (1977). *The urban question* [See esp. Part I and Ch. 2-3 on spatial inequalities, segregation, and urban social movements; links structural tensions to both segregation and urban protest/riots]. MIT Press.
- Castells-Quintana, D., Royuela, V., & Veneri, P. (2020). Inequality and city size: An analysis for oecd functional urban areas. *Papers in Regional Science*, 99(4), 1045–1064.
- Centraal Bureau voor de Statistiek (Ed.). (2016). *Jaarrapport integratie 2016*. Centraal Bureau voor de Statistiek.
- Centraal Bureau voor de Statistiek. (2019). 8.1.5 ongelijkheid (gini-coëfficiënt) inkomen, 2017 [Retrieved from <https://longreads.cbs.nl/welvaartnederland-2019/ongelijkheid-in-inkomen-en-vermogen/> in July 2022.].
- Centraal Bureau voor de Statistiek. (2020). Kerncijfers wijken en buurten 2020 [<https://www.cbs.nl/nl-nl/maatwerk/2023/50/kerncijfers-wijken-en-buurten-2020>, last accessed in January 2024.].
- Centraal Bureau voor de Statistiek. (2021). Microdata: Zelf onderzoek doen [<https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-enmicrodata/microdata-zelf-onderzoek-doen>, last accessed in June 2021.].
- Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106, 855–902. <https://doi.org/10.1257/aer.20150572>
- Chodrow, P. S. (2017). Structure and information in spatial segregation. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 11591–11596. <https://doi.org/10.1073/pnas.1708201114>
- Chyn, E., Haggag, K., & Stuart, B. A. (2022). *The effects of racial segregation on intergenerational mobility: Evidence from historical railroad placement* (tech. rep.). National Bureau of Economic Research.
- Cicchelli, V., Galland, O., de Maillard, J., & Misset, S. (2007). Retour sur les violences urbaines de l'automne 2005: Émeutes et émeutiers à Aulnay-sous-Bois. *Horizons stratégiques*, (1), 98–119.
- Clark, W. A. (1991). Residential preferences and neighborhood racial segregation: A test of the schelling segregation model. *Demography*, 28(1), 1–19.
- Clark, W. A., Anderson, E., Östh, J., & Malmberg, B. (2015). A multiscale analysis of neighborhood composition in los angeles, 2000–2010: A location-based approach to segregation and diversity. *Annals of the Association of American Geographers*, 105, 1260–1284. <https://doi.org/10.1080/00045608.2015.1072790>
- Cliff, A. D., & Ord, J. K. (1981). *Spatial processes: Models and applications* [Mathematical foundations for spatial diffusion processes]. Pion Limited.
- Cohen, J. P., & Coughlin, C. C. (2008). Spatial hedonic models of airport noise, proximity, and housing prices. *Journal of regional science*, 48(5), 859–878.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95–S120.

- Conway, D., Li, C. Q., Wolch, J., Kahle, C., & Jerrett, M. (2010). A spatial autocorrelation approach for examining the effects of urban greenspace on residential property values. *The Journal of Real Estate Finance and Economics*, 41(2), 150–169.
- Costa, R., & Valk, H. A. G. D. (2018). Ethnic and socioeconomic segregation in Belgium: A multiscalar approach using individualised neighbourhoods. *Eur J Population*, 34, 225–250. <https://doi.org/10.1007/s10680>
- Cottrell, M., Olteanu, M., Randon-Furling, J., & Hazan, A. (2017). Multidimensional urban segregation: An exploratory case study. *2017 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM)*, 1–7.
- Custers, G., & Engbersen, G. (2022). The urban class structure: Class change and spatial divisions from a multidimensional class perspective. *Urban Geography*, 43(6), 917–943.
- Davies, T. P., Fry, H. M., Wilson, A. G., & Bishop, S. R. (2013). A mathematical model of the London riots and their policing. *Scientific Reports*, 3(1), 1303.
- Davila, A., Colan, J., & Hasegawa, Y. (2024). Comparison of fine-tuning strategies for transfer learning in medical image classification. *Image and Vision Computing*, 146, 105012.
- Debrezion, G., Pels, E., & Rietveld, P. (2007). The impact of railway stations on residential and commercial property value: A meta-analysis. *The journal of real estate finance and economics*, 35(2), 161–180.
- Delmelle, E. C. (2017). Differentiating pathways of neighborhood change in 50 U.S. metropolitan areas. *Environment and Planning A*, 49, 2402–2424. <https://doi.org/10.1177/0308518X17722564>
- Denton, N. A. (2006). Segregation and discrimination in housing. *A right to housing: Foundation for a new social agenda*, 61, 68.
- der Nederlanden, R. (2024). Wet gemeentelijke taak mogelijk maken asielopvangvoorzieningen (spreidingswet) [Besluit van 31 januari 2024, houdende regels omtrent de gemeentelijke taak asielopvangvoorzieningen].
- Dijkstra, L., Poelman, H., & Veneri, P. (2019). *The EU-OECD definition of a functional urban area* (Data available at <https://www.oecd.org/cfe/regionaldevelopment/functional-urban-areas.htm>, accessed in June 2020.). Organization for Economic Development and Cooperation. OECD Publishing. <https://doi.org/10.1787/d58cb34d-en>
- DiPasquale, D., & Glaeser, E. L. (1998). The Los Angeles riot and the economics of urban unrest. *Journal of Urban Economics*, 43(1), 52–78.
- Dong, X., Morales, A. J., Jahani, E., Moro, E., Lepri, B., Bozkaya, B., Sarraute, C., Bar-Yam, Y., & Pentland, A. (2020). Segregated interactions in urban and online space. *EPJ Data Science*, 9, 1–22. <https://doi.org/10.1140/epjds/s13688-020-00238-7>
- Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2), 210–217.

- Duque, J. C., Ramos, R., & Suriñach, J. (2007). Supervised regionalization methods: A survey. *International Regional Science Review*, 30(3), 195–220.
- Durlauf, S. N. (1996). A theory of persistent income inequality. *Journal of Economic Growth*, 1, 75–93.
- Egede, L. E., Walker, R. J., Campbell, J. A., Linde, S., Hawks, L. C., & Burgess, K. M. (2023). Modern day consequences of historic redlining: Finding a path forward. *Journal of general internal medicine*, 38(6), 1534–1537.
- Ellis, M., Wright, R., Holloway, S., & Fiorio, L. (2018). Remaking white residential segregation: Metropolitan diversity and neighborhood change in the United States. *Urban Geography*, 39, 519–545. <https://doi.org/10.1080/02723638.2017.1360039>
- Erisen, C., & Kentmen-Cin, C. (2017). Tolerance and perceived threat toward Muslim immigrants in Germany and the Netherlands. *European Union Politics*, 18, 73–97. <https://doi.org/10.1177/1465116516675979>
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 96(34), 226–231.
- Farber, S., Neutens, T., Carrasco, J. A., & Rojas, C. (2014). Social interaction potential and the spatial distribution of face-to-face social interactions. *Environment and Planning B: Planning and Design*, 41, 960–976. <https://doi.org/10.1068/b120034p>
- Farrell, C. R., & Lee, B. A. (2011). Racial diversity and change in metropolitan neighborhoods. *Social Science Research*, 40, 1108–1123.
- Feitosa, F. F., Camara, G., Monteiro, A. M. V., Koschitzki, T., & Silva, M. P. (2007). Global and local spatial indices of urban segregation. *International Journal of Geographical Information Science*, 21(3), 299–323.
- Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics*, 31(7), 799–815.
- Florida, R. (2002). *The rise of the creative class: And how it's transforming work, leisure, community and everyday life* [See esp. Part I, Ch. 3–5 on diversity, tolerance, and cultural openness as urban assets]. Basic Books.
- Foley, L., Prins, R., Crawford, F., Humphreys, D., Mitchell, R., Sahlqvist, S., Thomson, H., Ogilvie, D., & Team, M. S. (2017). Effects of living near an urban motorway on the wellbeing of local residents in deprived areas: Natural experimental study. *Plos one*, 12(4), e0174882.
- Fossett, M. (2006). Ethnic preferences, social distance dynamics, and residential segregation: Theoretical explorations using simulation analysis. *Journal of Mathematical Sociology*, 30(3-4), 185–273.
- Fossett, M. (2017). *New methods for measuring and analyzing segregation*. Springer Nature.
- Fowler, C. S. (2016). Segregation as a multiscalar phenomenon and its implications for neighborhood-scale research: The case of South Seattle 1990-2010. *Urban Geography*, 37, 1–25.

- Gabriel, E. (2014). Estimating second-order characteristics of inhomogeneous spatio-temporal point processes: Influence of edge correction methods and intensity estimates. *Methodology and Computing in Applied Probability*, 16, 411–431.
- Galster, G., Tatian, P., & Accordino, J. (2006). Targeting investments for neighborhood revitalization. *Journal of the American Planning Association*, 72(4), 457–474.
- Galster, G. C. (2011). The mechanism (s) of neighbourhood effects: Theory, evidence, and policy implications. In *Neighbourhood effects research: New perspectives* (pp. 23–56). Springer.
- Gemeente Amsterdam. (2017). Woonagenda 2025 [<https://www.amsterdam.nl/en/policy/urban-development/living/>].
- Gemeente Den Haag. (2021). Segregatiemonitor (motie 1.3 sp) [<https://denhaag.raadsinformatie.nl/modules/6/Moties%20en%20amendementen%20en%20initiatieven/664268>].
- George, Y., Karunasekera, S., Harwood, A., & Lim, K. H. (2021). Real-time spatio-temporal event detection on geotagged social media. *Journal of Big Data*, 8(1), 91.
- Gordon, I., & Monastiriotis, V. (2006). Urban size, spatial segregation and inequality in educational outcomes. *Urban Studies*, 43, 213–236.
- Gracia, P., Vázquez-Quesada, L., & de Werfhorst, H. G. V. (2016). Ethnic penalties? the role of human capital and social origins in labour market outcomes of second-generation Moroccans and Turks in the Netherlands. *Journal of Ethnic and Migration Studies*, 42, 69–87. <https://doi.org/10.1080/1369183X.2015.1085800>
- Gramberg, P. (1998). School segregation: The case of Amsterdam. *Urban Studies*, 35(3), 547–564. <https://doi.org/10.1080/0042098984907>
- Grannis, R. (2009). From the ground up: Translating geography into community through neighbor networks. Princeton University Press.
- Granovetter, M. (1978). Threshold models of collective behavior. *American journal of sociology*, 83(6), 1420–1443.
- Gurr, T. R. (2015). Why men rebel. Routledge.
- Haandrikman, K., Costa, R., Malmberg, B., Rogne, A. F., & Sleutjes, B. (2023). Socio-economic segregation in European cities. a comparative study of Brussels, Copenhagen, Amsterdam, Oslo and Stockholm. *Urban Geography*, 44(1), 1–36.
- Haggett, P. (2000). *The geographical structure of epidemics*. Oxford University Press.
- Hartog, J., & Zorlu, A. (2009). Part 1: Ethnicity, religion, discrimination and exclusion: Ethnic segregation in the Netherlands: An analysis at neighbourhood level. *International Journal of Manpower*, 30, 15–25. <https://doi.org/10.1108/01437720910948366>
- Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1), 83–90.
- He, Z., Tao, L., Xie, Z., & Xu, C. (2020). Discovering spatial interaction patterns of near repeat crime by spatial association rules mining. *Scientific reports*, 10(1), 17262.

- Hedman, L., & Van Ham, M. (2011). Understanding neighbourhood effects: Selection bias and residential mobility. In *Neighbourhood effects research: New perspectives* (pp. 79–99). Springer.
- Hernández-Fuentes, A., & Monnier, A. (2022). Twitter as a source of information? practices of journalists working for the French national press. *Journalism Practice*, 16(5), 920–937.
- Hochstenbach, C., & Musterd, S. (2018). Gentrification and the suburbanization of poverty: Changing urban geographies through boom and bust periods. *Urban Geography*, 39, 26–53. <https://doi.org/10.1080/02723638.2016.1276718>
- Holdo, M., & Bengtsson, B. (2020). Marginalization and riots: A rationalistic explanation of urban unrest. *Housing, Theory and Society*, 37(2), 162–179.
- Home Office. (2011, October). An overview of recorded crimes and arrests resulting from disorder events in August 2011. Crown Copyright.
- Hoyt, H. (1939). *The structure and growth of residential neighborhoods in american cities*. US Government Printing Office.
- Hwang, J., & Ding, L. (2020). Unequal displacement: Gentrification, racial stratification, and residential destinations in philadelphia. *American Journal of Sociology*, 126(2), 354–406.
- Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2), 467–475. <https://doi.org/10.2307/2951620>
- Iyer, A., & Pryce, G. (2024). Theorising the causal impacts of social frontiers: The social and psychological implications of discontinuities in the geography of residential mix. *Urban Studies*, 61(5), 782–798.
- Jacobs, J. (1961). *The death and life of great American cities* [See esp. Ch. 2 on sidewalks and Ch. 7 on the generators of diversity]. Random House.
- Janssen, K. M. J., Cottineau, C., Kleinhans, R., & van Bueren, E. (2023). Gentrification and the origin and destination of movers: A systematic review. *Tijdschrift voor Economische en Sociale Geografie*. <https://doi.org/10.1111/tesg.12581>
- Janssen, K. M. J., Cottineau-Mugadza, C., Kleinhans, R., & Van Bueren, E. (2024). Spatial dynamics of incoming movers and the state-led gentrification process: The case of rotterdam. *Population, Space and Place*, 30(8), e2851.
- Jasper, J. M. (1998). The emotions of protest: Affective and reactive emotions in and around social movements. *Sociological forum*, 13(3), 397–424.
- Jongen, E., Been, J., Caminada, K., & Vethaak, H. (2024). Labour market and income inequalities in the Netherlands, 1977–2022. *Fiscal Studies*, 45(3), 343–357.
- Joshi, A., Horn, B. P., & Berrens, R. P. (2024). Contemporary differences in residential housing values along historic redlining boundaries. *Real Estate Economics*, 52(2), 514–551.
- Kauppinen, T. M., & van Ham, M. (2019). Unravelling the demographic dynamics of ethnic residential segregation. *Population, Space and Place*, 25. <https://doi.org/10.1002/psp.2193>

- Kawalerowicz, J., & Biggs, M. (2015). Anarchy in the uk: Economic deprivation, social disorganization, and political grievances in the London riot of 2011. *Social forces*, 94(2), 673–698.
- Kirkley, A. (2022). Spatial regionalization based on optimal information compression. *Communications Physics*, 5(1), 249.
- Koff, H., & Duprez, D. (2009). The 2005 riots in france: The international impact of domestic violence. *Journal of ethnic and migration studies*, 35(5), 713–730.
- Kramer, R. (2017). Defensible spaces in Philadelphia: Exploring neighborhood boundaries through spatial analysis. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 3(2), 81–101.
- Kramer, R. (2018). Testing the role of barriers in shaping segregation profiles: The importance of visualizing the local neighborhood. *Environment and Planning B: Urban Analytics and City Science*, 45(6), 1106–1121.
- Krupka, D. J. (2007). Are big cities more segregated? neighbourhood scale and the measurement of segregation. *Urban Studies*, 44, 187–197.
- Lachman, S. J. (1996). Psychological perspective for a theory of behavior during riots [PMID: 8969079]. *Psychological Reports*, 79(3), 739–744. <https://doi.org/10.2466/pr0.1996.79.3.739>
- Lan, T., Kandt, J., & Longley, P. (2020). Geographic scales of residential segregation in English cities. *Urban Geography*, 41, 103–123. <https://doi.org/10.1080/02723638.2019.1645554>
- Lapeyronnie, D. (2006). Révolte primitive dans les banlieues françaises: Essai sur les émeutes de l'automne 2005. *Déviance et société*, 30(4), 431–448.
- Laurence, J., Schmid, K., Rae, J. R., & Hewstone, M. (2019). Prejudice, contact, and threat at the diversity-segregation nexus: A cross-sectional and longitudinal analysis of how ethnic out-group size and segregation interrelate for inter-group relations. *Social Forces*, 97, 1029–1065. <https://doi.org/10.1093/sf/soy079>
- Le Parisien. (2023, June). Mort de Nahel : Des scènes de pillages et de saccages en plein Paris [<https://www.leparisien.fr/faits-divers/mort-de-nahel-a-paris-des-scenes-de-pillages-et-de-saccages-30-06-2023-AZNCXGFSUFHQLL2STJBMB3KMEU.php>, last accessed: September 2024].
- Leclerc, C., Vink, M., & Schmeets, H. (2022). Does residential context matter? neighborhood migrant concentration and citizenship acquisition in the Netherlands. *International Migration Review*. <https://doi.org/10.1177/01979183221115165>
- Levy, G., & Razin, R. (2019). Echo chambers and their effects on economic and political outcomes. *Annual Review of Economics*, 11(1), 303–328. <https://doi.org/10.1146/annurev-economics-080218-030343>
- L'obs. (2023, June). Mort de Nahel à Nanterre : Macron va présider une cellule de crise après une troisième nuit d'émeutes [<https://www.nouvelobs.com/societe/20230630.OBS75142/mort-de-nahel-a-nanterre-macron-va-presider-une-cellule-de-crise-apres-une-troisieme-nuit-d-emeutes.html>, last accessed: September 2024].

- Logan, J. R., & Stults, B. J. (2011). The persistence of segregation in the metropolis: New findings from the 2010 census. *Census brief prepared for Project US2010*, 24.
- Ludwig, J., Duncan, G. J., Genetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2012). Neighborhood effects on the long-term well-being of low-income adults. *Science*, 337.
- Luttik, J. (2000). The value of trees, water and open space as reflected by house prices in the netherlands. *Landscape and urban planning*, 48(3-4), 161–167.
- MacDonald, H., Galster, G., & Dufty-Jones, R. (2018). The geography of rental housing discrimination, segregation, and social exclusion: New evidence from sydney. *Journal of Urban Affairs*, 40(2), 226–245.
- Mahajan, A. (2024). Highways and segregation. *Journal of Urban Economics*, 141, 103574. <https://doi.org/https://doi.org/10.1016/j.jue.2023.103574>
- Maharaj, B. (2020). Urban and regional planning and development: 20th century forms and 21st century transformations. Springer.
- Malmberg, B., Andersson, E., & Östh, J. (2013). Segregation and urban unrest in Sweden. *Urban geography*, 34(7), 1031–1046.
- Marcińczak, S., Mooses, V., Strömgren, M., & Tammaru, T. (2021). A comparative study of immigrant-native segregation at multiple spatial scales in urban Europe. *Journal of Ethnic and Migration Studies*, 1–23.
- Marcińczak, S., Mooses, V., Strömgren, M., & Tammaru, T. (2023). A comparative study of immigrant-native segregation at multiple spatial scales in urban Europe. *Journal of Ethnic and Migration Studies*, 49(1), 43–65.
- Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social forces*, 67(2), 281–315.
- Massey, D. S., & Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass*. Harvard University Press.
- Mayor of London. (2021). The London Plan: The spatial development strategy for Greater London [<https://www.london.gov.uk/what-we-do/planning/london-plan/new-london-plan/london-plan-2021>].
- Milias, V., & Psyllidis, A. (2022). Measuring spatial age segregation through the lens of co-accessibility to urban activities. *Computers, Environment and Urban Systems*, 95, 101829.
- Mitchell, R., & Lee, D. (2014). Is there really a “wrong side of the tracks” in urban areas and does it matter for spatial analysis? *Annals of the Association of American Geographers*, 104(3), 432–443.
- Morales, A. J., Dong, X., Bar-Yam, Y., & Pentland, A. (2019). Segregation and polarization in urban areas. *Royal Society Open Science*, 6. <https://doi.org/10.1098/rsos.190573>
- Moran, M., & Waddington, D. (2016). Riots: An international comparison. Palgrave Macmillan London.
- Moreno-Monroy, A. I., & Veneri, P. (2018). Divided cities: Understanding intra-urban inequalities. OECD Publishing. <https://doi.org/https://doi.org/https://doi.org/10.1787/9789264300385-en>

- Müllner, D. (2011). Modern hierarchical, agglomerative clustering algorithms. *arXiv*. <http://arxiv.org/abs/1109.2378>
- Murray, G. (2006). France: The riots and the Republic. *Race & Class*, 47(4), 26–45.
- Musterd, S., Marcińczak, S., Van, M., & Tammaru, H. T. (2015). *Socio-economic segregation in European capital cities: Increasing separation between poor and rich* (tech. rep.). Institute of the Study of Labor, IZA.
- Musterd, S., Marcińczak, S., van Ham, M., & Tammaru, T. (2017). Socioeconomic segregation in European capital cities. increasing separation between poor and rich. *Urban Geography*, 38(7), 1062–1083. <https://doi.org/10.1080/02723638.2016.1228371>
- Myers, D. J. (2000). The diffusion of collective violence: Infectiousness, susceptibility, and mass media networks. *American Journal of Sociology*, 106(1), 173–208.
- Nägel, C., & Lutter, M. (2023). The 2017 French riots and trust in the police: A quasi-experimental approach. *European journal of criminology*, 20(1), 270–291.
- Natale, F., Scipioni, M., & Alessandrini, A. (2018). Divided cities: Understanding intra-urban inequalities. OECD Publishing. <https://doi.org/https://doi.org/https://doi.org/10.1787/9789264300385-en>
- Newburn, T. (2016). The 2011 England riots in European context: A framework for understanding the ‘life-cycle’ of riots. *European Journal of Criminology*, 13(5), 540–555.
- Newburn, T. (2021). The causes and consequences of urban riot and unrest. *Annual Review of Criminology*, 4(1), 53–73.
- Nicoletti, L., Verma, T., Warnier, M., & Santi, P. (2021). *Exploring the impact of hashtag activism on political processes using geolocated social media data* [Master’s thesis, TU Delft] [<https://repository.tudelft.nl/record/uuid:a4e3c609-c4ea-48cb-b268-d5b77549efe6>, last accessed: January 2025].
- Nielsen, M. M., & Hennerdal, P. (2017). Changes in the residential segregation of immigrants in Sweden from 1990 to 2012: Using a multi-scalar segregation measure that accounts for the modifiable areal unit problem. *Applied Geography*, 87, 73–84. <https://doi.org/10.1016/j.apgeog.2017.08.004>
- Nieuwenhuis, J., Tammaru, T., van Ham, M., Hedman, L., & Manley, D. (2020). Does segregation reduce socio-spatial mobility? Evidence from four European countries with different inequality and segregation contexts. *Urban Studies*, 57, 176–197. <https://doi.org/10.1177/0042098018807628>
- Nijman, J. (2007). Introduction - comparative urbanism. *Urban Geography*, 28, 1–6. <https://doi.org/10.2747/0272-3638.28.1.1>
- Noonan, D. S. (2005). Neighbours, barriers and urban environments: Are things ‘different on the other side of the tracks’? *Urban Studies*, 42(10), 1817–1835. <https://doi.org/10.1080/00420980500231720>
- OECD. (2018). *Divided cities: Understanding intra-urban inequalities*. <https://doi.org/https://doi.org/https://doi.org/10.1787/9789264300385-en>
- Olnér, D., Pryce, G., van Ham, M., & Janssen, H. (2024). The conflicting geographies of social frontiers: Exploring the asymmetric impacts of social frontiers on

- household mobility in rotterdam. *Environment and Planning B: Urban Analytics and City Science*, 51(3), 625–640.
- Olteanu, M., Hazan, A., Cottrell, M., & Randon-Furling, J. (2020). Multidimensional urban segregation: Toward a neural network measure. *Neural Computing and Applications*, 32, 18179–18191.
- Olteanu, M., Randon-Furling, J., & Clark, W. A. (2019). Segregation through the multiscalar lens. *Proceedings of the National Academy of Sciences*, 116(25), 12250–12254.
- Openshaw, S. (1981). The modifiable areal unit problem. *Quantitative geography: a British view*, 60–69. <https://cir.nii.ac.jp/crid/1572824498971908736>
- OpenStreetMap. (2021). OpenStreetMap data [Last accessed in June 2021, more info at <https://www.openstreetmap.org/>].
- Östh, J., Clark, W. A., & Malmberg, B. (2015). Measuring the scale of segregation using k-nearest neighbor aggregates. *Geographical Analysis*, 47, 34–49. <https://doi.org/10.1111/gean.12053>
- Owen, S. (2017). Monitoring social media and protest movements: Ensuring political order through surveillance and surveillance discourse. *Social Identities*, 23(6), 688–700.
- Owens, A. (2018). Income segregation between school districts and inequality in students' achievement. *Sociology of Education*, 91, 1–27. <https://doi.org/10.1177/0038040717741180>
- Panagiotopoulos, P., Barnett, J., Bigdeli, A. Z., & Sams, S. (2016). Social media in emergency management: Twitter as a tool for communicating risks to the public. *Technological Forecasting and Social Change*, 111, 86–96.
- Park, R. E., Burgess, E. W., & McKenzie, R. D. (1925). *The city* [See Ch. 1 for the city as a social laboratory; also focuses on group competition and human ecology for understanding segregation as a manifestation of social tension]. University of Chicago Press.
- Petrović, A., van Ham, M., & 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. <https://doi.org/10.1080/24694452.2017.1411245>
- Pires, B., & Crooks, A. T. (2017). Modeling the emergence of riots: A geosimulation approach. *Computers, Environment and Urban Systems*, 61, 66–80.
- Prabhu, A., Guhathakurta, D., Subramanian, M., Reddy, M., Sehgal, S., Karandikar, T., Gulati, A., Arora, U., Shah, R. R., & Kumaraguru, P. (2021). Capitol (pat) riots: A comparative study of Twitter and Parler. *arXiv preprint arXiv:2101.06914*.
- Procter, R., Vis, F., & Voss, A. (2013). Reading the riots on Twitter: Methodological innovation for the analysis of big data. *International journal of social research methodology*, 16(3), 197–214.
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.

- PyTorch Developers. (2024). Torchvision: Datasets, transforms, and models specific to computer vision [https://github.com/pytorch/vision. Last accessed in January 2025.].
- Reardon, S. F., & Bischoff, K. (2011). Income inequality and income segregation. *American Journal of Sociology*, 116, 1092–1153. https://doi.org/10.1086/657114
- Reardon, S. F., & Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology*, 32, 33–67.
- Reardon, S. F., & O’Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology*, 34, 121–162.
- Recht, B., Roelofs, R., Schmidt, L., & Shankar, V. (2019). Do imagenet classifiers generalize to imagenet? *International conference on machine learning*, 5389–5400.
- Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of applied probability*, 13(2), 255–266.
- Roberto, E., & Hwang, J. (2017). Barriers to integration: Institutionalized boundaries and the spatial structure of residential segregation. *Annual Meeting of the Population Association of America, Chicago, Illinois*.
- Roberto, E., & Korver-Glenn, E. (2021). The spatial structure and local experience of residential segregation. *Spatial Demography*, 9(3), 277–307.
- Robinson, J. (2011). Cities in a world of cities: The comparative gesture. *International Journal of Urban and Regional Research*, 35, 1–23. https://doi.org/10.1111/j.1468-2427.2010.00982.x
- Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing “neighborhood effects”: Social processes and new directions in research. *Annual review of sociology*, 28(1), 443–478.
- San Millán, J., Cottineau-Mugadza, C., & Van Ham, M. (2025). From flux to capital: Distinguishing patterns of income and wealth segregation in the Netherlands. *Population, Space and Place*, 31(8), e70127.
- Scarpa, S. (2015). The impact of income inequality on economic residential segregation: The case of Malmö, 1991–2010. *Urban Studies*, 52, 906–922. https://doi.org/10.1177/0042098014529347
- Schaeffer, Y., Cremer-Schulte, D., Tartiu, C., & Tivadar, M. (2016). Natural amenity-driven segregation: Evidence from location choices in french metropolitan areas. *Ecological Economics*, 130, 37–52.
- Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, 59, 488–493.
- Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology*, 1(2), 143–186.
- Schierup, C.-U., Ålund, A., & Kings, L. (2014). Reading the Stockholm riots – a moment for social justice? *Race & Class*, 55(3), 1–21.
- Schläpfer, M., Dong, L., O’Keeffe, K., Santi, P., Szell, M., Salat, H., Anklesaria, S., Vazifeh, M., Ratti, C., & West, G. B. (2021). The universal visitation law of human mobility. *Nature*, 593, 522–527. https://doi.org/10.1038/s41586-021-03480-9

- Semyonov, M., & Glikman, A. (2008). Ethnic Residential Segregation, Social Contacts, and Anti-Minority Attitudes in European Societies. *European Sociological Review*, 25(6), 693–708. <https://doi.org/10.1093/esr/jcn075>
- Sénat. (2024). Rapport d'information fait au nom de la commission des lois constitutionnelles, de législation, du suffrage universel, du règlement et d'administration générale (1), investie des pouvoirs d'une commission d'enquête, sur les émeutes survenues à compter du 27 juin 2023. *Session ordinaire de 2023-2024*, 10, 162. <https://www.senat.fr/rap/r23-521/r23-5211.pdf>.
- Skifter Andersen, H., Andersson, R., Wessel, T., & Vilka, K. (2016). The impact of housing policies and housing markets on ethnic spatial segregation: Comparing the capital cities of four nordic welfare states. *International Journal of Housing Policy*, 16(1), 1–30.
- Sleutjes, B., Ooijevaar, J., & de Valk, H. A. (2019). Residential segregation in the Amsterdam metropolitan region: A longitudinal analysis using scalable individualised neighbourhoods. *Tijdschrift voor Economische en Sociale Geografie*, 110, 359–377. <https://doi.org/10.1111/tesg.12356>
- Smith, M., & Miller, S. (2022). The ethical application of biometric facial recognition technology. *Ai & Society*, 37(1), 167–175.
- Sousa, S., & Nicosia, V. (2022). Quantifying ethnic segregation in cities through random walks. *Nature Communications*, 13(1), 5809.
- Spierenburg, L., van Cranenburgh, S., & Cats, O. (2022). A regionalization method filtering out small-scale spatial fluctuations. *AGILE: GIScience Series*, 3, 61. <https://doi.org/10.5194/agile-giss-3-61-2022>
- Spierenburg, L., Cats, O., & van Cranenburgh, S. (2024). Data related to the paper "Studying social unrest through the lens of social media" [Accessible at <https://doi.org/10.4121/649e8f5d-8e40-4ab7-9d07-b5ef53d810f0.v1>].
- Spierenburg, L., van Cranenburgh, S., & Cats, O. (2023). Characterizing residential segregation in cities using intensity, separation, and scale indicators. *Computers, Environment and Urban Systems*, 103, 101990. <https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2023.101990>
- Spierenburg, L., van Cranenburgh, S., & Cats, O. (2024). Evolution of residential segregation patterns in the Netherlands between 2015 and 2020. *Cities*, 150, 105038. <https://doi.org/https://doi.org/10.1016/j.cities.2024.105038>
- Stodden, V. (2014). Privacy, big data, and the public good: Frameworks for engagement. Cambridge University Press.
- Storper, M., & Scott, A. J. (2016). Current debates in urban theory: A critical assessment. *Urban Studies*, 53, 1114–1136. <https://doi.org/10.1177/0042098016634002>
- Tammaru, T., van Ham, M., Marcińczak, S., & Musterd, S. (2015). *Socio-economic segregation in European capital cities: East meets west*. Routledge.
- Tan, M. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*.
- Taylor, C., & Gorard, S. (2001). The role of residence in school segregation: Placing the impact of parental choice in perspective. *Environment and Planning A: Economy and Space*, 33(10), 1829–1852. <https://doi.org/10.1068/a34123>

- the United Nations. (2017). *New Urban Agenda* (<https://habitat3.org/the-new-urban-agenda>). UN-Habitat.
- Theebe, M. A. (2004). Planes, trains, and automobiles: The impact of traffic noise on house prices. *The Journal of Real Estate Finance and Economics*, 28(2), 209–234.
- 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*, 1(2), 187–193. <https://doi.org/10.1080/0022250X.1971.9989795>
- Theodoridis, S., & Koutroumbas, K. (2008). Pattern recognition. Academic Press.
- Thijssen, L., Lancee, B., Veit, S., & Yemane, R. (2021). Discrimination against Turkish minorities in Germany and the Netherlands: Field experimental evidence on the effect of diagnostic information on labour market outcomes. *Journal of Ethnic and Migration Studies*, 47, 1222–1239. <https://doi.org/10.1080/1369183X.2019.1622793>
- Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of political economy*, 64(5), 416–424.
- Till, J. (2013). The broken middle: The space of the london riots. *Cities*, 34, 71–74.
- Tilly, C. (2003). *The politics of collective violence* [See Ch. 1-2 on the framework of contentious politics and Ch. 3 on riots as episodic manifestations of underlying social tensions]. Cambridge University Press.
- Timoneda, J. C. (2018). Where in the world is my tweet: Detecting irregular removal patterns on Twitter. *PloS one*, 13(9), e0203104.
- Tonkin, E., Pfeiffer, H. D., & Tourte, G. (2012). Twitter, information sharing and the London riots? *Bulletin of the American Society for Information Science and Technology*, 38(2), 49–57.
- Tóth, G., Wachs, J., Clemente, R. D., Jakobi, Á., Ságvári, B., Kertész, J., & Lengyel, B. (2021). Inequality is rising where social network segregation interacts with urban topology. *Nature Communication*, 12. <https://doi.org/10.1038/s41467-021-21465-0>
- Trounstine, J. (2020). The geography of inequality: How land use regulation produces segregation. *American Political Science Review*, 114(2), 443–455.
- Tselios, V., Noback, I., van Dijk, J., & McCann, P. (2015). Integration of immigrants, bridging social capital, ethnicity, and locality. *Journal of Regional Science*, 55(3), 416–441.
- United States Census Bureau. (2025). Glossary [<https://www.census.gov/programs-surveys/geography/about/glossary.html>. Last accessed in September 2025.].
- van Leeuwen, N. (2020). *Statistische gegevens per vierkant en postcode 2019–2018–2017* [Retrieved from <https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode> in June 2021.]. Centraal Bureau voor de Statistiek.
- van de Werfhorst, H. G., & Heath, A. (2019). Selectivity of migration and the educational disadvantages of second-generation immigrants in ten host societies. *European Journal of Population*, 35, 347–378. <https://doi.org/10.1007/s10680-018-9484-2>

- van der Gref, S., & Fortuijn, J. D. (2017). Multiple disadvantage of older migrants and native Dutch older adults in deprived neighbourhoods in Amsterdam, the Netherlands: A life course perspective. *GeoJournal*, 82, 415–432. <https://doi.org/10.1007/s10708-015-9691-x>
- van der Laan Bouma-Doff, W. (2007). Confined contact: Residential segregation and ethnic bridges in the Netherlands. *Urban Studies*, 44(5-6), 997–1017.
- Veneri, P., Comandon, A., García-López, M. À., & Daams, M. N. (2021). What do divided cities have in common? an international comparison of income segregation. *Journal of Regional Science*, 61, 162–188. <https://doi.org/10.1111/jors.12506>
- Vervoort, M. (2012). Ethnic concentration in the neighbourhood and ethnic minorities' social integration: Weak and strong social ties examined. *Urban Studies*, 49, 897–915. <https://doi.org/10.1177/0042098011408141>
- Ville de Paris. (2021). Charte pour un aménagement durable et inclusif de Paris [<https://cdn.paris.fr/paris/2021/03/02/343e9528c3a88309b30920627c486c97.pdf>].
- Vis, F. (2013). Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 UK riots. *Digital journalism*, 1(1), 27–47.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Waddington, D. (2012a). The law of moments: Understanding the flashpoint that ignited the riots: David waddington on how the shooting of mark duggan led to the august rioting. *Criminal Justice Matters*, 87(1), 6–7.
- Waddington, D. (2012b). *Rioting in the UK and France: A comparative analysis* [See Ch. 5–7 for empirical examples of riots as expressions of social tension in urban contexts]. Routledge.
- Waddington, D. (2012c). *Rioting in the UK and France: A comparative analysis* [See Ch. 2–4 for analysis of riots as short-lived, localized, and episodic events in urban areas]. Routledge.
- Wang, Q., Phillips, N. E., Small, M. L., & Sampson, R. J. (2018). Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 7735–7740. <https://doi.org/10.1073/pnas.1802537115>
- Ward, K. (2010). Towards a relational comparative approach to the study of cities. *Progress in Human Geography*, 34, 471–487. <https://doi.org/10.1177/0309132509350239>
- Ward Jr., J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Watson, T. (2009). Inequality and the measurement of residential segregation by income in American neighborhoods. *Review of Income and Wealth*, 55(3), 820–844.
- Weiss, K., Khoshgoftaar, T. M., & Wang, D. D. (2016). A survey of transfer learning. *Journal of Big Data*, 3. <https://doi.org/10.1186/s40537-016-0043-6>
- White, M. J. (1983). The measurement of spatial segregation. *American journal of sociology*, 88(5), 1008–1018.
- Wilkinson, S. I. (2009). Riots. *Annual Review of Political Science*, 12, 329–343.

- Williams, D. R., & Collins, C. (2001). Racial residential segregation: a fundamental cause of racial disparities in health. *Public Health Reports*, 116, 404–416.
- Wilson, S. L. (2017). Detecting mass protest through social media. *The Journal of Social Media in Society*, 6(2), 5–25.
- Wilson, W. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. University of Chicago Press.
- Wong, D. W. S. (2004). Comparing traditional and spatial segregation measures: A spatial scale perspective. *Urban Geography*, 25(1), 66–82.
- Wright, R., Ellis, M., Holloway, S. R., & Wong, S. (2014). Patterns of racial diversity and segregation in the United States: 1990–2010. *Professional Geographer*, 66, 173–182. <https://doi.org/10.1080/00330124.2012.735924>
- Zimmer, F., Scheibe, K., Stock, M., & Stock, W. G. (2019). Fake news in social media: Bad algorithms or biased users? *Journal of Information Science Theory and Practice*, 7(2), 40–53.
- Zorlu, A., & Mulder, C. H. (2008). Initial and subsequent location choices of immigrants to the Netherlands. *Regional Studies*, 42, 245–264. <https://doi.org/10.1080/00343400601145210>
- Zuccotti, C. V., Lorenz, J., Paolillo, R., Sánchez, A. R., & Serka, S. (2024). Exploring the dynamics of neighborhood ethnic segregation with agent-based modelling: An empirical application to Bradford, UK. In *Computational research in ethnic and migration studies* (pp. 171–192). Routledge.
- Zuk, M., Bierbaum, A. H., Chapple, K., Gorska, K., & Loukaitou-Sideris, A. (2018). Gentrification, displacement, and the role of public investment. *Journal of Planning Literature*, 33(1), 31–44. <https://doi.org/10.1177/0885412217716439>
- Zwiers, M., van Ham, M., & Manley, D. (2018). Trajectories of ethnic neighbourhood change: Spatial patterns of increasing ethnic diversity. *Population, Space and Place*, 24.

Acknowledgements

The completion of this thesis marks the end of a journey that was as much about personal and professional growth as it was about scientific inquiry. While research is often viewed as a solitary endeavor, this work would not have been possible without the guidance, inspiration, and support of a diverse community of mentors, colleagues, and loved ones.

My sincere thanks go to Oded Cats for his organized and steadfast supervision. His commitment to academic rigor and his help in navigating the hurdles of the publication process provided the structure I needed to see this thesis through to its conclusion. I am grateful for the professional standards he instilled in me. My gratitude goes to Sander van Cranenburgh for his energetic and ambitious supervision. His focus on the underlying relevance of our research and his skill in drafting structured, impactful introductions have been invaluable to my development. I am especially thankful for his patience and expertise as I ventured into computer vision, and for his persistent drive to ensure our work reached its highest potential before submission.

Beyond my doctoral supervision, my path into research was paved by several key figures. I owe my foundational 'working hygiene' and a lack of intimidation toward complex problems to my high school mathematics teacher, Sébastien Jarraud. His rigorous insistence on mastering theorems and demonstrations taught me that persistence is the primary tool of any scientist. My transition into academia, however, was sparked at TU Delft within the Transport and Planning department. My Master's supervisor, Niels van Oort, provided an environment of trust and encouragement that I hadn't anticipated; by inviting me to present my findings to policymakers in Amsterdam, he showed me the real-world impact of research. Finally, my shift toward the social dimensions of engineering was inspired by Joaquim Oliveira Martins, whose insights into urban segregation and social tension gave my technical work a sense of urgent, human purpose.

While this thesis bears my name, the conceptual seeds were planted by others. I am particularly indebted to Madalina Olteanu, whose work on residential segregation provided the methodological blueprint for my use of unsupervised learning and regionalization. Similarly, the work of Balázs Lengyel on the intersection of segregation and urban fragmentation was the direct inspiration for the fourth chapter of this research. On a more practical note, I must thank my optician, Adrien Kauffman; his smart use of Snapchat to track urban unrest for his shop's safety in Paris provided the unexpected spark for the fifth chapter of my thesis. Finally, I owe an innovative way of sharing results to Leonardo Nicoletti, whose work as a data journalist challenged me to look beyond traditional text-based science communication and embrace web-based data storytelling.

To my colleagues in the Transport and Planning department Jesper, Matthew,

Nirvana, Jaime, Arjan, Lucia, Leonor, Yiru, Abhi, Arco, Callum, Ting, Nejc, Julia, Renzo, Alex, Nagarjun, and Sherman, thank you for the countless discussions, whether over lunch or in the office, that kept the work moving forward. I am also grateful to the 'roommates' of room A3.080 at the Faculty of Technology, Policy and Management: Francisco, Gabriel, Lion, Ali, Julien, and Floris. Thank you for welcoming me into your space and letting me borrow a desk whenever I needed a change of scenery.

Finally, my deepest thanks go to my partner Ieva, and to my mother, father, and sister. This thesis is the result of years of work, but it was your daily emotional support that made the process possible. I couldn't have done it without you.

Lucas Johannes SPIERENBURG
City, Date Month Year

About the Author

Lucas Spierenburg was born in Paris, France, in 1996. He received his Bachelor's degree in Microengineering in 2017 and his Master's degree in Energy Management and Sustainability in 2020, both from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. During his Master's studies, a research stay at the Transport and Planning department at TU Delft—where he conducted an accessibility impact assessment of the Amsterdam Noord-Zuidlijn metro line—shifted his career trajectory toward academia. Encouraged by the real-world impact of presenting his findings to policymakers in Amsterdam, he decided to pursue a doctoral degree.



In March 2020, Lucas joined the Department of Transport and Planning at the Faculty of Civil Engineering and Geosciences, TU Delft, as a PhD candidate within the CityAI Lab. Throughout his PhD, he developed a diverse methodological toolkit, including the application of unsupervised learning and regionalization to characterize segregation patterns across Dutch municipalities. A notable highlight of his work was the development of a computer vision model to track the spatiotemporal distribution of the 2023 riots in France using geolocated social media data, an approach inspired by real-world observations of digital footprints during urban crises.

Lucas has published his research in international journals such as *Cities*, *Journal of Transport Geography*, and *Computers, Environment and Urban Systems*. He has presented his work at several international venues, including the 11th International Conference on Computational Social Science (IC2S2) and the AGILE Conference on Geographic Information Science.

List of publications

Peer-reviewed journal articles

1. L. Spierenburg, E. Ralon-Santizo, S. van Cranenburgh, O. Cats, *Does residential segregation align with urban barriers?*, Scientific reports (in press).
2. L. Spierenburg, S. van Cranenburgh, O. Cats, *Studying riots through the lens of social media*, Journal of Big Data, **12** (1), 1-20 (2025)
3. L. Spierenburg, H. van Lint, N. van Oort, *Synergizing cycling and transit: Strategic placement of cycling infrastructure to enhance job accessibility*, Journal of Transport Geography, **116**, 103861 (2024)
4. L. Spierenburg, S. van Cranenburgh, O. Cats, *Evolution of residential segregation patterns in the Netherlands between 2015 and 2020*, Cities, **150**, 105038 (2024)
5. L. Spierenburg, S. van Cranenburgh, O. Cats, *Characterizing residential segregation in cities using intensity, separation, and scale indicators*, Computers, Environment and Urban Systems, **103**, 101990 (2023)

Submitted papers

1. B. van Horik, N. el Khatibi, M. van Galen, L. Spierenburg, *A privacy-preserving approach to identify riot-related footage on social media*

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series, For a complete overview of more than 400 titles see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Spierenburg, L., *Advances in the Analysis of Residential Segregation and Urban Riots*, T2026/9, April 2020, TRAIL Thesis Series, The Netherlands

Boot, M., *Evaluating Experiences with Smart Cycling Technologies: Sensor-based evaluations of outdoor cycling experiences with Smart Cycling Technologies*, T2026/8, March 2026, TRAIL Thesis Series, the Netherlands

Wen, X., *Data-Driven Spatial-Temporal Modeling for Bicycle Traffic Prediction*, T2026/7, March 2026, TRAIL Thesis Series, the Netherlands

Wang, Z., *Optimising Performance of Automatic Train Operation on Railway Networks*, T2026/6, March 2026, TRAIL Thesis Series, the Netherlands

Hadi, A.H., *DEM Modelling of Multi-Component Segregation in the Blast Furnace Charging System*, T2026/5, February 2026, TRAIL Thesis Series, the Netherlands

Farhani, M., *Demand Management Strategies for Operations of Shared Mobility Services*, T2026/4, February 2026, TRAIL Thesis Series, the Netherlands

Yao, X., *Driving Heterogeneity in Traffic Flow Theory: An action-based framework for identification, modelling, and simulation*, T2026/3, January 2026, TRAIL Thesis Series, the Netherlands

Versluis, N.D., *Optimising Railway Traffic Management under Radio-Based Distance-to-Go Signalling*, T2026/2, January 2026, TRAIL Thesis Series, the Netherlands

Jiao, Y., *Proactive Collision Risk Quantification in Multi-directional Traffic Interactions*, T2026/1, January 2026, TRAIL Thesis Series, the Netherlands

Asadi, M., *Accessibility and Road Safety: Integration of road safety in accessibility evaluation*, T2025/20, November 2025, TRAIL Thesis Series, the Netherlands

Akse, R., *Understanding and untangling the uncertainty knot: How to catalyse decision-*

making in mobility innovations, T2025/19, November 2025, TRAIL Thesis Series, the Netherlands

Führer, K., *Participatory Decision-making under Deep Uncertainty: Modelling mobility transitions*, T2025/18, November 2025, TRAIL Thesis Series, the Netherlands

Picco, A., *Monitoring and Feedback in Driving*, T2025/17, October 2025, TRAIL Thesis Series, the Netherlands

Cebeci, M.S., *Behaviour of Prosumers in Last-mile Logistics: The case of crowdshipping*, T2025/16, September 2025, TRAIL Thesis Series, the Netherlands

Kuijpers, A., *Enabling Inter-Organizational Collaboration Through Platforms: The role of trust*, T2025/15, September 2025, TRAIL Thesis Series, the Netherlands

Song, R., *Human-MASS Interaction in Decision-Making for Safety and Efficiency in Mixed Waterborne Transport Systems*, T2025/14, June 2025, TRAIL Thesis Series, the Netherlands

Destyanto, A.R., *A Method for Evaluating Port Resilience in an Archipelago*, T2025/13, June 2025, TRAIL Thesis Series, the Netherlands

Karademir, C., *Synchronized Two-echelon Routing Problems: Exact and approximate methods for multimodal city logistics*, T2025/12, May 2025, TRAIL Thesis Series, the Netherlands

Vial, A., *Eyes in Motion: A new traffic sensing paradigm for pedestrians and cyclists*, T2025/11, May 2025, TRAIL Thesis Series, the Netherlands

Chen, Q., *Towards Mechanical Intelligence in Soft Robotics: Model-based design of mechanically intelligent structures*, T2025/10, April 2025, TRAIL Thesis Series, the Netherlands

Eftekhari, Z., *Exploring the Spatial and Temporal Patterns in Travel Demand: A data-driven approach*, T2025/9, June 2025, TRAIL Thesis Series, the Netherlands

Reddy, N., *Human Driving Behavior when Interacting with Automated Vehicles and the Implications on Traffic Efficiency*, T2025/8, May 2025, TRAIL Thesis Series, the Netherlands

Durand, A., *Lost in Digitalisation? Navigating public transport in the digital era*, T2025/7, May 2025, TRAIL Thesis Series, the Netherlands

Dong, Y., *Safe, Efficient, and Socially Compliant Automated Driving in Mixed Traffic: Sensing, Anomaly Detection, Planning and Control*, T2025/6, May 2025, TRAIL Thesis Series

Series, the Netherlands

Droffelaar, I.S. van, *Simulation-optimization for Fugitive Interception*, T2025/5, May 2025, TRAIL Thesis Series, the Netherlands

Fan, Q., *Fleet Management Optimisation for Ride-hailing Services: from mixed traffic to fully automated environments*, T2025/4, April 2025, TRAIL Thesis Series, the Netherlands

Hagen, L. van der, *Machine Learning for Time Slot Management in Grocery Delivery*, T2025/3, March 2025, TRAIL Thesis Series, the Netherlands

Schilt, I.M. van, *Reconstructing Illicit Supply Chains with Sparse Data: a simulation approach*, T2025/2, January 2025, TRAIL Thesis Series, the Netherlands

Ruijter, A.J.F. de, *Two-Sided Dynamics in Ridesourcing Markets*, T2025/1, January 2025, TRAIL Thesis Series, the Netherlands



Summary

This dissertation investigates the spatial structure of social tensions in cities, examining both the chronic patterns of residential segregation and the acute eruptions of urban riots. By developing innovative analytical methods ranging from regionalization frameworks to computer vision for social media analysis, the research reveals how these phenomena manifest in space. Key findings demonstrate that residential segregation in Europe does not systematically align with physical urban barriers; furthermore, the work uncovers the highly localized and short-lived nature of unrest events during the 2023 French riots.

About the Author

Lucas Spierenburg holds a Master's from EPFL and a PhD from TU Delft's CityAI Lab. His research leverages unsupervised learning and computer vision to analyze urban unrest and segregation. Lucas has published several papers at the intersection of machine learning and urban science in leading journals.

TRAIL Research School ISBN 978-90-5584-384-8



Radboud University



rijksuniversiteit
 groningen



UNIVERSITY OF TWENTE.

TU/e

Technische Universiteit
 Eindhoven
 University of Technology