# Frontiers in residential segregation: Understanding neighbourhood boundaries & their impacts

*Nema Dean,* Guanpeng Dong,**Aneta Piekut,*** & Gwilym Pryce ****

* School of Mathematics & Statistics, University of Glasgow
** School of Environmental Sciences, University of Liverpool
*** Sheffield Methods Institute, University of Sheffield

*Working Paper, 2nd Feb 2018*

## Abstract

"Social frontiers"—places of sharp difference in social/ethnic characteristics between neighbouring communities—have largely been overlooked in quantitative research. Advancing this nascent field first requires a way of identifying social frontiers in a robust way. Such frontiers may be "open"—an area may contrast sharply with a neighbourhood in one direction, but blend smoothly into adjacent neighbourhoods in other directions. This poses some formidable methodological challenges, particularly when computing inference for the existence of a social frontier, an important goal if one is to distinguish true frontiers from random variation. We develop a new approach using Bayesian spatial statistical methods that permit asymmetries in spatial effects and allow for spatial autocorrelation in the data. We illustrate our method using data on Sheffield and find clear evidence of 'open' frontiers. Permutations tests and Poisson regressions with fixed effects reveal compelling evidence that social frontiers are associated with higher rates of crime.

**Keywords**: Segregation; social frontiers; neighbourhood boundaries; social cohesion; territoriality; crime; neighbourhood conflict

## Introduction

Fissures in social relations have been researched at length in the geography and sociology literatures. For example, there is a very large literature on explaining (Schelling 1971) and measuring (Massey & Denton 1988) segregation, with recent emphasis on capturing multi-scale aspects (Jones *et al.* 2015) and quantifying uncertainty (Lee *et al.* 2015; Kavanagh *et al.* 2016).

In the human geography literature, attention has been given to the geography of difference, the significance of boundaries in shaping power relations, and social exclusion of those who are defined to be the 'outsiders' in relation to the normatively dominant group (Sibley 1995). Social frontiers can be thought of as the spatial fault-lines of 'social tectonics' (Butler & Robson 2001) where different social groups may at times move past each other "like tectonic plates below the Earth's crust, with little contact" (Jackson & Butler 2015: 2350), but at other times become epicentres of pent-up social tensions and conflict. Given the size, scope and importance of these interrelated literatures, it is surprising to find an almost complete absence of robust quantitative research on 'social frontiers', both with respect to how social frontiers can be identified, and in terms of estimating their impacts on individuals and society.

We define social frontiers as boundaries between adjacent neighbourhoods where there are steep differences in the racial, ethnic, religious, cultural or social characteristics. These frontiers represent cliff edges in the complex landscape of segregation. Shaped by processes of homophily (McPherson *et al.* 2001) and market sorting (Kuminoff *et al.* 2013) they can embody "objectified forms of social differences manifested in unequal access to and unequal distribution of resources (material and nonmaterial) and social opportunities" (Lamont & Molnar 2002: 168).

The pioneering work of Logan *et al*. (2011) called for developing alternative approaches in identifying ethnic neighborhood boundaries. Since then, Spielman and Logan (2013), Kramer (2017) and Legewie and Schaeffer (2016) all remark on deficiencies in existing research on segregation which "does not address what happens at places where groups border" (Legewie & Schaeffer 2016: 131). Similarly, Kramer laments how "empirical research on neighborhood boundary making is practically non-existent" (Kramer 2017: 2). The lack of attention to boundaries is reflected in the way measures of segregation have been developed and deployed in the literature (Massey & Denton 1988) which tend to overlook the nature of transitions between neighbourhoods. For example, for a given value of the index of dissimilarity—perhaps the most widely used measure of segregation—there are many different possible spatial configurations of neighbourhoods. So two cities can have exactly the same level of segregation as defined by the index of dissimilarity but have very different degrees and frequency of social frontiers: it may be possible to either arrange the neighbourhoods in such a way that there are no sharp differences between any two contiguous neighbourhoods, or in such a way that there are many such differences.

There is therefore an imperative to address the dearth of research on neighbourhood transitions and their potentially important impacts. Legewie and Schaeffer (2016), for example, have argued that conflicts may arise along contested "fuzzy" borders as competing groups contend for territory. However, this is perhaps best thought of as part of the dynamic evolution of social frontiers, rather than an indication that such frontiers are desirable. Well-defined frontiers can become places of long-term conflict—consider the 'peace lines' of Belfast. So while conflict may be most intense during particular stages of frontier formation where neighbourhood borders are still being contested, they may nevertheless have long term deleterious consequences. Social frontiers may also emerge under relatively benign circumstances (Schelling 1971) and may lead to problems other than those arising from inter-group contact. Given the multiple phases and aspects of social frontiers, we seek to position the Legewie and Schaeffer proposition along with other hypotheses as part of larger schema comprised of multiple interconnected processes and phases of change. We hope this schema will not only highlight the importance of social frontiers and the need for further research in this area, but also provide a platform for a more capacious framework for understanding the causes and consequences of social frontiers, and stimulate new directions in this nascent field.

For this literature to develop, however, we first need a robust method for identifying frontiers. This is the main methodological and empirical contribution of the paper. A small number of recent studies have attempted to address this by developing empirical strategies that capture the nature of transition between neighbourhoods. Spielman and Logan (2013) use high-resolution population data to define neighbourhoods in selected 1880 US cities by looking at the distribution of groups defined by ethnicity and class. Kramer (2017) uses a combination of GIS and kernel density analysis to locate frontiers, while Legewie and Schaeffer (2016) deploy edge detection algorithms borrowed from the image processing literature. Legewie and Schaeffer (2016: 125), however, go a step further and consider the impact of social frontiers both theoretically and empirically, challenging "the 'aspatial' treatment of neighborhoods as isolated areas in research on ethnic diversity". Their work makes a valuable contribution, particularly since it is possibly the first serious attempt to quantify the impacts of neighbourhood boundaries. Additionally, they demonstrate that the relationship between neighbourhood conflict and boundary/edge density is not linear but reverse-u-shaped, with a higher number of tensions being recorded around 'fuzzy' boundaries.

We seek to address two methodological issues with respect to frontier detection. First, we tackle the issue of uncertainty – how do we gauge how reliable an estimated social frontier is? Ideally, we would like to quantify the uncertainty to help us distinguish true frontiers from those that are the product of random error. Application of statistical inference is made considerably more challenging, however, when spatial autocorrelation is present. This is likely to be the case in the kind of demographic data used in segregation research. Research in related areas has demonstrated how problematic ignoring spatial autocorrelation can be, leading to large inaccuracies in the measurement of uncertainty (Lee, Minton & Pryce 2014). This issue has yet to be addressed in the social frontier literature as far as we are aware. Second, there is a need to find a method for frontier detection that does not impose 'closed' boundaries. It is possible, for example, that a particular neighbourhood has a steep social frontier along one section of its boundary segment but gradually blends into neighbouring communities along other sections. Existing approaches to this tend to use spatial smoothing approaches that overlook the kind of asymmetries in neighbourhood transitions that we observe in real life.

As far as we are aware, no previous study has successfully addressed these issues in a unified way. The most relevant developments are those made in the areal Wombling literature where the aim is to identify areas or areal edges of abrupt changes in the distribution of a spatial outcome (Womble 1951). A comprehensive review on Wombling approaches or algorithms for point-referenced or image data is provided in Jacquez *et al.* (2000). Most relevant to our approach is the recent Bayesian areal Wombling literature where the issues of uncertainty and directionality can be better accommodated (Lu & Carlin 2005). At its heart, Bayesian hierarchical spatial models are employed. Based on the rich estimates on the fitted outcomes or residuals usually through the Markov chain Monte Carlo

(MCMC) approach, differences in geographically neighbouring areas and the associated empirical distributions can be obtained. These differences, also referred to as boundary likelihood values, are then compared to a threshold specified by a researcher to define boundaries (Lu & Carlin 2005). The uncertainty and spatial correlation in data are explicitly accounted for in this model-based boundary detection approach, though they are not intuitively modelled in the edge detection approach employed in Legewie and Schaeffer (2016). The key concern, however, is in relation to the spatial smoothing process induced by a global spatial model, which might mask local discontinuity when crossing areal borders, and subsequently affect the identification of boundaries. The same limitation is also applied to the approach in Legewie and Schaeffer (2016), which imposes a smoothing algorithm to the original data via spatial kernels. In addition, the two-stage boundary detection approach is isolated in the sense that the boundaries identified (i.e. local spatial discontinuity) are not utilized to update the local structure of spatial correlations and refine the estimates of model parameters. To address these issues, recent advances in Bayesian areal Wombling studies adopt a full Bayesian approach that treats the adjacency structure of areas (or elements of a spatial weights matrix) as random quantities to be updated along with other model parameters (Ma et al. 2010; Rushworth et al. 2017). However, the approach is computationally challenging and the practical applications to large data sets are highly restricted.

Our approach to social frontier detection is in line with the broad Bayesian areal Wombling literature. The uncertainty and global spatial smoothness of data is captured by a spatial autoregressive model while the boundaries are detected as locations where significant step changes in the data take place. More specifically, a locally adaptive spatial conditional autoregressive model is developed in which the estimation of the spatial model allows for the spatial weights matrix to be updated using information on identified boundaries. The estimation process is iterated until a termination criterion is met. To ease the computational burden, the spatial autoregressive model is estimated by using an approximate Bayesian inference approach—an integrated nested Laplace approximation (INLAs, Rue et al. 2009) instead of the MCMC simulations that are used in most Bayesian areal Wombling studies (e.g. Lu & Carlin 2005, Ma et al. 2010). Thus, the approach proposed in this study can cope with large data sets. To facilitate inter-study comparison, the method also needs to be replicable (as opposed to a bespoke case-study approach that relies heavily on researcher judgement). We illustrate our proposed method using data on Sheffield and demonstrate the potential usefulness of the derived boundaries by testing whether neighbourhoods joined by a social frontier tend to have higher rates of crime.

Note that, for clarity, we use the term ‘social frontier’ to distinguish steep social shifts in geographical space both from (1) arbitrary administrative or aerial unit borders, and from (2) ‘social boundaries’ which may be non-spatial. A social boundary is what Tilly defines as “any contiguous zone of contrasting density, rapid transition, or separation between internally connected clusters of population and/or activity” (2004: 214). ‘Social boundaries’ can be thought of as a broader concept which include social frontiers but also include non-spatial divisions such as those across different social and labour market spheres, or virtual daily interactions that cluster by particular attributes. We use the term ‘borders’ to refer to the perimeters of administrative areas such as local authority areas or statistical zones, defined by national or statistical authorities for administrative purposes.

The remainder of this paper is structured as follows. In section 2 we set out our theoretical reasons for the viewing of social frontiers as malignant features of the social landscape, representing an important and under-explored domain for future research. In section 3 we describe our method for empirical estimation of social frontiers using locally adaptive spatial conditional autoregressive Bayesian estimation. In section 4 we present the results of our frontier detection algorithm applied to lower super output areas (LSOAs) in Sheffield, England. In Section 5 we report the results of a permutation test and fixed effects Poisson model to estimate the impact of these social frontiers on local crime rates. Section 6 concludes with a brief summary of our findings and a discussion of their implications.

## The Malignancy of Social Frontiers

We begin this section with a summary of the overall structure of our argument framed as a series of summary statements, which we then relate to existing literatures. These statements describe the causes (S1 & S2), and possible consequences (S3-S5) of social frontiers.

*The Causes of Social Frontiers*

S1: Social frontiers arise as the consequence of strong aversion to living at the interface of communities in conflict; this aversion is the product of sustained antipathy between groups (Yinger 1976). The cliffs in the social landscape that emerge as a result, represent important indicators of entrenched social division.

S2: Social frontiers may alternatively be the product of benign forces, such as the unintended macro consequence of micro decisions (Schelling 1971).

*The Consequences of Social Frontiers*

S3: Social frontiers can represent places of settled difference, zones of temporary stability in the sorting process. They may denote lines of equilibrium where there are no net forces at work to compel further segregation. In contrast, places where frontiers are emerging or being contested, may be where most friction is generated, albeit fleetingly (Legewie & Schaeffer 2016). As such, it may be the process of frontier development, rather than their longstanding existence, that generates the most acute conflict.

S4: Nevertheless, by physically separating dissimilar populations, social frontiers may serve to entrench spatial impediments to intergroup contact and inter-group relations (Allport 1954). Households who live outside their own community or between diverse can have a bridge-building effect by linking diverse social networks and providing a buffer that cushions and alleviates inter-group tensions and misunderstandings. Social frontiers imply the absence or sparsity of such households, particularly at the transitions between neighbourhoods, reducing the potential for social networks to connect and understand each other. Combined with the defensive nature of territorial boundaries, social frontiers are therefore likely to imply fleeting or negative contact between communities. Social frontiers therefore have the capacity to impede social relations in the long run, heightening the sense of social division and territoriality; laying the grounds for future conflict.

S5: Residents in closest proximity to social frontiers are necessarily distant from the core of the communities of which they are a part. It is at the core where community norms and social hierarchies are most firmly established and help maintain order. Social frontiers may therefore represent zones where processes of social control are least potent, fostering deviant behaviour more generally, not just crimes of inter-group conflict.

These summary statements describe a set of complementary interactive processes that potentially reinforce or mitigate each other. Benign forces can strengthen malignant ones and vice versa, so that frontiers which began as uncontested boundaries become zones of conflict and threatened expansion. We now expand on how existing literatures relate to these highlighted processes, with a focus on the consequences of social frontiers which motivate the need for rigorous measurement.

### *Social frontiers as the Product of Antipathy (S1) and Benign Forces (S2)*

Yinger (1976: 370) presented theoretical models "of racial prejudice and household location" which "predict that the black area in a city will be of the shape that minimises the length of the black-white border". The minimum border length (MBL) hypothesis is the logical outcome of strong aversion to leaving near those from another ethnicity, religion or social group. I also offers a powerful mechanism for cliffs in the social landscape to emerge as a result of underlying deeply-felt antipathy.

In contrast to this hypothesis are Schelling's (1971) models of residential sorting, one of the most striking implications of which is the emergence of segregation as an unintended consequence. Even when all households "prefer a degree of integration" (1969: 489), if they have a relatively weak preference for homogeneity—such as not wanting to be in the minority in their immediate locality—individual decisions to move preferred neighbourhoods will result in a highly segregated society, even if no one wants that outcome. The implication is that residential mix is an unstable state—even modest levels of homophily can cause it to unravel.

Social frontiers may also arise as accidents of history and geography. Planning decisions to locate migrants in particular housing schemes, physical obstacles, infrastructure or environmental objects or features – such as railway tracks, highways, rivers, green spaces – that happen to divide communities and reduce interaction, can all provide unintended starting conditions for segregation (Hipp et al. 2014). Path dependencies emerge as community identity forms around these spatial features. Territorial borders between neighbourhoods or regions may be important for construction of meaningful difference on the other side of the border, especially if it overlaps with distinctive local, regional or ethnic identities (Lamont & Molnar 2002; Brantingham *et al.* 2012). Environmental barriers might also play a role in increasing dissimilarity among residents on both sides of the barrier (Noonan 2005; Van Gent *et al*. 2016). Hence barriers may, over time, become frontiers.

### *Social Frontiers as places of Settled Difference (S3), Barriers to Personal contact (S4) and Deviant Behaviour (S5)*

Whether social frontiers are the product of benign and/or malignant causes, the spatial division they create has the effect of reducing the opportunity for social interaction between groups *cet par*. This could have both good and bad consequences for social relations between communities. For example, following Allport's (1954) contact hypothesis, a physical separation of residents is likely to lead to greater misunderstanding, distrust and prejudice between groups, and, in turn, produce less often and more fleeting encounters between representatives of the separated groups. This is not, of course, unique to social frontiers – similar concerns have been raised about other forms of segregation (Uslaner 2010). However, frontiers may imply a particular spatial configuration of contrasting neighbourhoods that entails an absence or shortage of "bridge-builders" – those willing to live at the interface between communities. Bridge-builders create granular and more gradual transitions between two neighbourhoods which act as buffer zones and provide nodes of connection between isolated social networks. While dissimilarity indices measure how evenly a minority group is distributed, they overlook the spatial juxtaposition of contrasting neighbourhoods which is potentially more important. For a given level of unevenness, there are many possible levels of social frontier depending on how areas are arranged in space. How neighbourhoods are joined could be vital for inter-group relations. The fewer the number of 'bridge-builders', the steeper the frontier and vice versa. Opportunities for 'meaningful encounter' – contact which would allow questioning one's prejudice and have a transforming role for attitudes (Valentine 2008) – will be limited both by the spatial separation, and by the absence of individuals willing and able to connect the two communities. Being more exposed to the members of another community, but without actual contact could have negative consequences for inter-group relations (Stolle *et al.* 2008). Between-group violence might therefore cluster around territorial boundaries, especially if there are hostile relations between those groups resulting in competition over territory (Brantingham *et al.* 2012). Contact at the frontier in the context of between-group conflict could have harmful results, as it does not fulfil all requirements of the contact hypothesis: equal status, common goals, cooperation and an institutional support (Paolini *et al.* 2010; Pettigrew, & Tropp 2006).

On the other hand, the lack of contact might bring peaceful separation between dissimilar communities by helping to reduce opportunities for 'negative contact' between adversaries. Legewie and Schaeffer's (2016: 125) hypothesis is that "neighbourhood conflict is more likely to occur at fuzzy boundaries defined as interstitial or transitional areas sandwiched between two homogeneous communities". These areas are contested in the sense that "they threaten homogeneous community life and foster ambiguities about group rank. […] Well-defined boundaries, by contrast, are accepted divisions between one group's turf and another's and are thus less contested" (p. 126). Intergroup competition theories (Blumer 1958; Olzak 1992) suggest that "people feel threatened by the presence of out-group members because of real or perceived competition […] for scarce resources" (Legewie & Schaeffer 2016: 129). Ambiguous boundaries threaten this sense of cohesion and lead to defensive responses, while "sharp or well-defined boundaries do not threaten the integrity of neighborhood communities" (ibid: 132; see also Lim *et al.* 2007: 1543).

But tensions along contested boundaries are unlikely to leave the social landscape unchanged. Rather, they will catalyse further rounds of residential sorting and increased segregation. Location of perceived boundaries overlaps with patterns of daily activities and interactions (Van Gent *et al.* 2016). So the Legewie & Schaeffer hypothesis can be interpreted not as an end point, but as part of the sorting process that generates social frontiers. The greater the perception of threat from the presence of out-group members, the more rapidly social frontiers will emerge.

Another consequence (S5) of social frontiers is that residents living in their shadow are distant from the core of their respective community where social norms and hierarchies are most clearly established. According to 'environmental criminology', crime is related to social (dis)organisation of local community and is a function of neighbourhood dynamics (Bottoms & Wiles 2002). High level of crime indicates the lack of social cohesion between subpopulations in a community (Sampson *et al.* 1997). Crime happens because of a community's 'inability' to maintain social order and control, which is more unsuccessful in neighbourhoods with high social deprivation (e.g. Herbert 1977, Laurence 2015), but also in more social heterogeneity (Hirschfield & Bowers 1997). Crucially, residents living at the core of a community (rather than at the frontier between communities) may have a higher level of collective efficacy – residential ability to hold social control and related willingness to intervene in a neighbourhood – which is associated with a lower level of crime (Goudriaan *et al.* 2006; Sampson *et al.* 1997). Social isolation contributes to escalated violence within that peripheral zone, because it is located outside the reach of community institutions (Griffiths & Tita 2009) and thus social frontiers may be associated with deviant behaviour more generally, not just crime arising from inter-group conflict. Hirschfield *et al.* (2014), for example, finds that the level of burglaries is lower between similar areas than in areas that share borders with dissimilar neighbourhoods.

### *Summary of the main argument*

We have highlighted a number of key features of social frontiers that together scope out a broad framework for future research. These can be distilled into four interlocking processes: malignant separation, unintended segregation, temporary equilibrium, and widening division through lack of contact. While these processes may be well understood in isolation (at least in theory), little is known about how they combine to effect particular outcomes for social frontiers. For example, under what circumstances, stages of evolution, and combinations of interaction, do frontiers become places of conflict rather than settled difference? What are the implications for social frontiers of multiple groups and super-diversity? How stable are social frontiers and under what circumstances do they shift or dissolve? To what extent are neighbourhoods delineated by 'closed boundaries', encircled by social cliffs on every side? And to what extent do neighbourhoods have 'open boundaries' – their perimeters characterised by a combination of cliffs and slopes in social space? To what extent, and under what circumstances, do social frontiers exacerbate social tensions and disorder (S3-S5) and to what extent are they the product of them (S1-S2)? What are the impacts on the life course of living close to a social frontier, compared with living at the core of a neighbourhood (S5)?

It is beyond the scope of a single paper to rigorously evaluate these conjectures. Rather, they are presented as a way of delineating a new research agenda within which we can situate the current contribution. Against this backdrop, we argue, there is a clear imperative to find a robust way to identify social frontiers. In particular, we need a method for estimating social frontiers that both allows for 'open boundaries' (cluster analysis, for example, would delineate social boundaries without distinguishing between social cliffs and slopes in the way neighbourhoods intersect) and also tackles the issue of autocorrelation in spatial data—there is a need to gauge the level of uncertainty associated with the identification of a particular social frontier. It is to this challenge that we now turn with a view to providing the practical tools needed to invigorate research on social frontiers.

## The detection of boundaries in ethnicity segregation

### *Data*

We seek to illustrate the identification of social frontiers using LSOA[1] Census data (see Supplementary Material) on ethnicity and country of birth (CoB) for Sheffield Local Authority District. Figures 1 and 2 depict the spatial distributions of non-white population in

[1] LSOAs are lower layer super output areas used by UK Office for National Statistics, designed to cover minimum 1000 people or 400 households and maximum 3000 people or 1000 households. In our study area, there are 339 and 345 LSOAs in 2001 and 2011, respectively.

2001 and 2011 with the quintiles of each variable providing the cut points for the contour shading. At both census years, the non-white population are concentrated in the city centre area and the north-east of Sheffield. There is also increasing concentration of non-white population to the south-east and south-west of Sheffield. The geographical distribution of the non-UK born population is similar to that of the ethnic minorities (Supplementary Material, Figures SM1 and SM2).

### *Method for detecting social frontiers*

We propose a two-step approach for identifying frontiers in the distribution of non-white and non-UK born populations at each census year. The first step is to identify the step changes in the distribution of the non-white population (or non-UK born population). This entails finding geographic borders shared by two areas (LSOAs) that have statistically significant differences in the proportion of non-white population. These step changes are detected using a locally adaptive spatial conditional autoregressive model proposed by Lee and Mitchell (2013). Their approach was originally developed in the context of Poisson modelling of respiratory disease risk. We adapted the approach to a binomial probability distribution to model proportions. Intuitively, our approach establishes whether arbitrary aerial boundaries coincide with step changes across space in the social makeup of neighbourhoods. To do this, we fit a spatial model that initially assumes smoothness in the spatial distribution of ethnicity (or whatever variable is of interest) and then searches for abrupt changes in that surface by identifying sections along the LSOA borders where the smoothed surface poorly fits the data.

Suppose that the LSOA aerial units are indexed from $A_1$ to $A_n$ where $n$ is the total number of LSOAs in the study area. Within each LSOA we denote the count of people in the group of interest (e.g. non-UK born) as $Y_k$, where $k$ indexes the $k$-th LSOA. $N_k$ is the count of the total population in aerial unit $k$. This gives a binomial distribution of the number of non-UK born in area $k$ with proportion, $p_k$, of the total population in that area. We then take a logit transformation of this proportion, to give us, $\ln(p_k/1-p_k)$, which we set equal to a linear function of a spatial random effect for each of the aerial units, $u_k$:

$$\ln(p_k/1-p_k) = \beta_0 + u_k \tag{1}$$

The random effect is assumed to be spatially autocorrelated, which means that the variation in the proportion that are non-UK born in area $k$ is affected by the proportion that are non-UK born in the surrounding neighbourhoods that share a border with $k$. We start off by assuming that the proportion of non-UK born in each of the neighbourhoods contiguous with area $k$ has an equal effect on the size of the proportion of non-UK born in $k$. So we compute an average effect on each aerial unit $k$ of the contiguous neighbourhoods. This average effect becomes our "prior" in the Bayesian model. The strength of the spatial autocorrelation – how much the proportion of non-UK born in $k$ is affected by the surrounding neighbourhoods – is measured by parameter $\lambda$, which we need to estimate.

In the vast majority of spatial models used in social science, it is assumed that the spatial variation is smooth and symmetrical across contiguous aerial units. This is unrealistic because of the existence of step changes across space in many social variables, including ethnicity. Rather than assuming an average spatial effect across all contiguous neighbourhoods, ideally we want to compute the average only for the contiguous areas that are similar to $k$. Another way of saying this is that we need to allow the spatial weights matrix, W, to change as we identify these step changes. W is usually defined as a fixed matrix of 0s and 1s where a 1 indicates that a neighbourhood is contiguous with $k$ and therefore included in the average spatial effect estimate for $k$. In our method, we allow W to change as the model identifies contiguous areas that need to be ignored in the average spatial effect computation, because it is noticeably different to area $k$—i.e. there is a social frontier with $k$. This entails converting 1s to 0s in the W matrix when there is a statistically significant step change between contiguous areas. When this happens we know we have identified a potential social frontier.

Following Lee and Mitchell (2013), a Bayesian spatial conditional autoregressive model for a binomial response variable is specified as,

$$
\begin{gathered}
Y_k \sim Binomial\ (N_k, p_k);\ \ k = 1, \dots, n \qquad (2) \\
\ln(p_k/1 - p_k) = \beta_0 + u_k \\
u_k | \boldsymbol{u_{-k}}, W, \lambda, \tau^2 \sim N(\frac{\lambda \sum_{k \sim l} u_l}{1 - \lambda + \lambda w_{k+}}, \frac{1}{\tau^2(1 - \lambda + \lambda w_{k+})}) \\
\beta_0 \sim N(0, b);\ \tau^2 \sim \mathrm{gamma}(e', f');\ \mathrm{logit}(\lambda) \sim N(0,100).
\end{gathered}
$$

This more formal presentation of our modelling approach is described in more detail in the Supplementary Material.

However, finding discontinuities that are statistically significant does not necessarily imply they are substantive step changes between neighbourhoods. For example, a small difference (say of 0.01) in non-white proportions between two adjacent LSOAs could be statistically significant if the difference is measured with precision (i.e. estimated with low levels of uncertainty), but would not represent a substantial difference between the two areas. To account for this issue, a second hurdle defined in terms of substantive difference (the first one being the statistical significance) needs to be cleared for a border to be classified as a true social frontier.

### *Testing for an Association between Social Frontiers and Crime*

To allow for the possibility of different effects of social frontiers on different types of crime, we considered a number of crime categories including violence and sexual offences, burglary, shoplifting and vehicle crimes. To test the impact of social frontiers on crimes, we compare the crime rates among pairs of contiguous LSOAs joined by a social frontier against the crime rates of pairs of contiguous LSOAs not joined by a social frontier. The crime data (freely available at https://data.police.uk/) records the geo-coordinates of the streets where crimes happen. The crime data was further aggregated to LSOAs to calculate the LSOA-scale crime rates (see Supplementary Material for more information). In the following analyses, borders and frontiers are based on the geographical boundaries of LSOAs in 2011 as crime data are only available from December 2010. We obtained all the crime committed in Sheffield from December 2010 to December 2012, which gives more than 180,000 crime records, of which about 9,684 are violent crimes (violence and sexual offences), 14,142 burglary crimes, 11,942 vehicle crimes and 5,630 shoplifting crimes.

The test for comparing the difference in frontier-paired LSOAs and border-paired LSOAs is

$$\frac{C_F}{N_F * P_F} - \frac{C_B}{N_B * P_B}. \qquad (3)$$

In the equation, $C_F$ and $P_F$ represent the counts of crimes and the total population of paired LSOAs on the opposite sides of social frontiers identified above. $C_B$ and $P_B$ represent the counts of crimes and the total population of paired LSOAs on the opposite sides of borders. $N_F$ and $N_B$ represent the number of the frontiers and borders. Therefore, this test adjusts both the unequal amount of borders and frontiers (more borders than frontiers) and the population base. As an LSOA may have several bordering LSOAs, some of which are separated by frontiers while others are not, the above comparison will inevitably involve the focal LSOA multiple times. To address this issue, a further comparison is performed between two mutually exclusive sets of areas: frontier-paired LSOAs and LSOAs that are separated by borders and not included in the set of frontier-paired LSOAs.

As the counts of crimes for some LSOAs are repeatedly used in calculating Equation (3), it is difficult to test the statistical significance of the test statistic parametrically due to the dependency in the data. Instead, a permutation procedure is used to produce statistical inference on calculated differences in crime counts. In each permutation, the "borders" were randomly changed to "frontiers", and *vice versa*, and from this process the statistic in Equation (3) was calculated. The statistical significance of the actual statistic from data was derived from the distribution of the statistics calculated from permutated data. An R function to implement the permutation procedure is provided in the Supplementary Material.

For a more rigorous test of the social frontier effect on crime, we take advantage of the detailed locational information of the crime data. A spatial grid with a resolution of 100 meters by 100 metres was created for Sheffield, and was then overlaid with the point-referenced crime data to calculate the crime counts per grid. The spatial distribution of the total crime counts is provided in Figure SM3 in the Supplementary Material. Based on the fine-resolution gridded crime data, we estimate the effect of proximity to social frontiers on crime while controlling for the LSOA fixed effects. The proximity of grids to social frontiers is measured by whether a grid is located within 200 metres of a frontier. As unemployment has been found a significant predictor of crime, an inverse distance interpolation procedure was employed to interpolate the unemployment rates for each grid by using the output area scale unemployment rate data in 2011. We note that the inclusion of LSOA fixed effects is important as it will capture the sociodemographic, economic, and other unobservable contextual influences on crime and their potential correlations with the proximity variable. Poisson regression models specified in Equation (4) were used.

$$crime_{i,j}|\mu_{i,j} \sim \text{Poisson}(\mu_{i,j} = E_{i,j}R_{i,j}) \qquad (4)$$

$$\ln(\mu_{i,j}) = \beta x_{i,j} + \gamma z_{i,j} + \delta_j + \ln(E_{i,j})$$

In the equation, $crime_{i,j}$ is the crime count for grid $i$ in LSOA $j$ and follows a Poisson distribution with mean $\mu_{i,j}$. $E_{i,j}$ represents the expected crime count for a grid, which is calculated by using an internal standardization approach and is based on the population count of each grid, i.e. $E_{i,j} = \frac{\sum crime_{i,j}}{\sum pop_{i,j}} \times pop_{i,j}$ where $pop_{i,j}$ is the interpolated population count for the grid (the same procedure as in calculating the grid unemployment rates was used). $R_{i,j}$ is the risk of crime. $x_{i,j}$ is a binary proximity variable with a value of 1 representing grid $i$ within 200 metres of a frontier and a value of 0 otherwise. We also run a model with a distance threshold of 100 meters. $z_{i,j}$ measures the unemployment rate of grid $i$ in LSOA $j$. $\beta$ and $\gamma$ are estimated effects of proximity to social frontiers and unemployment on crime. The vector $\delta$ captures all the observed and unobservable LSOA effects, which is useful for the identification of the partial effect of proximity to social frontiers.

## Results of Boundary Detection

### *Frontiers of ethnicity and country of birth in Sheffield*

Several interesting features are revealed from Figure 3 which plots the estimated ethnic frontiers in the two census years. First, in 2011 the ethnicity frontiers tend to be linked with each other especially in the city centre area. This contrasts with the social frontiers estimated in 2001 which appear more discretely distributed. This seems to suggest that there was an increasing concentration of non-white population in the city centre area during the two periods (also shown in Figure 2). Secondly, enclosed frontiers are gradually emerging in 2011 in the north-west of the city centre where elevated concentration of non-white population was observed in areas enclosed by these frontiers, and was statistically significantly and substantially different from that of surrounding areas.

Figure 4 visualises CoB (country of birth) frontiers in Sheffield in 2001 and 2011 with proportions of non-UK born population superimposed. Comparing to the ethnicity frontiers, the overall spatial pattern of the CoB frontiers seems to be more stable during the two census years. There seems to be a significant and persistent northwest-southeast divide in the distribution of non-UK born population in Sheffield. In addition, the enclosed ethnicity frontiers are also observed in the CoB frontiers. As with the spatial pattern of ethnicity frontiers, there are both new CoB frontiers emerging and previous CoB frontiers diminishing due to population dynamics in Sheffield during last decade.

### *The Association between Social frontiers and Crime Rates in Sheffield*

After identifying social frontiers, we tested whether neighbourhoods joined by social frontiers tend to have higher crime rates. Table 1 summarises the results of the permutation test described above that seeks to establish whether there is *de facto* evidence for a relationship between social frontiers and crime. The second column of the table provides the differences in adjusted crime counts between ethnicity frontier-paired LSOAs and border-paired LSOAs (Equation (3)). The third column gives the statistical significance associated with each difference measure based on 1,000 permutations. The last two columns of the table show the differences in adjusted crime counts between CoB frontier-paired LSOAs and border-paired LSOAs and their statistical significances. The results suggest that there is an elevated risk of crimes in areas joined by social frontiers.

The positive differences in all crimes, burglary, violent, vehicle and shoplifting crimes indicate that the adjusted crime counts are larger than those for neighbourhoods joined by a social frontier (rather than an administrative border). These differences in crime counts, between neighbourhoods joined by social frontiers and those that are not, are all statistically significant at the 95% confidence level (Figure SM4 of the Supplementary Material). The comparisons between frontier-paired neighbourhoods and neighbourhoods that are separated by borders and not included in the set of frontier-paired neighbourhoods reveal the same results (Table 2).

The overall intensity of crimes, except shoplifting, is higher for the proximity to ethnicity frontiers than CoB frontiers. This result indicates that frontiers defined in relation to white ethnicity/race might have a stronger association with violent crime than frontiers defined in relation to being born in the UK. Burglary and vehicle crimes are more likely than violent crime and shoplifting to occur in neighbourhoods joined by social frontiers than those joined by administrative borders. These crimes do not involve personal interactions between an offender and a victim. Hence, there may be something about social (dis)organisation of community life at the peripheries which makes crimes against property more likely to occur.

Table 3 reports the results from estimating Equation (4) for all crimes. We do not estimate separate models for each crime category because of the issue of small sample size (for each specific crime category discussed above, there are not many grids with values). Proximity to social frontiers is statistically significantly associated with elevated crime rates at the 0.001 significance level. Grids within 200 metres of social frontiers are associated with an increased crime rates of about 6% (exp(0.056) - 1) than their counterparts after controlling for the LSOA fixed effects and the unemployment effects. The same finding holds in case that a 100 metres threshold distance is used. We also note unemployment is significantly associated with crime in both model specifications. With the inclusion of the LSOA fixed effect, our estimates on the effect of proximity to social frontiers are relatively credible.

## Discussion & Directions for Future Studies

*We construct borders, (…) to fortify our sense of who we are; and we cross them in search of who we might become*. (Stonor Saunders 2016)

We conceptualised social frontiers as spatial zones of contrasting density and separation between clusters of dissimilar populations (Tilly 2004). We aimed to stimulate research on this phenomenon by (1) establishing a more capacious conceptual framework for understanding social frontiers and (2) developing the practical tools needed to estimate social frontiers with statistical rigour. While social frontiers can emerge for any category of social difference, our focus has been on the consequences of ethnic segregation and neighbourhood ethnic diversity. Despite the enormous growth of segregation research, previous empirical studies in this field have tended to overlook the role of social frontiers (Kramer 2017; Legiewe & Schaeffer 2016) which are potentially important manifestations of the widely discussed processes of out-group threat, inter-group contact, or a lack of thereof, and neighbourhood disorganisation and anomie.

Whether social frontiers grow as a result of pre-existing group antipathies or as unintended outcomes of benign processes, they are believed to exert a powerful influence on social relations between divided communities and other processes taking place in the city. Living in proximity to social frontier entails exposure to concentrated clusters of the opposite community. This exposure accompanied by fleeting and superficial contact might contribute to the feeling of threat towards the other community members, especially in situations of 'scare resources' (Esses *et al.* 2001). Yet, if the exposure involves positive social contact, it might foster mutual respect and peaceful coexistence (Stolle *et al.* 2008). As such, there is ambiguity about the role of social frontiers which might play out differently in different communities and at different stages of their development. There is an imperative, therefore, for empirical researchers to establish detailed evidence on the nature and consequences of social frontiers, leading to a taxonomy of these effects and phases in different contexts.

While the existence of social frontiers might bring positive or negative outcomes for relations between communities, the adjacent areas might also become 'zones of no-one' –

distanced from core hubs of communities and under lesser social control. Hence, we argued that areas around social frontiers create favourable conditions for deviant behaviour to flourish. The existence of social tensions between two communities, however, might additionally increase social isolation of the area, which could 'attract' further crime. Social frontiers may also reflect an absence of 'bridge-builders' – those willing to live beyond the territory of their own group and who have the potential to connect isolated social networks and cushion inter-group tensions. The absence of 'bridge-builders' therefore makes neighbourhood relations more fragile with misunderstandings more likely to escalate.

Our conceptual framework is ripe for development into a more formal theoretical representation and raises a wide range of research questions that we hope will stimulate further research. The framework provides an imperative for developing a robust way of estimating social frontiers and quantifying the uncertainty embedded in social-spatial data. We proposed a Bayesian approach that not only takes into account spatial autocorrelation in the data but also permits asymmetries in spatial effects. Capturing this asymmetry is essential if we are to distinguish the cliffs from the slopes in spatial patterns of segregation. Neither do we want to impose 'closed boundaries' which force neighbourhoods to be encircled by social cliffs when in fact their borders may comprise a mixture of gradual slopes and precipitous edges.

To illustrate, we applied our approach to ethnicity and country of birth data on Sheffield. We found clear evidence of open boundaries. We tested whether the proximity to social frontiers is more strongly associated with crime intensity than proximity to administrative LSOAs borders. We found that all types of crime – violent crime, burglary, shoplifting and vehicle crimes – were reported at higher levels in neighbourhoods joined by social frontiers than neighbourhoods joined by area borders.

Our findings do not necessarily contradict the analysis conducted by Legewie and Schaeffer (2016). They argued that more within-community tensions would occur in areas adjacent to 'fuzzy' not 'clear-cut' borders, with neighbourhood tensions less likely to occur in areas around high-edge borders, but least likely to happen in areas without any borders. Note that some of our frontiers might fall within the fuzzy boundary threshold, as in our models the frontiers were defined as a step change difference of about 8% between two populations. Note also that Legewie and Schaeffer (2016) examined the impact on the incidence of relatively minor complaints and anti-social behaviours, whereas our interest was in the impact on more serious offences such as violent crime.

There are a number of important limitations to our methods and data. First, the crime data contains only crime which is reported, so many minor crimes are unlikely to be included in the data. Also, we cannot tell from the data which ethnic group committed a crime or against which ethnic group the crime was committed, so we cannot distinguish within and between group crimes from each other. Secondly, in the method looking at the link between social frontiers and crime intensity, we did not fully control for other differences between adjacent areas, such as social class differences, existence of environmental barriers, which might coexist with ethnic divisions. Future work will seek to develop multi-factorial approaches to estimating the relationship between crime and social frontiers. Thirdly, we do not look at the side on which the crime occurs, but at the overall intensity of crime in areas joined by a frontier. Hence, the next step in the analysis would be to explore cross-border asymmetries in crime intensity. Future research could build on the social frontier detection tools we developed to test this, as well as apply the method to detect other sorts of frontiers: socio-economic, demographic and religious, or intersectional; and whether some combinations of difference reinforce or mitigate crime outcomes. There is also the potential of applying our boundaries algorithms to new types of transactional and online data, such as social media or mobile phone data.

Further research is also needed on the evolution and dynamics of social frontiers. Social differences and their spatial configuration could have an important role in shaping power relations between groups, perpetuating unequal access to social, economic and political resources, with potentially deleterious long term effects on community cohesion. Yet, social frontiers are not fixed entities. They likely emerge and decline at different rates and in different ways in different social contexts. However, very little is understood about the spatial dynamics of social frontiers, partly because hitherto we have not had a reliable way of identifying them empirically. More research is also needed on the appropriate spatial scale for measuring social frontiers. We used LSOAs as our unit of analysis, but we do not know whether the nature and impact of estimated frontiers would change significantly if we had access to data at a finer spatial scale. When combined with the other research opportunities raised by the research agenda we set out in section 2, these questions highlights a very significant unexplored field of enquiry.

We hope our study not only signposts this uncharted territory but also provides a useful conceptual structure, complemented by a set of practical empirical tools to assist quantitative investigation. In time, we anticipate this emerging programme of research yielding important insights for social theory and policy.

## Funding

This work was funded by the Economic and Social Research Council (ESRC) through the Applied Quantitative Methods Network: Phase II, Grant Number ES/K006460/1.

## References

Allport, G. (1954), *The Nature of Prejudice*. Reading, MA: Addison-Wesley.

Blumer, H. (1958), Race prejudice as a sense of group position. *Pacific Sociological Review* 1, pp. 3-7.

Bottoms, A. E., & P. Wiles. (2002), *Environmental criminology* (pp. 620-656). The Oxford handbook of criminology.

Butler, T. & G. Robson. (2001), Social capital, gentrification and neighbourhood change in London: a comparison of three south London neighbourhoods. *Urban Studies* 38, pp. 2145-2162.

Brantingham, P. J., G. E. Tita., M. B. Short. & S. E. Reid, (2012), The ecology of gang territorial boundaries. *Criminology* 50, pp. 851-885.

Goudriaan, H., K. Wittebrood. & P. Nieuwbeerta, (2006), Neighbourhood characteristics and reporting crime. Effects of Social Cohesion, Confidence in Police Effectiveness and Socio-Economic Disadvantage. *British Journal of Criminology* 46, pp. 719-742.

Griffiths, E. & G. E. Tita. (2009), Homicide In and Around Public Housing: Is Public Housing a Hotbed, a Magnet, or a Generator of Violence for the Surrounding Community? *Social Problems* 56, 3, pp. 474-493.

Herbert, D. (1977), Crime, delinquency and the urban environment. *Progress in Human Geography* 1, pp. 208-239.

Hipp, J.R., J. Corcoran, R. Wickes. & T. Li. (2014), Examining the Social Porosity of Environmental Features on Neighborhood Sociability and Attachment. *PLOS ONE* 9, pp. 1-13.

Hirschfield, A. & K. J. Bowers. (1997), The Effect of Social Cohesion on Levels of Recorded Crime in Disadvantaged Areas. *Urban Studies* 34, pp. 1275- 1295.
Hirschfield, A., M. Birkin., C. Brunsdon., N. Malleson. & A. Newton. (2014), How Places Influence Crime: The Impact of Surrounding Areas on Neighbourhood Burglary Rates in a British City. *Urban Studies* 51, pp. 1057-1072.
Jackson, E., & T. Butler. (2015), Revisiting 'social tectonics': The middle classes and social mix in gentrifying neighbourhoods. *Urban Studies*, *52*(13), pp. 2349-2365.
Jacquez, G. M., S. Maruca. & M. J. Fortin. (2000), From fields to objectives: a review of geographic boundary analysis. *Journal of Geographical Systems*, 2, pp. 221-241.
Jones, K., R. Johnston., D. Manley., D. Owen. & C. Charlton. (2015), Ethnic residential segregation: A multilevel, multigroup, multiscale approach exemplified by London in 2011. *Demography* 52, pp. 1995–2019.
Kavanagh, L., D. Lee. & G. Pryce. (2016), Is Poverty Decentralizing? Quantifying Uncertainty in the Decentralization of Urban Poverty. *Annals of the American Association of Geographers* 106*,* pp. 1286-1298.
Kramer, R. (2017), Defensible Spaces in Philadelphia: Exploring Neighborhood Boundaries Through Spatial Analysis. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 3, pp.81-101.
Kuminoff, N., K. Smith. & C. Timmins. (2013), The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets. *Journal of Economic Literature* 51, pp. 1007-1062.
Logan, J. R., S, Spielman., H. Xu. & P. N. Klein. (2011), Identifying and Bounding Ethnic Neighborhoods. *Urban Geography* 32, pp. 334–359.
Lamont, M. & V. Molnar. (2002), The Study of Boundaries in the Social Sciences. *Annual Review of Sociology*, 28, pp. 167-195.
Laurence, J. (2015), Community Disadvantage and Race-Specific Rates of Violent Crime: An Investigation into the "Racial Invariance" Hypothesis in the United Kingdom. *Deviant Behavior* 36, pp. 974-995.
Lee, D. & R. Mitchell. (2013), Locally adaptive spatial smoothing using conditional auto-regressive models. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 62, pp. 593-608.
Legewie, J. & M. Schaeffer. (2016), Contested Boundaries: Explaining Where Ethnoracial Diversity Provokes Neighborhood Conflict. *American Journal of Sociology*, 122, pp. 125-161.
Leroux, B., X. Lei, & N. Breslow. (1999), Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical models in epidemiology, the environmental and clinical trials* (eds M. Halloran and D.Berry), pp. 135-178. New York: Springer.
Lim, M., R. Metzler. & Y. Bar-Yam. (2007), Global Pattern Formation and Ethnic/Cultural Violence. *Science* 317, pp. 1540-1544.
Lu, H. & B. P. Carlin. (2005), Bayesian areal Wombling for geographical boundary analysis. *Geographical Analysis*, 37, pp. 265-285.
Ma, H., B. P. Carlin. & S. Banerjee. (2010), Hierarchical and joint site-edge methods for medicare hospice service region boundary analysis. *Biometrics*, 66, pp. 355-364.
Massey, D. S. and N. A. Denton. (1988), The Dimensions of Residential Segregation, *Social Forces*, 67, pp. 281-315.
Noonan, D.S. (2005), Neighbours, Barriers and Urban Environments: Are Things 'Different on the Other Side of the Tracks'? *Urban Studies* 42, pp.1817-1835.

Olzak, S. (1992), *The Dynamics of Ethnic Competition and Conflict*. Stanford, Calif.: Stanford University Press.

Paolini, S., J. Harwood, & M. Rubin (2010), Negative intergroup contact makes group memberships salient: explaining why intergroup conflict endures. *Personality and Social Psychology Bulletin* 36, 12, pp. 1723-1738.

Pettigrew, T.F., & L.R. Tropp (2006). A Meta-Analytic Test of Intergroup Contact Theory. *Journal of Personality and Social Psychology* 90, 5, pp. 751-783.

Rue, H, S. Martino, F. Lindgren, D. Simpson, A. Riebler & E.T. Krainski (2014), INLA: functions which allow to perform a full Bayesian analysis of structured additive models using integrated nested Laplace approximations. URL: http: //www.r-inla.org/.

Rue, H., S. Martino, & N. Chopin (2009), Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 71, pp. 319-392.

Rushworth, A., D. Lee. & C. Sarran. (2017), An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 62, pp. 593-608.

Sibley, D. (1995). *A Geography of Exclusion*. London: Routledge.

Schelling, T. C. (1971). Dynamic models of segregation. Journal of mathematical sociology, 1(2), 143-186.

Spielman, S. E. & J. R. Logan. (2013), Using High-Resolution Population Data to Identify Neighborhoods and Establish Their Boundaries. *Annals of the Association of American Geographers*, 103, pp. 67–84.

Stonor, S. F. (2016), Where on Earth are you? *London Review of Books*, 38(5), pp. 7-12. Retrieved from http://www.lrb.co.uk/v38/n05/frances-stonorsaunders/where-on-earth-are-you.

Stolle, D., S. Soroka, & R. Johnston. (2008). When does diversity erode trust? Neighbourhood diversity, interpersonal trust and the mediating effect of social interactions. *Political Studies* 56, pp. 57-75.

Valentine, G. (2008), Living with difference: reflections on geographies of encounter. *Progress in Human Geography* 32, pp. 321-335.

Van Gent, W. P. C., W. R. Boterman. & M. W. van Grondelle. (2016), Surveying the Fault Lines in Social Tectonics; Neighbourhood Boundaries in a Socially-mixed Renewal Area. *Housing, Theory and Society*, 33, pp. 247-267.

Womble, W. (1951), Differential systematics. *Science* 114, pp. 315-322.

Yinger, J. (1976), Note on the Length of the Black-White Border. *Journal of Urban Economics* 3, pp. 370-382.

## Figures and Tables

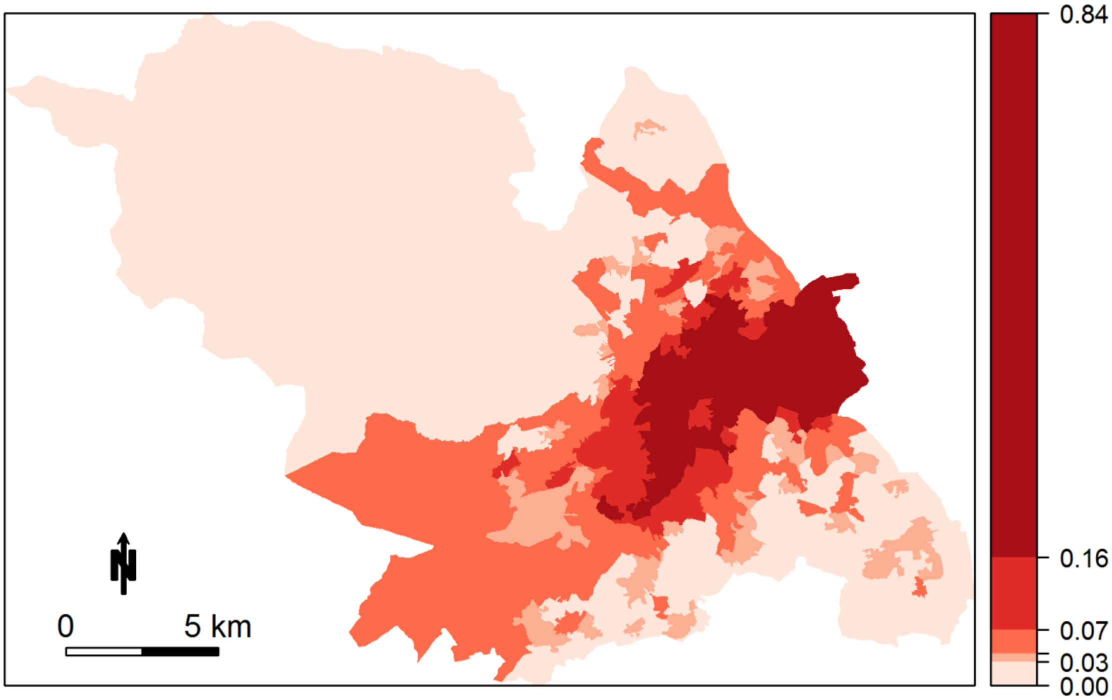



**Figure 1.** The distribution of the proportion of non-white population in Sheffield in 2001

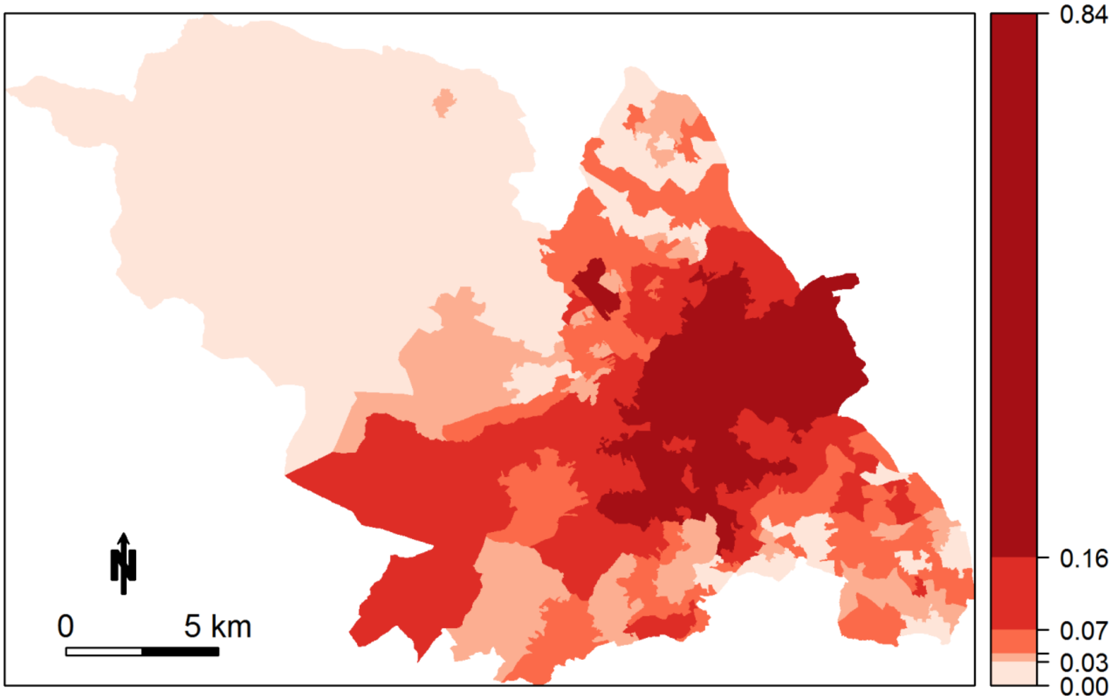
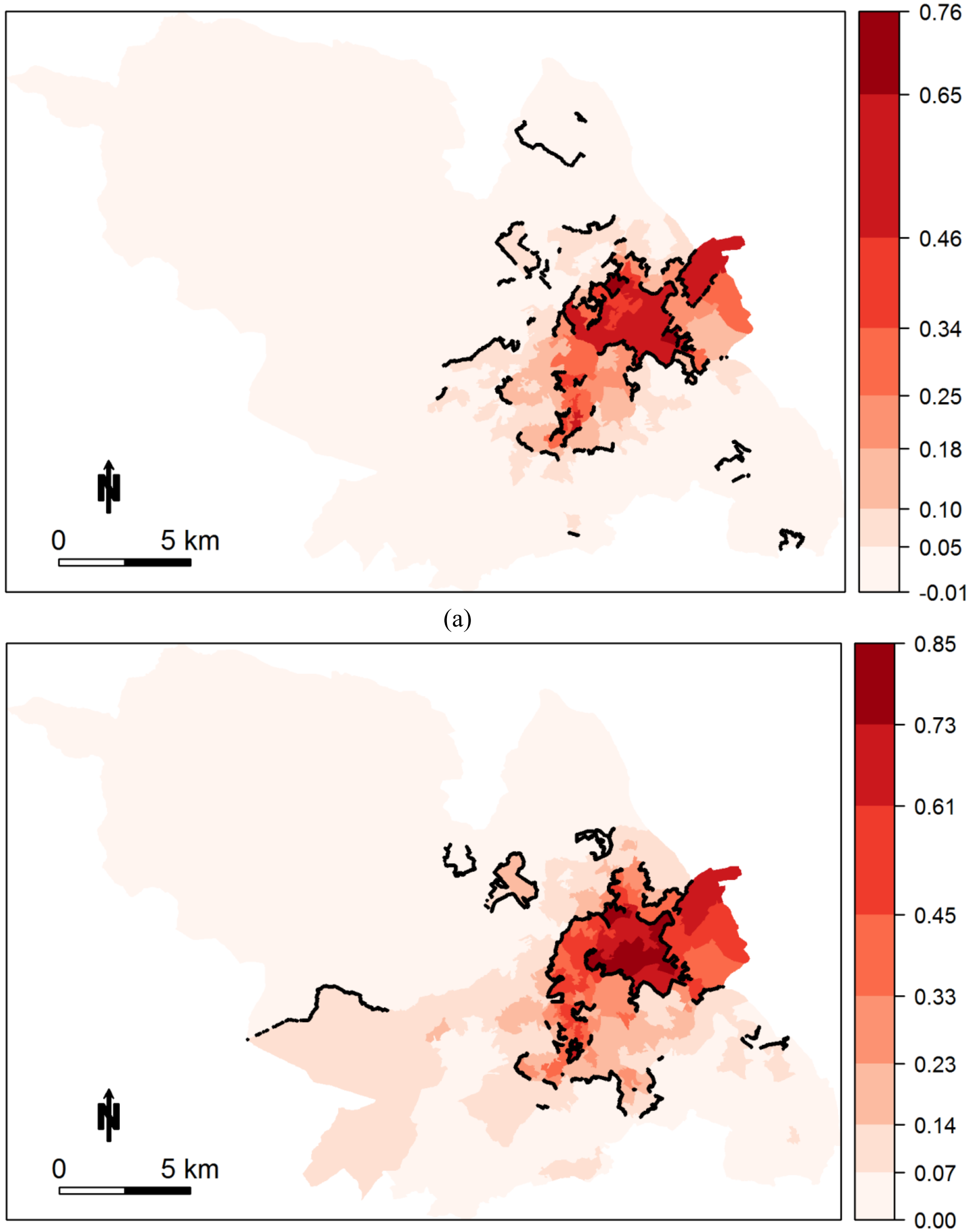



**Figure 2.** The distribution of the proportion of non-white population in Sheffield in 2011

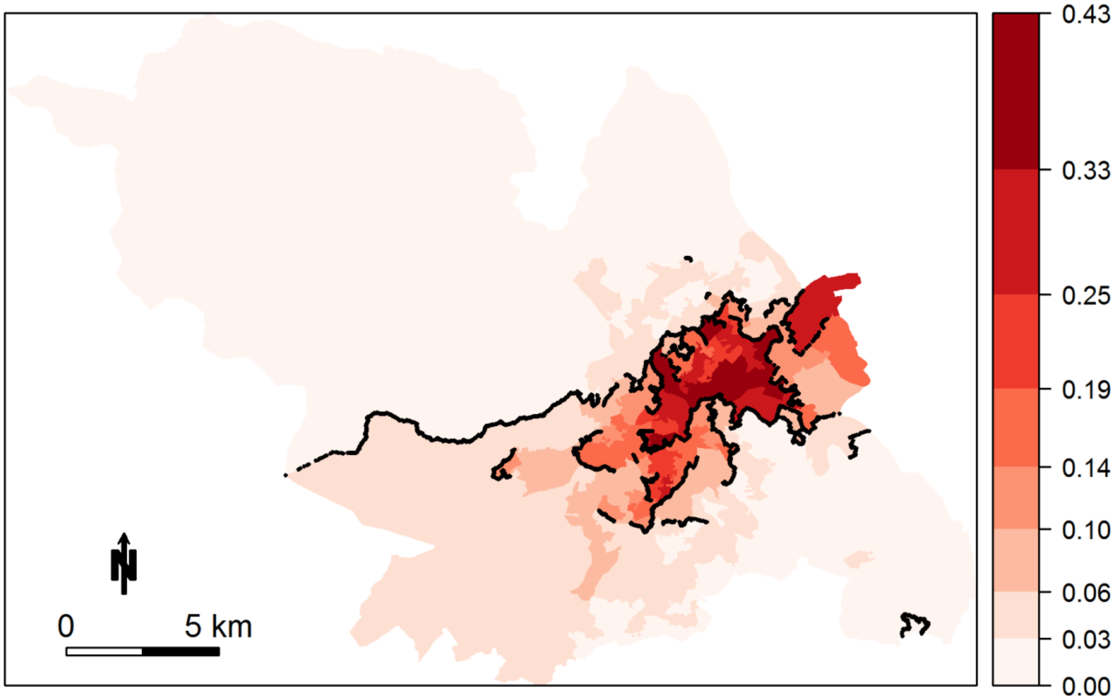



(a)

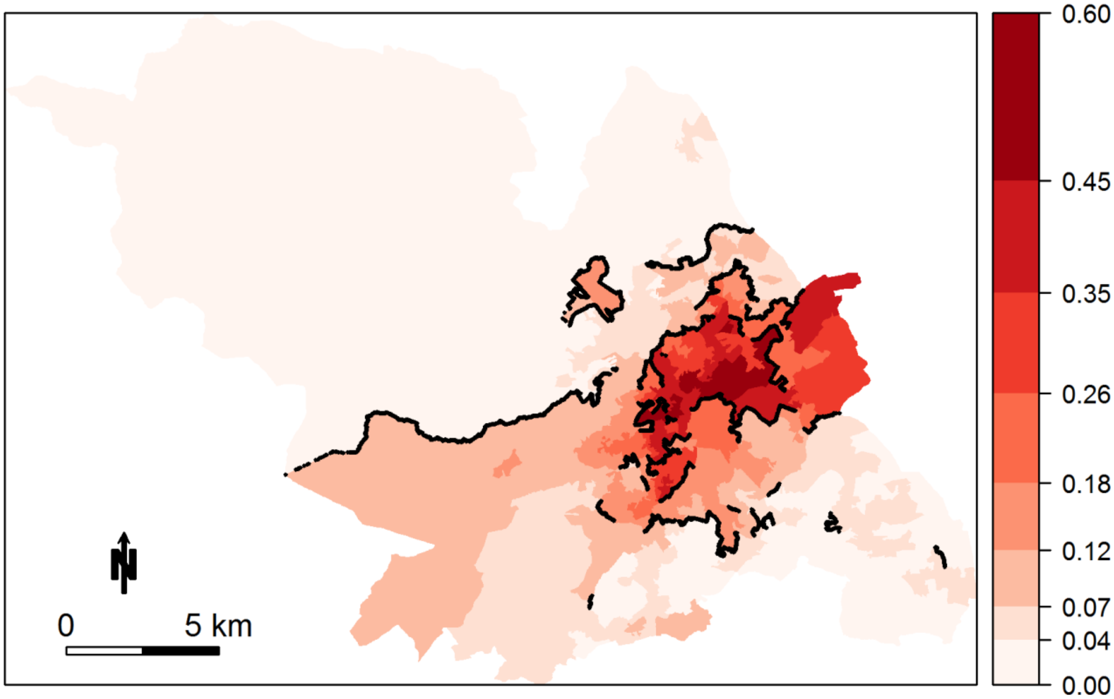


(b)

**Figure 3.** Model-identified ethnicity frontiers in Sheffield in 2001 (panel a) and 2011 (panel b)




(a)




(b)

**Figure 4.** Model-identified country of birth frontiers in Sheffield in 2001 (panel a) and 2011 (panel b)

**Table 1**. The differences in adjusted crime counts of between frontier-paired and border-paired LSOAs. The statistical significance tests are based on 1,000 permutations.

| | Ethnicity frontiers | | Country of birth frontiers | |
|---|---|---|---|---|
| Units: *Counts / per 1000 persons* | Differences as in Equation (3) | *p*-values | Differences as in Equation (3) | *p*-values |
| All crimes | 1.428 | 0.002 | 1.358 | 0.001 |
| Burglary crimes | 0.096 | 0.002 | 0.090 | 0.001 |
| Violent crimes | 0.083 | 0.011 | 0.084 | 0.001 |
| Vehicle crimes | 0.096 | 0.001 | 0.089 | 0.001 |
| Shoplifting crimes | 0.054 | 0.046 | 0.054 | 0.024 |

**Table 2**. The differences in adjusted crime counts of between frontier-paired and border-paired LSOAs. The statistical significance tests are based on 1,000 permutations. The border-paired LSOAs are the geographically bordering areas that are not included in the set of frontier-paired LSOAs.

| | Ethnicity frontiers | | Country of birth frontiers | |
|---|---|---|---|---|
| Units: *Counts / per 1000 persons* | Differences as in Equation (3) | *p*-values | Differences as in Equation (3) | *p*-values |
| All crimes | 1.337 | 0.001 | 1.207 | 0.001 |
| Burglary crimes | 0.085 | 0.001 | 0.074 | 0.001 |
| Violent crimes | 0.080 | 0.004 | 0.077 | 0.001 |
| Vehicle crimes | 0.088 | 0.001 | 0.080 | 0.001 |
| Shoplifting crimes | 0.053 | 0.044 | 0.052 | 0.013 |

**Table 3**. Estimation results on the effect of proximity to ethnic frontiers on crime.

| | *Threshold distance of 200 metres* | | *Threshold distance of 100 metres* | |
|---|---|---|---|---|
| | Estimate | Std.Error | Estimate | Std.Error |
| Intercept | 1.216* | 0.053 | 1.216* | 0.053 |
| Proximity dummy variables | 0.056* | 0.007 | 0.034* | 0.008 |
| Unemployed | 6.487* | 0.113 | 6.474* | 0.113 |
| LSOA fixed effects | YES | | YES | |
| AIC | 150813 | | 150856 | |
| Deviance | 125771 | | 125814 | |
| Sample size | 5145 | | 5145 | |

Note: The "*" symbol represents a significance level smaller than 0.001.

# Frontiers in residential segregation:
# Understanding neighbourhood boundaries and their impact
# Supplementary Material

*Further Information on the Data and Aerial Units*

The basic spatial unit analysed is the Lower Layer Super Output area (LSOA), which was created by grouping a few of Output Areas (typically four to six) based on the 2001 census data in England and Wales, and is the main geography through which the Office for National Statistics releases neighbourhood statistics (www.neighbourhood.statistics.gov.uk/). LSOAs have an average population of about 1,500 with lower and upper population thresholds of 1,000 and 3,000 and are relatively homogenous in terms of housing characteristics such as types of dwellings and housing tenure. Few LSOAs' geographic boundaries were updated based on the 2011 census data (about 1% of LSOAs in 2001), taking account of significant changes in population during the period. In our study area, there are 339 and 345 LSOAs in Sheffield in 2001 and 2011, respectively.

Our analyses are based on the ethnicity and country of birth (CoB) characteristics of each LOSA in Sheffield. The census data provide detailed information on categories of ethnicity and country of birth for each LSOA. For the purpose of an easy exposition of our methodology, we dichotomise ethnicity into two categories: white population and non-white population. Therefore, we have counts of the two ethnicity groups at each LSOA. Similarly, we categorise CoB into UK born population and non-UK born population. The aggregation of subcategories of ethnicity and CoB also alleviates the issue of excessive zero counts in the two variables in model estimation. The spatial distribution of non-UK born population in 2001 and 2011 is presented in Figures SM1 and SM2 below.

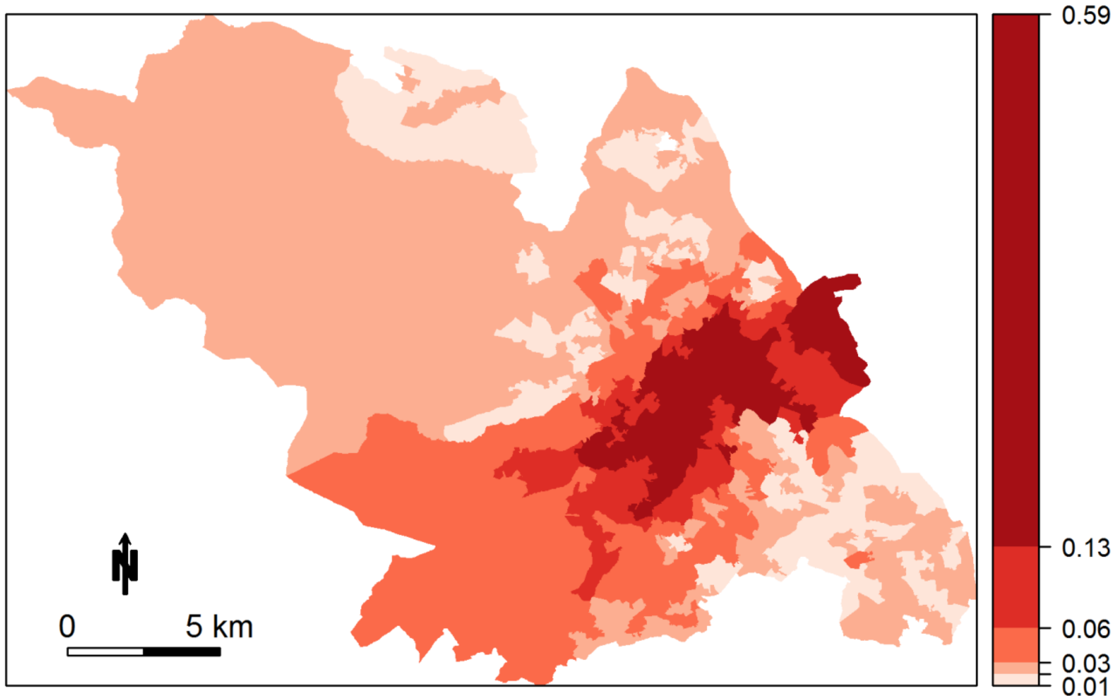


**Figure SM1.** The spatial distribution of non-UK born population in Sheffield in 2001

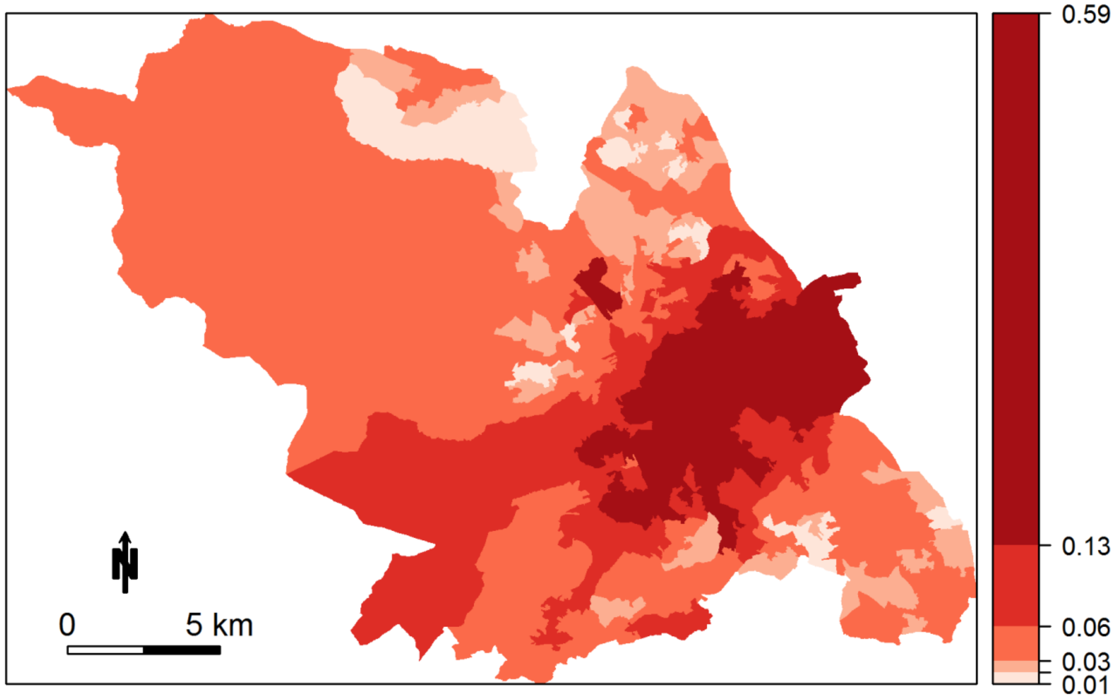



**Figure SM2.** The spatial distribution of non-UK born population in Sheffield in 2011

*Crime data*

The street-level crime data is publicly and freely accessible from https://data.police.uk/. It provides data on crime types such as violent, burglary and vehicle crimes for each police force area in England, Wales and Northern Ireland. The temporal coverage of the data is from December of 2010 to present. In the present study, crime data from December of 2010 to December 2012 in the South Yorkshire Police force area (where Sheffield city is located) are used. Although the geographical coordinates (longitude and latitude) of crime records are provided, they are not the exact locations where crimes happened but the locations of nearest centre points of roads or other locally relevant featured points to the actual crime sites. For this reason, it is not recommended to analyse the point-referenced crime data directly.

The study employs two ways to explore the crime data and estimate the social frontier effect on crime. First, we aggregated the point-referenced crime data to the LSOA scale and counted the number of crimes for each LSOA in Sheffield. We then compared the differences in crime rates between frontier-paired LSOAs and border-paired LSOAs (detailed in the main text).

Second, we created a spatial grid with a fine spatial resolution of 100 metres by 100 metres and counted the number of crimes for each grid cell. The spatial distribution of crime counts (all crimes) is presented in Figure SM3. By exploiting the geographical scales of the crime data, we are able to build statistical models to rigorously test the effect of proximity to social frontiers on crime while controlling for the sociodemographic, economic and other unobservable contextual differences of grids by incorporating the LSOA fixed effect.

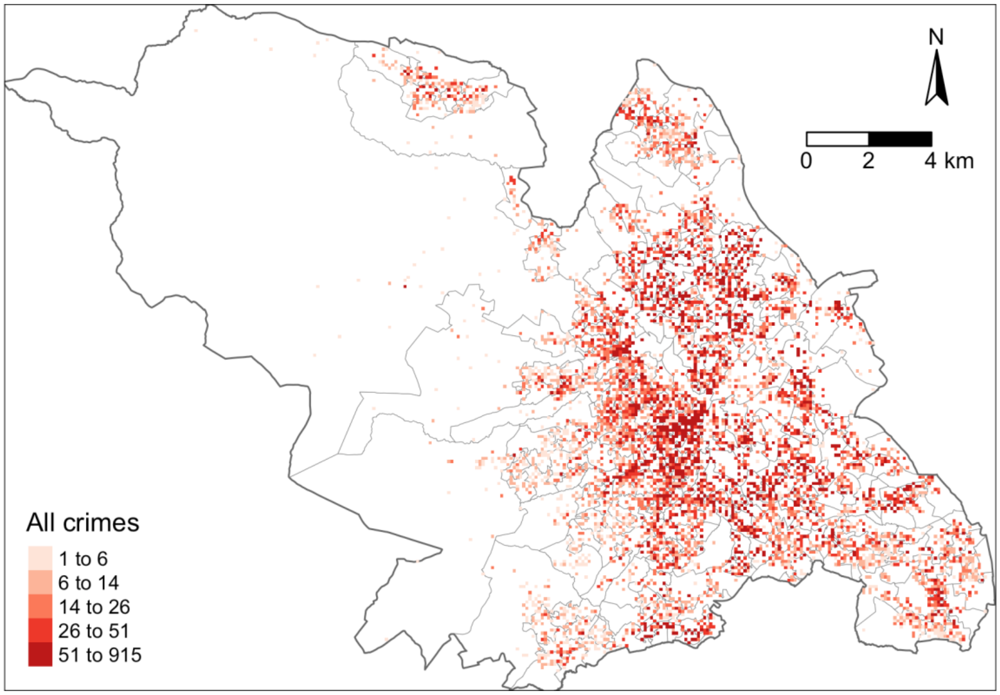



**Figure SM3**. The spatial distribution of crime at the 100 metres by 100 metres grid scale in Sheffield.

*Identifying Social Frontiers: Formal Description of Modelling Strategy*

Our modelling strategy entails estimating two sets of parameters: (1) the random effects $u_k$ which capture the change across space in proportion non-UK born, along with the associated parameters including $\lambda$ needed to estimate this spatial variation, and (2) the spatial weights matrix itself. W therefore becomes part of the estimation, rather than something we assume from the outset and keep fixed. Estimation of this flexible model requires an iterative procedure, one that starts with a standard spatial Conditional Autoregressive (CAR) model – i.e. one that assumes W is fixed, and then treats W as something that needs to be estimated.

Consider a study area (e.g. Sheffield, England) partitioned into $n$ areal units (LSOAs), $\boldsymbol{A} = \{A_1,\ldots, A_n\}$, on which the proportions of non-white population are quantified by $(Y_k, N_k)$. Following Lee and Mitchell (2013), a Bayesian spatial conditional autoregressive model for a binomial response variable is specified as,

$$
\begin{gathered}
Y_k \sim Binomial\ (N_k, p_k);\ \ k = 1, \ldots, n \\
\ln(p_k/1 - p_k) = \beta_0 + u_k \\
u_k | \boldsymbol{u_{-k}}, W, \lambda, \tau^2 \sim N(\frac{\lambda \sum_{k\sim l} u_l}{1 - \lambda + \lambda w_{k+}}, \frac{1}{\tau^2(1 - \lambda + \lambda w_{k+})}) \\
\beta_0 \sim N(0, b);\ \tau^2 \sim \text{gamma}(e', f');\ \text{logit}(\lambda) \sim N(0,100).
\end{gathered}
\tag{2}
$$

where the estimated proportion of non-white population in $A_k$, $p_k$, is modelled on the logit scale by a global constant $\beta_0$ and a separate random effect $u_k$. The whole set of random effects for all areal units is denoted by $\boldsymbol{u} = (u_1, \ldots, u_n)$. As is clearly shown in the Figures 1 and 2, there is a spatial clustering pattern (spatial autocorrelation) in the distribution of non-white and non-UK born population. Therefore, the possible spatial correlation effect is incorporated into $\boldsymbol{u}$ via a conditional autoregressive model developed by Leroux et al. (1999), the same approach used in Lee and Mitchell (2013). The spatial weights matrix W was a binary $n$ by $n$ matrix with $w_{jk} = 1$ if units $A_j$ and $A_k$ share a geographical border and $w_{jk} = 0$ otherwise. $\tau^2$ is a precision parameter measuring the conditional uncertainty of $u_k$ given the random effects from surrounding areas of $A_k$. The conditional autoregressive model implies that the conditional expectation (mean) of $u_k$ at area $A_k$ is a weighted average of random effects of areas surrounding $A_k$. $\lambda$ is a spatial correlation parameter measuring how strongly random effects $\boldsymbol{u}$ are correlated with each other. The model is completed by the specification of prior distributions for $\beta_0$, $\tau^2$ and $\lambda$. We use the same prior distributions for these parameters as in Lee and Mitchell (2013).

The joint posterior distribution that needs to be estimated is $f(\boldsymbol{\theta}|Y,\mathrm{W})$, where $\boldsymbol{\theta}=[\beta_0,\tau^2, \lambda, \boldsymbol{u}]$. The specification of areal random effects described above is overly restrictive as it assumes a global spatial correlation in the outcome variable across the whole study area. In other words, correlations (or similarities) in outcomes would be imposed as long as two areas share a geographic border. However, it is quite common that spatial data could contain both abrupt and gradual changes between bordering areas.

To tackle both the smoothness and step changes in spatial outcomes, Lee and Mitchell (2013) developed a locally adaptive spatial autoregressive modelling approach, in which elements of the spatial weights matrix W are also estimated along with other model parameters $\boldsymbol{\theta}$. The basic idea is that a strong global spatial smoothing across the study area was first enforced (by setting $\lambda$ close to 1), which could then be adapted locally by estimating the elements of W. For example, by setting $w_{jk}$ equal to zero during the model calibration process where areas $A_j$ and $A_k$ share a border ($w_{jk}$ was equal to 1 originally), the random effects of the two areas $u_j$ and $u_k$ would be conditionally independent even if there was a global spatial correlation or smoothing trend. In short, model parameters ($\boldsymbol{\theta}$, **W**) in the locally adaptive spatial autoregressive model were estimated using an iterative algorithm, which cycles between updating $\boldsymbol{\theta}$ and W conditioning on the other until certain termination criteria are met. Detailed information on the iterative algorithm and the R function used to implement the model are provided below.

*Social Frontier Estimation*

Their proposed estimation algorithm is illustrated below to facilitate the understanding of the model estimation process. In short, there are three steps for implementing the proposed locally adaptive spatial conditional autoregressive binomial model.

***Step 1***: estimate starting values of $\boldsymbol{\theta}$ by assuming the random effects $\boldsymbol{u}$ to be independent (fixing $\lambda$ to zero), denoted by $f(\boldsymbol{\theta}^{(0)}|Y,\mathrm{W}^{(0)})$ where $\boldsymbol{\theta}^{(0)} = [\beta_0^{(0)}, \tau^{2(0)}, \boldsymbol{u}^{(0)}]$

***Step 2***: iterate the update of $f(\boldsymbol{\theta}|Y,\mathrm{W})$ and $f(\mathrm{W}|Y,\boldsymbol{\theta})$.

(1) Estimate $\mathrm{W}^{(t+1)}$ deterministically based on the current estimation of random effects $\boldsymbol{u}^{(t)}$. Set $w_{jk}^{(t+1)} = 1$ if the marginal 95% posterior credible intervals for $u_j^{(t)}$ and $u_k^{(t)}$ overlap and areas $j$ and $k$ share a geographic border. Otherwise, $w_{jk}^{(t+1)}$ is set to zero. The rationale is that if the random effects of two geographically adjacent areas are substantially

different (non-overlapping 95% credible intervals of random effects), it gives clear evidence that these two area effects should be considered as conditionally independent.

***(2)*** Estimate the posterior distribution $f(\boldsymbol{\theta}^{(t+1)}|Y,\mathrm{W}^{(t+1)})$ based on Equation (1) and the updated spatial weights matrix $\mathrm{W}^{(t+1)}$.

***Step 3***: the iteration stops if either of these two termination conditions is met. The first case is that the sequence of estimated W is such that $\mathrm{W}^{(t^*+1)} = \mathrm{W}^{(t^*)}$. The second case is more complicated that the sequence of estimated W cycles between $k$ different states ($\mathrm{W}^{(t^*)}$, $\mathrm{W}^{(t^*+1)}$,…, $\mathrm{W}^{(t^*-1+k)}$, $\mathrm{W}^{(t^*)}$). In this case the estimated W giving the smallest level of spatial autocorrelation in model residuals from Equation (1) would be selected as the final spatial weights matrix.

Based on the final spatial weights matrix W*, the final estimation of model parameters $\boldsymbol{\theta}[\beta_0,\tau^2,\boldsymbol{u}]$ was obtained. Denote $d\_u_{k\sim l} = |u_k - u_l|$ as the absolute difference of the model fitted random effects $\boldsymbol{u}$ at areas $A_k$ and $A_l$ that share a common geographic border. We further define a threshold value of $mean\,(d\_u_{k\sim l}) + 1 * sd(d\_u_{k\sim l})$ to define social frontiers (ethnicity and COB) as it represents a level of difference considered important. Therefore, a geographic border shared by two areas $A_k$ and $A_l$ becomes a social frontier *if and only if* $\mathrm{w}^*_{jk} = 0$ and $d\_u_{k\sim l}$ is larger than the threshold value.

Based on the converged spatial weights matrix W*, the final estimation of model parameters $\boldsymbol{\theta}[\beta_0,\tau^2,\boldsymbol{u}]$ was obtained. Denote the absolute difference of the model fitted random effects $\boldsymbol{u}$ at areas $A_k$ and $A_l$ that share a common geographic border. We further define a threshold value to define social frontiers (ethnicity and CoB) as it represents a level of difference considered important. Therefore, a geographic border shared by two areas $A_k$ and $A_l$ becomes a social frontier *if and only if* $\mathrm{w}^*_{jk} = 0$ and is larger than the threshold value. In the whole model calibration process, Bayesian models with W-matrices were estimated using an integrated nested Laplace approximations approach (INLAs, Rue *et al.* 2009) through the R package INLA (Rue *et al.* 2014). It is useful to note that one of the two termination rules is guaranteed to be met after a large number of iterations as the sample space for W (or the possible different forms of W) is finite (Lee & Mitchell 2013). In the present study, only a small number of iterations (all less than five) are needed to reach the first termination condition.

### *Results of the Social Frontier Model*

The estimated ethnic frontiers in the two census years are shown in Figure 3 of the main text, which plots the results for Sheffield with proportions of non-white population superimposed. The threshold values we imposed to identify frontiers from step changes in the random effect surface give differences of 7.6% and 8.1% in the proportion of non-white population on the opposite sides of a frontier in 2001 and 2011, respectively.[2] There are 1,932 pairs of LSOAs sharing borders in 2001, about 41.8% and 13.1 % of which are identified as step changes and frontiers in the distribution of non-white population,

[2] The differences of 7.6% and 8.1% were evaluated when the random effect equals to zero, roughly the mean of the random effects in both years.

respectively. In 2011 there are 1,970 borders shared by LSOAs and 53.6% and 13.4% of them are identified as step changes and frontiers in the non-white population distribution.

*Association between Social frontiers and Crime Rates: Permutations Test*

The following figure (Figure SM4) shows how differences in crime counts, between neighbourhoods joined by social frontiers and those that are not, are all statistically significant at the 95% confidence level.

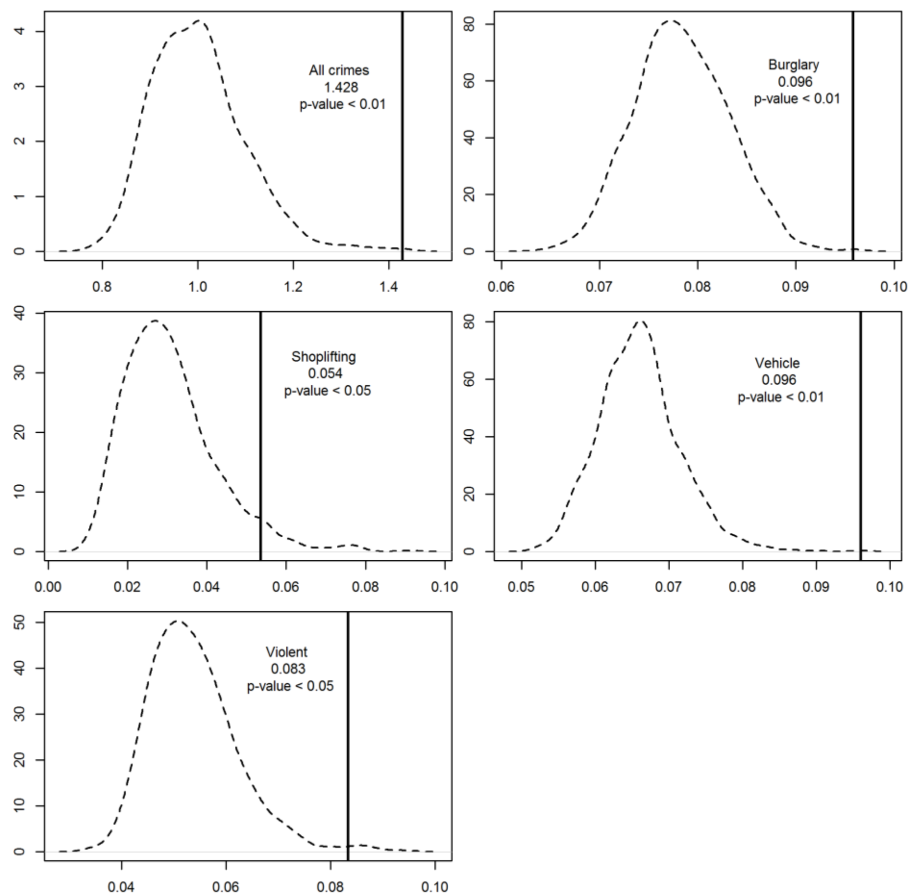



**Figure SM4**. The permutation test of the ethnicity frontiers impacts on different types of crimes. The solid lines in the figure present the observed differences in adjusted crime counts between frontier-paired LSOAs and border-paired LSOAs whereas the density distribution of the same statistic using 1,000 permuted data is represented by a dashed line.

*Association between Social frontiers and Crime Rates: Fixed Effect Poisson Estimation*

As discussed in the main text, a more robust and model-based approach was employed to test the potential frontier effect on crime by estimating a Poisson model with LSOA fixed effects. The dependent variable is the count of crimes for each grid cell with a resolution of 100 metres by 100 metres covering the study area, as illustrated in Figure SM3. The independent variable of particular interest is the proximity of each grid cell to the nearest frontier, which is measured by whether a cell is located within the 200-metre buffer of an identified frontier (the variable $x$ in Equation (4) of the main text). The coefficient of $x$, $\beta$, gives the frontier effect on crime, with a positive (or negative) coefficient sign indicating elevated (or declined) crime counts or density near frontiers. The inclusion of LSOA fixed effects is expected to control for possible impacts on crime counts of LSOA-scale sociodemographic, economic and other unobservable contextual differences. In terms of model estimation, the LSOA fixed effects are accounted for by adding LSOA dummy variables into Equation (4). The Poisson model was estimated by using the *glm* functionality in the core *stats* package of R (Venables and Ripley, 2002).

*R functions*

*binomial.localisedINLA.R*: The file implements a binomial locally adaptive spatial autoregressive model discussed in the methodology section of the paper.

*Permutation_test.R:* The file implements a permutation procedure discussed in the paper.

*References for Supplementary Material*


Lee, D. & R. Mitchell. (2013), Locally adaptive spatial smoothing using conditional auto-regressive models. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 62, pp. 593-608.

Rue, H, S. Martino, F. Lindgren, D. Simpson, A. Riebler & E.T. Krainski (2014), INLA: functions which allow to perform a full Bayesian analysis of structured additive models using integrated nested Laplace approximations. URL: http: //www.r-inla.org/.

Rue, H., S. Martino, & N. Chopin (2009), Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 71, pp. 319-392.

Venables, W. N. and B. D. Ripley (2002), Modern Applied Statistics with S. New York: Springer.