Journal of Safety Science and Resilience 3 (2022) 115–127

Contents lists available at ScienceDirect

Journal of Safety Science and Resilience

journal homepage: http://www.keaipublishing.com/en/journals/journal-of-safety-science-and-resilience/

Could social medias reflect acquisitive crime patterns in London?

Zenghui Wang [a], Yijing Li [b,*]

[a] CIM, Warwick University, Coventry, UK
[b] CUSP London, King's College London, London, UK

ARTICLE INFO

Keywords:
Crime opportunity
Acquisitive crime
Social disorganization
Broken windows
Twitter
Foursquare
Big data

ABSTRACT

Embraced within the framework of crime opportunities integrated with Social Disorganization theory and Broken Windows theory, this paper intends to explore the patterns of four types of acquisitive crimes, using social media data i.e. Twitter, Foursquare and cross-sectional data acquired through text analysis technique. With Greater London as the study area, models like negative binominal regression (NBR) and geographically weighted regression (GWR) are performed to illustrate the aggregated relationships between acquisitive crimes and crime opportunities at London-wide and sub-regional MSOAs levels respectively. The results work towards to hypotheses that: the tweets sentiment could reflect property-related crime rates positively in light of Broken Windows Theory; more tweets with negative sentiment may incur increases of acquisitive crimes. It contributed to existing studies in (1) providing empirical evidence for integrating these three theories; (2) complementing current research on local discrepancies of acquisitive crimes by utilising both GWR and NBR models; (3) challenging the traditional stereotypes about racial disparities with the finding that ethnic heterogeneity and instrumental crimes have counterintuitive association, especially taking education factor into consideration; (4) implicating some localised acquisitive crime prevention strategies to policy makers in light of the reality that the relationship between local variations and different crime types may vary by place.

1. Introduction

Acquisitive crimes have long been focused among crime prevention research, in light of their premeditated, purposeful and intentional nature, which is contrasting with expressive crimes' impulsive, unthinking and irrational characteristics [1]. To better predict the patterns of acquisitive crimes, theories relating to crime opportunities, e.g. Rational Choice [2], Routine Activity [3] and Crime Pattern [4], were proposed to indicate associations using a variety of data from different countries and regions at multiple scales [5,6,7]; whilst interact with each other to inspire explorations on crime opportunity theories, i.e. to examine the patterns of acquisitive crimes and decaying social structures (e.g. residential instability) in neighborhoods referring to social disorganisation theory; jointly considering broken windows theory which concentrates on visible signs of incivilities (e.g. public drinking, trash, insulting graffiti) [8,9], Sampson and Raudenbush [10] confirmed that higher crime rates are originated directly from the reduced informal social control rather than disorders; similarly Miethe and Meier [11] proposed that informal social control[1] should be treated as essential factors in crime occurrences, which was further combined with broken window theory by Weisburd et al. [12].

In digital city era, social medias, like Twitter, Facebook and Foursquare, are becoming one of the most important data sources along with the always evolving measures. Data from social medias has valuable insights into human behavior patterns, includes not only individual daily activities, but also people's emotional changes and attitudes towards events. Such massive open-sourced data have been applied into crime literature in assistance with techniques like text analysis, sentiment analysis and machine learning classifiers [13,14,15], which turned out to be very time-saving and reflective; however, as an emerging research field, majority of the established literature is short of consideration of changes derived from the utilization of massive social media data [16,17], but rather more focused on predictive models building-up. Hence, this paper attempts to better test the spatial relationships of social media data with acquisitive crimes taking Greater London as the case study area; in light of relevant criminologial theories [10], indictors matrix was extracted from social media and cross-sectional surveys data as independent variables (Table 1), and local acquisitive crime rates as dependent variable were compiled using text classification and sentiment analysis.

2. Background and literature

Within the framework of crime opportunities, routine activity theory (RAT hereafter) had been initially applied in criminology [2] under the assumption that potential offenders have strict and monotonic prefer-

* Corresponding author.
E-mail address: yijing.li@kcl.ac.uk (Y. Li).

https://doi.org/10.1016/j.jnlssr.2021.08.007
Received 25 January 2021; Received in revised form 12 August 2021; Accepted 25 August 2021

Table 1
Global Moran's I statistic comparison at borough and MSOA levels.

Crime Type	Borough-Level		MSOA-Level	
	Global Moran's I Statistic	*p*-value	Global Moran's I Statistic	*p*-value
Bicycle Theft	0.033	0.052	0.598***	1.47E-222
Burglary	0.015	0.257	0.382***	1.036E-92
Theft from the Person	0.078	0.097	0.544***	1.328E-214
Vehicle Crime	-0.086	0.703	0.370***	6.724

*** indicate significance test over 99%

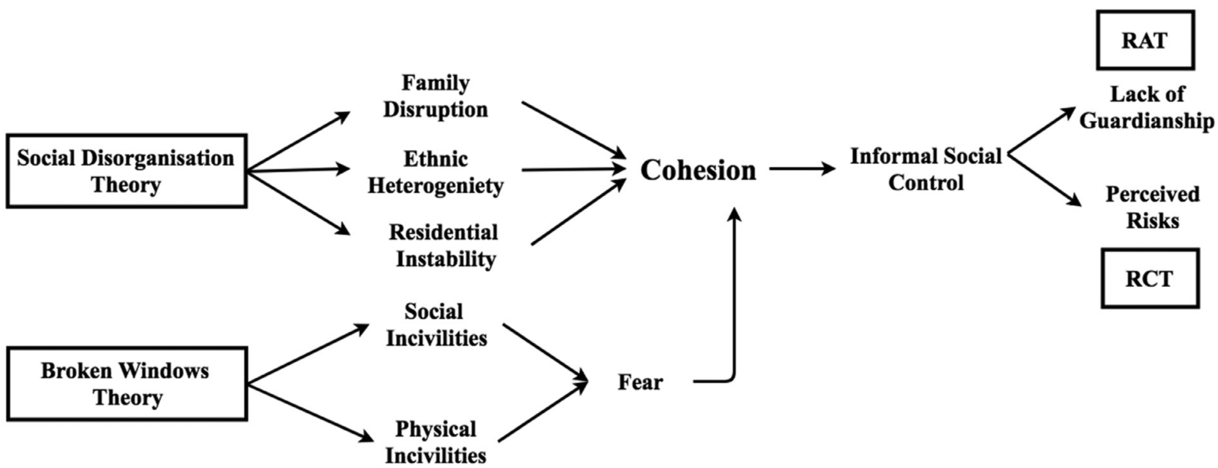

Fig. 1. Revised from Steenbeek and Hipp [20].

ence orderings for all possible outcomes [1], which renders a higher probability of criminals choosing to commit a crime once the corresponding rewards outweigh the costs, with the latter more refers to getting arrested or subjectively perceived probability of being arrested. Cohen and Felson [3] assembled the explainers of changes in crime rates at neighborhood level into three converged strands: motivated offenders, suitable targets and without capable guardianship (see Appendix 1), to interpret the occurrence of acquisitive crimes via temporal and spatial interactions. Scaling up from individual to neighbourhood level, disorders denoted by both social disorganization theory [8] due to unrestricted behaviors of residents and broken windows theory [9] on visible signs of disorders encouraging criminalities, had been emphasized; but still under the presumption that, higher crime rates are originated directly from the reduced informal social control rather than disorders [10]. Apart from the emphasis on human elements in routine activity theory, crime pattern theory focuses on the traits of places and categorizes locations into crime attractors (non-residential facilities that attracting many individuals, including both would-be offenders and potential targets) and crime generators (venues offering well-known opportunities for crime) [18].

Although aforementioned theories have varied emphases, they have a shared hinge on "Cohesion" (Fig. 1). Cohesion in communities will decline if (1) informal social control reduced [19], and further result in increased difficulty for people from disruptive households in getting connected with others; (2) higher residential instability due to frequent movements, which makes people interact with their neighbours inactively; (3) higher compound effect from ethnic heterogeneity and cultural variation, which hinders the communications among residents. In another word, mistrust, prejudices, and misunderstanding incurred by lack of interactions, would weaken social ties and cohesion in neighbourhoods.

In broken window theory, those tangible signs of social incivilities (e.g. public drinking, public peeing) and physical incivilities (e.g. trash, cigarette butts, abandoned vehicles) can create fear and a tense atmosphere in communities [21], inducing residents' more apt to concern their ambient environment for personal safety against perceived dangers [22], and more likely to take others as untrusted and suspicious [23]; in return, residents tend to avoid interactions with one another [24], isolate themselves [21] and further lead to lower cohesion in neighborhoods. This weak cohesion was explained by Wilson and Kelling [25] as declined "sense of mutual concerns and responsibilities" of residents, and this "no one cares" atmosphere will further encourage higher crime opportunities in that, motivated offenders may perceive less risks in being arrested and thus inclining to making criminogenic decisions; and the loss of mutual supervisions may result in the lack of guardianship for properties in neighborhoods. Therefore, crime would see a rise when informal social control declines.

Concerning about acquisitive crimes, there are many literature demonstrated the essentiality of crime opportunities, with emphasis on the significant relationship between various types of acquisitive crimes and three factors (cost, return and perceived risks) illustrated in routine activity theory [26,7,27], which also substantiates the generality of rational choice theory in instrumental crimes [6]. To measure returns and cost before commiting crimes, the potential criminals would investigate into deterrence and individual-level perceived risks, with measurements like the term of imprisonment [28,29,30] as deterrent factor and extracted data from retrospective surveys as individual-level perceived risk, especially at the sharp period when crimes happen [6,7,31].However, retrospective surveys data may result in "the causal ordering of the variables contracting their temporal order of measurement." [1], hence, to cope with this misalignment problem, Tweets data will be employed as corresponding indicators. To measure influential factors for acquisitive crimes, routine activity theory has suggested that, motivated offenders are often measured by the percentage of young males, unemployed populations or improvised groups [32,33]; the level of guardianship could be measured by varied indicators against different types of crime. For example, Larsson [34] chose the unemployment rate to explore patterns of burglary when occupants are away from their residences, while Louderback and Sen [35] used the number of residents in neighborhoods for street robbery reckoning that inhabitants are likely to notice and recognize strangers; last but not the least, the measures for suitable targets in this work are derived from Twitter and Foursquare data by location, which is also consistent with crime pattern theory, with number of visitors around attractions (e.g. recreational venues, college facilities), and the frequency of visits as measurements. In addition, binary variable on numbers of different crime-prone venues, such as bars, convenience stores and schools, has been considered to explore the potential associations with crimes like theft from persons, shoplifting and bicycle theft, etc. [36,18,37,5,38], and further identify the importance of places accounting for crime hot spots [12].

Measures appropriate to test either social disorganization theory or broken window theory were rare but keeping evolving and integrating, besides of the initial measurement from survey data like British Crime Survey [39] and resident surveys [40,41], several methods intending to collect "big data" have emerged. For example, Sampson and Rausenbush [19] derived visible incivilities data from vehicle's videotamping in over 80 communities in Chicago; Steenbeek and Hipp [20] recruited many volunteers to score "broken windows" for each 100-meter in Amsterdam; O'Brien and Sampson [42] took an unprecedentedly innovative attempt to manually annotate over 200,000 requests from non-emergent 911 calls, based on incivilities from Boston's official datasets. O'Brien et al. [42] further suggested the worthiness of using crowdsourced datasets such as social media messages and cell phone records to measure "broken windows". In recent years, criminologists increasingly apply social media data, e.g., geo-tagged tweets, into of crime patterns analysis [16,43].

Accordingly, techniques like Latent Dirichlet Allocation (LDA) had been applied to extract information on crime events in tweets for more accurate crime prediction modelling [14], and to predict crime incidents better from nosier tweets' content [17]; sentiment analysis (SA) on tweets' content had been applied for crime analysis [13], with the assumption that human's emotions and attitudes can reflect their past and potential behaviors [44]; data mining techniques, e.g. SVM (Support Vector Machine) and RF (Random Forest) classifiers, are employed to classify voluminous numbers of tweets about hate crime [15]; Williams et al. [45] also demonstrated that the number of tweets on disorder can be used as an effective measurement for illustrating crime rates, on basis of broken windows theory using fixed/random effects model. Besides of tweets data, Foursquare data also received some attention from researchers. Quercia and Pompeu [46] successfully substantiated those different kinds of venues have certain types of associations with neighborhood deprivation using London's Foursquare data. Inspired by the work, Kadar et al. [47] drew a conclusion that crowd-source are better proxies to human daily activities via predicting New York City's crime rates using check-ins data derived from Foursquare; Boni and Matthew [48] further combined them with tweets to construct a predictive model by using the geotagging functionality.

It could be arrived at this stage that, to measure the cohesion at neighbourhood level towards crime prevention evidence, requires a comprehensive integration of solid criminological theories, both crowdsourcing and officially recorded data, and cutting-edge optimal methods. However, majority of empirical studies only focus on supporting sole or limited mixture of theories, hitherto, few studies investigated into crime patterns using mixed frameworks. Besides, although the application of social media data by criminological researchers witnessed increases in recent decades, only locational information was largely extracted for majority of such studies, whilst valuable textual content of tweets was regretfully discarded without efforts to try to interpret changes in crime incidence. Therefore, this paper managed to realise such comprehensive integration of criminological theories, spatial analysis point of view and multiple sourced data, especially those crowdsourced, with London as the pilot study, and try to sew up the "patchy mat" through cutting-edge techniques. Pertained with the theoretical framework discussed above and data derivation techniques with social media data, they could be translated into 2 elements for crime opportunities, subjective perceived risks, and the level of guardianship, and further applied to construct the measures on basis of social media data and cross-sectional surveys (shown in Fig. 2), with 2 hypotheses to test with in this paper:

- *Hypothesis 1*: the increase of Tweets relating to broken window theory could lead to higher rates of acquisitive crimes
- *Hypothesis 2*: more negative sentiments in Tweets may cause increases on the level of acquisitive crimes.

3. Data & methods

The study area is London city at MSOA (Middle Super Outcome Area) level, a finer scale than the widely applied Borough level for empirical studies, with steady results on significant relationships [49,45]. Although it could be ideal to conduct crime analysis at smaller geographical units [50], i.e. LSOA (Lower Super Outcome Area) level, in consideration of spatial autocorrelation [51], the regression results might expect tremendous distortion due to zero inflation of social media data. Hence, MSOA level could be ideal as analytical unit for this paper, which has also verified by Global Moran's I^3 test (Table 1), with more significant spatial association at MSOA level than at Borough level.

3.1. Data

Crime incidents point-data in London from police records (01/02/2017 to 31/01/2018) had been aggregated into measure "crime rates (*per 1,000 persons*)" for four types of acquisitive crimes, bicycle theft, burglary, theft from the person and vehicle crime, and further been applied in the model as dependent variables; whilst independent variables were derived from:

(1) **Cross-sectional data**: UK Census 2011 data, Indices of Multiple Deprivation (IMD) in 2015, and estimated population in 2016 from Office for National Statistics; such socio-economic and demographic data were the most up to date for research area, being consistent among London MSOAs to suggest contextual influences on local crime incidents, hence were taken to be sufficient and align with current research.
(2) **Twitter data**: Geo-tagged tweets within defined bounding box of London from February 2017 to January 2018 (detailed methods were provided in Appendix). After geocoding and removing duplicates, 1,836,747 tweets with geographical information were identified to derive two indicators on:
    - "broken windows" (*BW Tweets*): calculates the number of tweets mentioning any visible incivilities at MSOA scale. Inspired by the text classification from Burnap and Williams [15] and Williams et al. [45], supervised machine learning classifiers are trained and used to distinguish the tweets between *BW* categorization and/or non-*BW* ones based on LDA (Latent Dirichlet Allocation) technique: (1) based on the designated number of topics, to generate a matrix of the probabilities of a document belonging to each one; (2) an iterative process to construct classifiers for those posts. Details of building the classifiers could be found in Appendix 2. Areal tweets will be labelled with positive, negative and neutral via sentiment score, which is widely used improving the accuracy of models for crime prediction [13]; and acting as a reflection of offline sentiment [52] and local environment.
    - For example, "*Spotted on a #street in #Kennington. You know who you are, so make sure you dump a black bag of* #rubbish *on a #Wednesday*"; "*We finally escaped the confines of our lurgy infested house! Fresh air feels so good*" and "*I know it's vandalism illegal graffiti for traffic signs #London*".
    - Normalized average sentiment scores. A universal classifier has been built based on the completed Stanford Sentiment 140 Tweets Corpus containing 1.6 million classified tweets, where

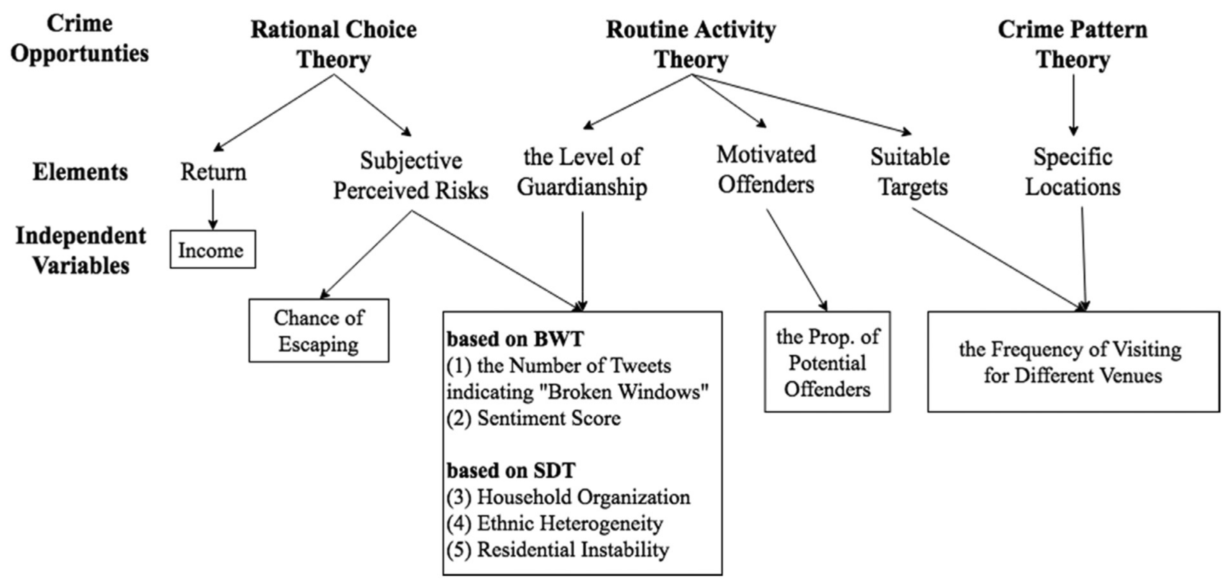

Fig. 2. Measures for independent variables.

Table 2
Weighted number of venues categorized into 10 aggregated categories.

Venue Category	Counts (%)
Residence	5135.665(3.992%)
College and University	3682.666(2.862%)
Professional and Other Places	26558.334(20.641%)
Arts and Entertainment	4345.328(3.377%)
Outdoors and Recreation	10077.333(7.832%)
Food	27107.164(21.068%)
Nightlife and Sport	6988.83(5.432%)
Shop and Service	29728.997(23.106%)
Travel and Transport	15014.666(11.670%)
Event	16(0.012%)
Total	128665

emoticons have been and appropriately handled. Inspired by method of Bryl's (https://analyzecore.com/2017/02/08/twitter-sentiment-analysis-doc2vec/) and Mitchell et al. [53], this study utilizes the probabilities ranging from 0 to 1 as substitute for three categorizations—positive, neutral and negative in original corpus. Hereby tweets with probabilities from 0.35 to 0.65 can be regarded as neutral based on the distribution of sentiments in corpus. In addition, *doc2vec* algorithm [54] was used to draw the context of each phrase and processing negative expressions (e.g. "not bad"). Further data processing will be discussed in following section.

(3) **Foursquare data**: Foursquare API is employed to access basic information about each venue' names, categories, GPS and check-ins. Finally, 132,996 out of 147,994 venues were scraped in Greater London area (Detailed methods were provided in Appendix), which follow into 5,250 individual categories and 10 aggregated classifications (Table 2).

In Table 2, "Event" is discarded due to small count, thus 9 different explanatory variables are calculated through dividing weighted number of venues for each type by corresponding count of visiting. The frequency of visiting to these 9 venue categories could be used as predictors to measure suitable targets illustrated in routine activity theory, with tweets data applied for calculating the average frequency of visiting by venue types; specific types of locations could also indicate crime generators and attractors[5] as illustrated in crime pattern theory. The underlying rationale is that, for crime types like thefts from person, thefts from vehicles and bicycle thefts, suitable targets normally refer to unattended belongings which largely relates to visitors' probabilities of using vehicles [55]. As the larger figures for people's visiting non-public-transportation facilities, the higher possibility in using vehicles hence higher exposure to would-be offenders, and further higher crime rates on bicycle theft and vehicle crimes. For burglary crime, the suitable targets for potential burglars are mostly unattended houses and flats with valuable items. Rather than measuring the average number of tweets around residences, as some unexpected noise data could be generated by temporary passers who can't be regarded as the effective guardianship [35], instead the average frequency of visiting non-residential areas can be utilized as measure with the assumed guardianship reduction accordingly [34]. In consequence, burglary incidence may increase when people arrive more at non-resident facilities. To reflect multiple visits for geocoded venues in each category, people posting tweets are counted as "visitors" when their distance are less than 10m, which is identified by experimental regression analysis (4 threshold levels at 5 m, 10 m, 15 m and 20 m respectively were set due to the positioning errors and prior research [48], but 10 m turned out to performance best hence being the optimal threshold level); it rendered some of venues are labelled with multiple tags, thus weighted vectors for each venue are constructed. For example, a venue called "Club Africa" may have three different tags named as "African Restaurant", "Music Venue" and "Nightclub", which respectively corresponds to "Food", "Art and Entertainment" and "Nightlife and Spot" categories in Table 2, hence the weighted vector for this venue is (0,0,0,1/3,0,1/3,1/3,0,0,0); another venue "Hair by Laura" with single label "Shop and Service" accordingly vectored as (0,0,0,0,0,0,0,1,0,0). Hypotheses on relationships between specific type of crime and people's visiting to venues were proposed accordingly.

The independent variables were derived from cross-sectional datasets to measure 5 dimensional indicators on:

(1) Perceived risks / the levels of guardianship

Indicators like household organization, residential instability and ethnic heterogeneity [35,56,20] are constructed using dataset derived from IMD 2015 and Census UK 2011. (1) The household organization index is a scaled sum of the percentages of various households except for couple with dependent child household and couple without dependent household; (2) The residential instability index is calculated by summing up the proportion of private rents and household spaces; (3) The

Table 3
AIC derived from NBR and GWR.

	Bicycle Theft			Burglary		
	NBR	GWR		NBR	GWR	
Akaike Information Criterion (AIC)	raw rate 3282.1	raw rate 4732.64	normalized 2064.47	raw rate 5077.90	raw rate 5783.83	normalized 2491.20
Akaike Information Criterion (AIC)	Theft from the Person			Vehicle Crime		
	NBR	GWR		NBR	GWR	
	raw rate 4043.70	raw rate 8360.85	normalized 2677.58	raw rate 5820.50	raw rate 6173.23	Normalized 2397.12

ethnic heterogeneity is calculated via Eq.1 below [57] then standardized against census data.

$$\text{EthnicHeterogeneity} = 1 - \sum_{i=1}^{n} \left(\frac{k_i}{N} \right)^2 \tag{1}$$

where N is the number of all population, k represents for the number of each race and i denotes five different races (white, mixed/multiple ethnic groups, Asian/Asian British, Black/African/Caribbean/Black British and other ethnic group). The ethnic heterogeneity has been taken to be with positive relations with crime rates.

(2) Chance of escaping

Many studies have demonstrated that rapid spread of public transport stations such as newly constructed bus stations [55] and metro stations [58] can encourage crimes in neighborhoods [59,60], in light of rational choice and crime pattern theories, that transit stations can attract massive commuters (crime generators) thus facilitate would-be offenders' access to potential targets. Hence, the density of public transportations would be taken as the measure for offenders' chance to get escaped. Being different from prior empirical studies, this research took into consideration of various types of public transportations, like underground, bus station, tram station, taxi stand and train station, derived from Foursquare, then further divided by MSOA area accordingly followed by standardisation to generate the density.

(3) Return

It is assumed that the potential rewards from committing acquisitive crimes have positive relations with income index, which is used to indicate the potential return if would-be offenders choose to commit crimes. Namely, would-be offenders tend to commit a crime when the level of income rising because more rewards are likely to be obtained. Using IMD 2015 data, income index for each MSOA is generated under the assumption that, higher level of income index represents for increasing deprivation level.

(4) Motivated offenders

Motivated offender index is a standardized sum of the percentage of males from 18 to 35 years old [32]; the employment index is calculated by using IMD 2015 at MSOA level [33], which was believed that greater motivation on committing crimes could be generated with higher unemployment rates; and the proportion of low-income households indicated by routine activity theory, is regarded as reflecting the motivation for committing crime due to demand of more money for living.

(5) Control variables

Local population density at MSOA is used as control variable in the analysis.

3.2. Negative binominal regression

Since the selected variables are skewed and over-dispersed, negative binominal regression is more appropriate to model the crime rates of four different types based on previous studies [35,61], with crime rates variable being normalized considering reduced AIC nature of the data (see in Table 3). Furthermore, in light of the geographical disparities in socioeconomic conditions and demographic situations across space [62,63], it has long been observed of the uneven spatial distribution of crimes. However, majority of empirical studies only apply at aggregated level (e.g., boroughs in London) whilst neglecting local variations; thus, in this research, local Moran's I has been utilised to investigate the spatial patterns (clusters or outliers) of four crime types respectively [64,65]. In addition, GWR is used to present the local disparities regarding the associations between outcome variable and predictors across the space [66].

4. Results & discussions

From descriptive of variables in Table 4 below, there is a considerable variation of crime rates (per 1,000 residents) among four different types of crime, and significant skewed distribution supporting the utilization of negative binominal regression, as well as the necessity in testing their geographical associations by applying spatial analysis (i.e. Global *Moran's I* Statistic in Table 1).

Fig. 3 below presents 4 types of crime respectively in that, the (a) spatial distributions of crimes and the (b) Local Moran's I, presenting statistically significant high-high clustering of crime rates. Universal spatial pattern regardless of crime types could be spotted that, clusters of high crime rates tend to accumulate in the northern side of the Thames (especially in the City of London and Westminster Borough), indicating comparatively higher levels of property-related crimes in these densely populated and economically developed MSOAs.

However, there was obviously substantial disparities between the northern side and the southern side of the Thames, and the distribution pattern varied by crime type as well in that, vehicle crime and burglary tend to be more scatter-distributed than bicycle theft and theft from the person crimes. The possible reason related to crime's spill-over effect incurred by offender's escaping over MSOA boundaries, which further made explainable annotation on the low-high outliers' existence for burglary and vehicle crime [67].

4.1. Negative binomial regression (NBR)

To explore the underlying driving forces for such crime distribution pattern, while being referred by social disorganisation and broken window theory, indicators for crime opportunities were processed and integrated into the negative binomial regression model. Each type of crime rates (per 1,000 residents, 02/2017–01/2018) was taken as the dependent variable, while independent variables derived from Twitter, Foursquare and cross-sectional surveys datasets were processed for each of the four models with varied crime type, and estimated coefficients could be found from Table 5.

Firstly, the increasing number of tweets tagged with "broken windows" had significant positive association with the incidences of all four acquisitive crimes as expected [45], but once aggregating the tweets by

Table 4
Descriptive statistics for variables at MSOAs in London.

MSOA (Middle Super Outcome Area): n=983				
Variables	Mean	SD	Min.	Max.
Dependent Variables				
Bicycle Theft	2.508	3.885	0.000	47.547
Burglary	9.021	5.338	1.599	73.977
Theft from the Person	5.598	18.008	0.000	333.703
Vehicle Crime	12.287	6.826	2.575	62.352
Independent Variables				
Motivated Offenders				
Potential Offenders Index	0.700	0.249	0.219	2.094
Suitable Targets				
F.Residence	1.046	3.573	0.000	49.667
F.College.and.University	2.209	29.907	0.000	853.333
F.Professional.and.Other.Places	1.287	3.564	0.000	61.629
F.Art.and.Entertainment	1.530	9.828	0.000	193.333
F.Outdoor.and.Recreation	1.199	3.06	0.000	43.097
F.Food	1.341	3.861	0.000	89.429
F.Nightlife.and.Sport	1.752	9.663	0.000	262.300
F.Shop.and.Service	1.325	2.928	0.000	37.577
F.Travel.and.Transport	1.404	5.572	0.000	96.81
Lack of Guardship/Perceived Risks				
BWT: Traditional Measures				
BW Tweets	98.761	216.097	0.000	3960.4
Sentiment.Score	0.662	0.027	0.501	0.743
SDT: Traditional Measures				
Households.Organization.Index	0.758	0.103	0.456	0.952
Ethnic.Heterogeneity.Index	0.507	0.153	0.074	0.741
Residential.Instability.Index	0.278	0.122	0.051	0.771
Chance of Escaping without Arrest				
Transport.Station	7.018	11.383	0.000	142.463
Return(Committing Crime)				
Income.Index	0.812	0.402	0.075	2.963
Control Variable				
Population.Density	87.120	50.301	2.883	266.601

area, it was found that no significant relationships between crimes and average sentiment score of areal tweets. It could be interpreted that more online tweets regarding of “broken windows” might signify the rises in offline incivilities and turn out to magnify the “no one cares” atmosphere; in the meantime, potential offenders may feel less risky in getting caught and less worried about being stopped even committing crimes (reduced level of guardianship). The sentiment score could reflect levels of online rudeness, discourtesies and negative attitudes, which might exert adverse effect on users’ emotion and further be converted into violence-related or hate crimes [13].

Secondly, the household organization and residential instability indicators, both of which were measures for information social control in community, showed significantly positive relations to crime incidences regardless of their types. Empirical studies found that people from broken families are more prone to have weaker sense of connection with their neighbourhood; higher level of residential instability indicates more frequent movements which further discourages people to communicate with their neighbors proactively. Consequently, reduced cohesion from the lack of mutual interactions would lead to less informal social control, hence lower risks for offenders and more crime occurrences.

Thirdly, ethnic heterogeneity was significantly negative with bicycle theft, burglary and theft from the person, but no significant relationship with vehicle crimes. This finding contradicts to majority of empirical studies in that, higher level of racial disparities would contribute to growing crime occurrences [68,69,70]. Traditionally, ethnic heterogeneity is regarded as an important factor contributing to the reduced informal social control, because the cultural or language disparities reflected in theories might produce prejudices, discriminations and shortage of communications; besides, ethnical minorities have long been considered as disadvantaged in terms of economic status, employment and etc., which gradually built up the stereotype on link between minorities and higher crime rates. Adapted to contemporary change, Churchill and Laryea [71] proposed reliable justification for this counterintuitive phenomenon from education perspective: the minorities receiving education in the Greater London became to take up more and more proportions over time [72]. This may improve their understanding, tolerance and respect for diversity through intensive education hence reduce and further eliminate the misunderstandings and mistrust originated from lack of communications. In addition, from economic perspective, these well-educated minority groups tend to have higher possibility in guaranteeing decent and well-paid jobs, thus reduce the chance of getting into poverty or property-related crimes.

Fourthly, measure for perceived risks of escaping after committing crimes is the density of public transport stations (e.g. bus stations, taxi stand, underground stations). It can be noticed from Table 5 that, in consistent with the theoretical hypothesis, denser transportation hubs may attract more crimes except for vehicle crimes. Because offenders normally are apt to escape from criminal spots right after criminalities (i.e., bicycle theft, burglary and theft from the person), and densely populated public transport stations could further facilitate the coverage of their traces, reduce their risks of getting arrested, thus increase the probabilities of would-be offenders in committing crimes. The exceptional case for vehicle crimes could be explained by the inability for offenders in getting escaped by public transportation with the stolen cars/car parts/valuable items from cars, which contributed to their insignificant relationship.

Lastly, income index displayed significantly negative associations with all types of crime upon controlling of all other factors. It is been taken that higher income level denotes more rewards in committing

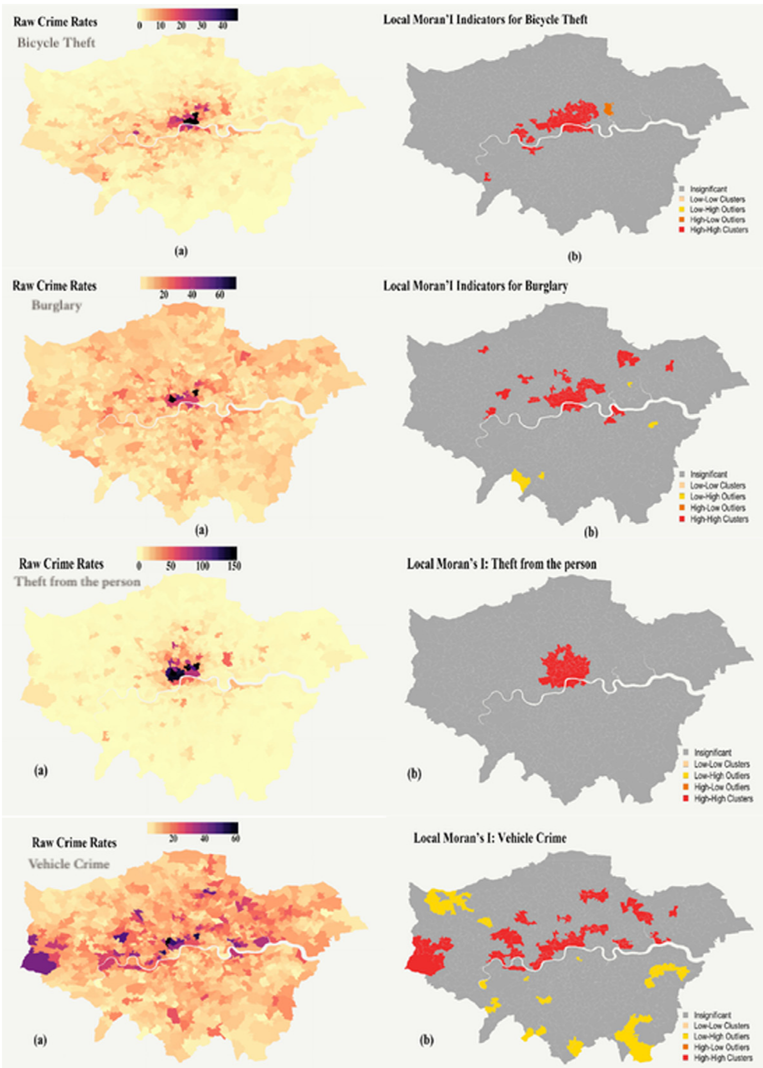

Fig. 3. (a) Spatial distribution for 4 types of crimes among MSOAs respectively; (b) Local Moran's *I* Values for 4 types of crimes among MSOAs respectively.

crimes, hence encourages would-be offenders to make criminogenic decisions. In consistent with hypotheses from routine activity theory, the normalized sum of males aged between 18 and 35, the unemployed and those households with low incomes were found to be positively related with four types of crime. Multiple collinearity issue was also considered with taking the variance inflation factor values into consideration, but it seems that the results are significantly away from the threshold, hence current coefficients estimates in both Negative Binominal Regression and Geographical Weighted Regression models are reliable.

In Table 5, the venues labelled as College and University (e.g. teaching buildings, students' residence hall) had significantly positive relationships with all types of crime rates, which is consistent with both crime pattern theory and routine activity theory. For the former theory, College and Universities can attract a large number of individuals (i.e., students, teaching staffs and other members, visitors and tourists) thus increase the probability of acquisitive crimes occurrence. Since majority of the campuses are fenceless, any motivated offenders can easily get nearer to ideal targets; for the latter theory, college students are relatively vulnerable due to lack of awareness of guarding their valuable belongings without parents' endless reminders [73], rendering higher exposure to motivated theft offenders. Besides of thefts, burglary is on the same track due to weak mutual guardianship, low level residential stability, weak bonds between short-term occupants and individual carelessness in students' dormitories.

The venues for Art-and-Entertainment (e.g. theater, gallery, museum) and Outdoor-and-Recreation (e.g. park, pitch, beach) had both presented positive associations with bicycle theft, whilst the latter further relates positively to vehicle crimes. Borrowed from crime pattern theory, these venues can generate more crimes upon attracting massive individuals; however, the former category, Art-and-Entertainment, normally has more managed parking lots with CCTV, whilst the latter category, Outdoor-and-Recreation, is near to unenclosed, fenceless and open-air parking lots hence simultaneously play the role of crime attractors providing more opportunities to potential offenders. For former category solely, the number of visitors has significantly positive correlation with burglary by providing more attractive targets—a growing number of vacated houses and unattended flats while the occupants are out at entertainment venues.

The theft from the person crime would see a rise if an increase in the frequencies of visiting venues pertaining to residential areas, college and university, art and entertainment as well as travel and transport. With most of occasions individuals as the suitable targets, potential offenders choose such direct aim in their familiar areas, thus the increasing exposure from these suitable targets will produce higher crime rate. Those aforementioned venues are crime attractors appealing to a large number of suitable targets, especially those with greater accessibility like Travel and Transport [74]. Finally, the only one significantly negative relationship with crimes could be spotted from professional and

Table 5
Results of negative binomial regression.

	(MSOA, N=983)	Bicycle Theft	Burglary	Theft from the Person	Vehicle Crime	
ROUTINE ACTIVITY THEORY	INTERCEPT	0.166 (0.098)	2.026*** (0.112)	0.428*** (0.112)	2.363*** (0.046	CRIME PATTERN
	Motivated Offenders Potential.Offenders.Index	0.520*** (0.07)	0.385*** (0.036)	0.591*** (0.087)	0.36*** (0.039)	
	Suitable Targets/Venue Types F. Residence	0.058 (0.033)	0.012 (0.016)	0.139*** (0.038)	-0.004 (0.017)	
	F. College. and. University	0.095*** (0.031)	0.046*** (0.016)	0.106** (0.036)	0.035* (0.017)	
	F.Professional.and.Other.Places	0.028 (0.040)	-0.017 (0.019)	-0.024 (0.046)	-0.038* (0.019)	
	F.Art.and.Entertainment	0.126*** (0.033)	0.045** (0.016)	0.099* (0.038)	0.033 (0.017)	
	F.Outdoor.and.Recreation	0.066* (0.035)	0.032 (0.017)	0.051 (0.041)	0.048** (0.017)	
	F.Food	0.046 (0.040)	0.021 (0.019)	-0.017 (0.046)	0.019 (0.019)	
	F.Nightlife.and.Spot	0.037 (0.034)	0.006 (0.017)	0.005 (0.040)	-0.008 (0.017)	
	F.Shop.and.Service	-0.038 (0.040)	-0.018 (0.019)	0.001 (0.046)	-0.013 (0.019)	
	F.Travel.and.Transport	0.060 (0.038)	0.011 (0.018)	0.092* (0.043)	0.019 (0.019)	
	Guardianship/Perceived Risk **(1) BWT: Social Media Measures** BW Tweets	0.100*** (0.025)	0.034** (0.012)	0.107*** (0.029)	0.028* (0.012)	RATIONAL CHOICE THEORY
	Sentiment.Score	0.043 (0.027)	0.012 (0.013)	0.028 (0.032)	0.02 (0.013)	
	(2) SDT: Traditional Measures Households.Organization.Index	0.275*** (0.046)	0.051* (0.021)	0.478*** (0.054)	0.086*** (0.022)	
	Ethnic.Heterogeneity.Index	-0.262*** (0.038)	-0.046** (0.017)	-0.193*** (0.045)	-0.032 (0.018)	
	Residential.Instability.Index	0.143*** (0.029)	0.09*** (0.015)	0.34*** (0.036)	0.146*** (0.016)	
	Risk (Chance of Escaping without Arrest) Transport.Station	0.098*** (0.021)	0.044*** (0.013)	0.242*** (0.029)	-0.036* (0.016)	
	Return (Committing Crime) Income.Index	-0.437*** (0.069)	-0.442*** (0.035)	-0.567*** (0.085)	-0.384*** (0.037)	
	Control Variable Population.Density	0.012 (0.033)	-0.127*** (0.017)	0.016 (0.039)	-0.136*** (0.018)	

***p<0.05, **p<0.01, ***p<0.001**

Table 6
Variations of independent variables for four acquisitive crime types.

	Bicycle Theft				Burglary				Theft from the Person				Vehicle Crime			
	mean	sd	min.	max.	mean	sd	min.	max.	mean	sd	min.	max.	mean	sd	min.	max.
(INTERCEPT)	1.520	1.685	-5.718	11.444	7.867	2.498	-0.193	17.618	1.764	24.906	-250.07	219.82	11.034	2.193	1.945	31.757
Motivated Offenders	1.552	2.243	-5.039	10.544	4.329	3.161	-7.120	15.632	1.644	19.088	-211.78	82.796	5.393	3.420	-21.13	18.642
F.Residence	0.199	0.584	-3.626	3.597	0.088	1.111	-13.91	2.942	0.521	4.262	-31.820	52.795	0.070	0.628	-4.411	2.281
F.College.and.University	0.184	0.492	-2.242	2.241	0.164	1.260	-10.12	2.825	0.909	5.334	-13.408	86.781	0.514	1.442	-7.704	20.599
F.Professional.and.Other.Places	-0.089	0.520	-2.211	1.637	-0.287	0.872	-3.505	1.879	-0.533	10.074	-52.673	220.92	-0.759	0.866	-4.119	1.214
F.Art.and.Entertainment	0.224	0.473	-1.383	3.154	0.632	1.342	-9.118	11.117	1.520	6.555	-19.049	98.971	0.598	1.025	-6.079	5.200
F.Outdoor.and.Recreation	0.078	0.396	-1.997	1.220	0.215	0.662	-1.753	6.646	-0.406	4.920	-76.217	12.909	0.403	0.744	-3.091	2.818
F.Food	0.145	0.531	-2.131	3.114	0.190	0.841	-3.630	4.042	-0.512	6.753	-105.54	44.837	0.246	0.878	-2.149	7.125
F.Nightlife.and.Sport	0.029	0.516	-2.693	2.588	0.026	0.928	-9.878	2.650	0.974	5.988	-87.726	71.789	0.031	0.768	-2.464	2.430
F.Shop.and.Service	-0.222	0.729	-3.732	4.156	-0.176	1.344	-3.531	17.355	-0.994	6.552	-79.843	19.982	-0.246	1.075	-2.109	8.613
F.Travel.and.Transport	0.156	0.472	-1.900	2.266	0.213	0.900	-3.501	5.497	1.449	6.421	-39.912	61.936	0.359	0.996	-3.304	2.374
BW Tweets	0.148	0.371	-1.259	1.870	0.126	0.563	-2.186	1.894	-0.25	3.877	-50.754	13.305	0.347	0.428	-0.573	2.077
Sentiment.Score	-0.056	0.401	-2.386	1.891	0.002	0.476	-4.371	1.360	-0.499	3.798	-35.708	19.319	0.295	0.558	-1.278	2.400
Households.Organizations.Index	0.733	1.552	-5.600	9.237	0.487	1.957	-4.570	9.652	3.885	13.791	-48.726	115.9	1.302	1.474	-5.283	9.080
Ethnic Heterogeneity	-0.856	1.706	-8.338	1.587	-0.802	1.473	-5.038	2.953	-1.972	9.951	-81.875	81.945	-0.967	1.705	-4.347	3.051
Residential.Instability	0.321	0.890	-3.194	3.850	0.574	1.342	-7.779	7.452	0.928	4.488	-21.445	32.751	1.179	1.446	-3.685	6.994
Transport.Density	1.148	1.517	-4.605	16.169	0.685	1.871	-7.614	12.251	2.407	7.940	-31.570	115.66	-0.348	2.142	-6.757	29.164
Income Index	-1.416	2.245	-10.675	5.680	-4.602	3.271	-16.17	5.251	-0.719	21.832	-73.581	258.04	-5.677	2.820	-16.69	13.033
Population.Density	-0.32	1.055	-6.504	3.742	-1.298	1.477	-6.210	6.942	-0.845	5.579	-40.536	32.452	-1.753	1.330	-6.406	11.316

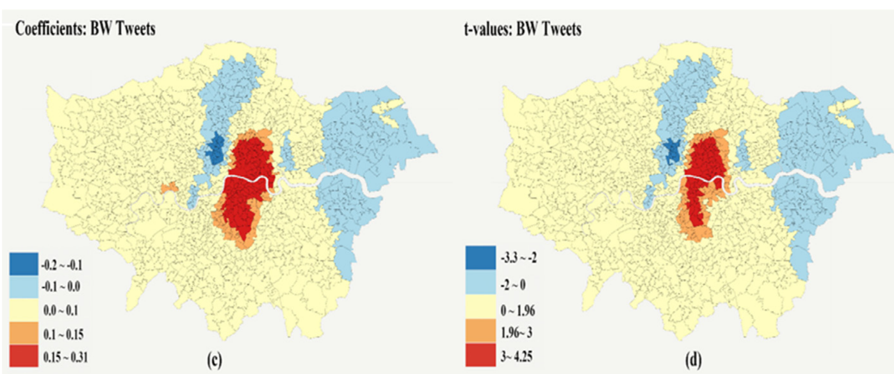

Fig. 4. Coefficients of BW tweets and corresponding t-values from GWR with bicycle theft.

other places. These venues mainly refer to governmental buildings like council hall, police stations, etc., indicating stricter management and surveillance, which might be interpreted as crime deterrent venues to offenders.

4.2. Geographically weighted regression (GWR)

Taking the spatial spill-over effect reflected in Fig. 3 into consideration of regression model, the corresponding Geographically Weighted Regression had been implemented and further compared with Negative Binomial Regression results as shown in Table 5.

It could be arrived that GWR (for normalized rates) is complementary to NBR results upon estimating the local variations with dramatical reduction on AICs values. Furthermore, spatial pattern of crime rates showed discrepancies across the 983 MSOAs (Table 6) due to varied characteristics, which were depicted by Twitter, Foursquare and cross-sectional data.

With varied spatial distribution patterns of each predictor, some exemplary extracts were derived and presented below from Figs 4 to 7 for illustration. In Figs. 4 and 5, the Tweets indicator (BW Tweets) indicated by broken window theory could be used as measure for the level of guardianship from routine activity theory and subjective perceived risks from rational choice theory.

In Fig. 4, the value for BW Tweets ranged from -0.21 to 0.31 with 793 MSOAs showed positive coefficients, among which 114 were significant (t-value > 1.96). It could be located on the map that, all these

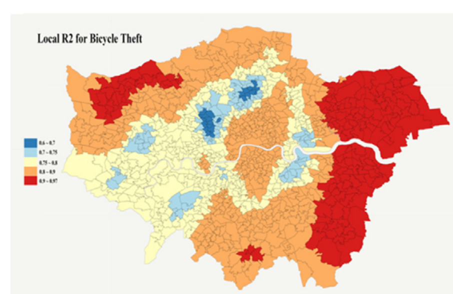

Fig. 5. The geographical distribution of local R^2 values for bicycle theft.

114 MSOAs presenting significant positive relations between BW tweets and bicycle thefts (orange and red in Fig. 4d) were majorly around the Thames midstream areas (i.e., City of London, Tower Hamlets, Hackney, Southwark and Lambeth), with the highest cycling flows according to Strategic Cycling Analysis [72]. Remaining areas with lower utilization of bikes [72] hence can't tempt potential offenders with enough suitable targets, which might arise from residents' feeling of unsafety or disorder due to flooding tweets information on incivilities. GWR model used for predicating their relationships (Fig. 5) showed varied results over space (Local R^2=0.6~0.97).

Similarly, upon controlling other variables, influences from residential instability on burglary varied from -3.112 to 9.110 (Fig. 6), with

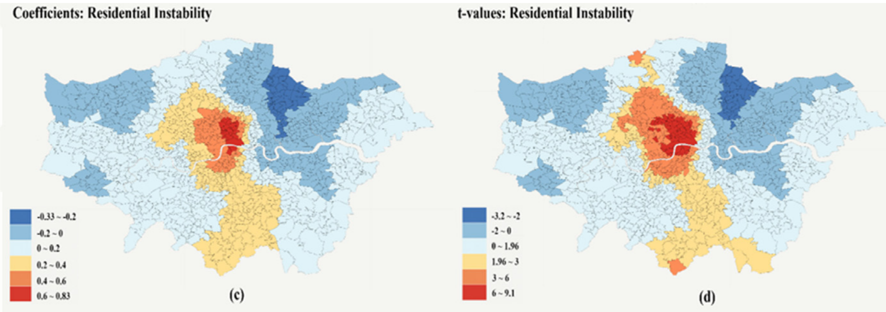

Fig. 6. Coefficients of residential instability and corresponding t-values from GWR with burglary.

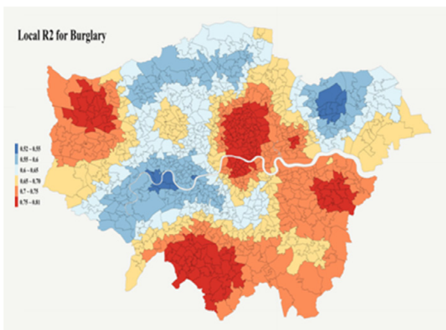

Fig. 7. The geographical distribution of local R^2 values for burglary.

over 70% (692 out of 983) MSOAs showed positive associations, within which 277 MSOAs were significantly positive (t-value > 1.96). it is well in support of translating social disorganization theory into an indicator for crime opportunities. For those remaining MSOAs (blue and lighter blue in Fig. 6) with relatively lower house prices, they didn't present significant relations with burglary due to their being lack of crime-attractive valuable items for potential burglars [75]. Besides of the aforementioned, there were some outlier places where house prices are high but have no significant relations with burglary. It might be induced by the spillover effects and infectious characteristics of burglary [67] in that, properties suffering from repetitious burglaries, or so called "near repeat victimization", tend to cluster in high crime rates for burglary [76,77,78]; hence would-be offenders prone to travel across the boundaries of MSOAs in their vicinities where have higher historical burglary rate. It could also be reflected in GWR model (Fig. 7) with the fitness value range from 0.52 to 0.81 signifying fitted and expected predictions.

5. Conclusions

This paper explored four types of acquisitive crimes via combined framework of crime opportunities, social disorganization and broken windows theories, using incorporated datasets derived from both open-sourced data and traditional data. Negative binominal and geographical weighted regressions are performed to examine the hypotheses against integrated framework. Significant relationships were demonstrated between corresponding predictors and property-related crime rates; Twitter and Foursquare were considered as reliable sources for crime analysis. The incorporated data model for this paper had complimented the current research focuses, which were conducted in London context, on applying social media data (i.e., Twitter and Foursquare) to measure human mobility incurred violent crimes [43], and on utilising only the geolocational information whilst dismissal large chunk of the informative tweets' texts [45] referring to criminological theories.

Applying the NBR and GWR models, which were better tailored for the selected datasets, helped to reach at some reversed findings from empirical studies in that, higher level of ethnic heterogeneity were found significantly negative with bicycle theft, burglary and theft from the person, totally different from traditional recognition of racial disparities; so did educational indicator, i.e. the frequency of visiting College and Universities, was found to be significantly positive with all instrumental crimes, especially considering campus (crime generator) and students (weak awareness of guardianship)' characteristics contributing to crime occurrences. Spatially, there was significant spatial autocorrelation spotted at MSOA level rather than at Borough level, implying the existence of local discrepancies and necessity in adding geographical element into regression models, hence GWR model as complement to the NBR model. Such methodological models were optimal and appropriate for the research question and datasets, embracing the data's crowdsourcing features, the localised spatial dependency to the utmost extent, and the evidence-based policy making value from the research findings in London by varied local situations and be indicative in:

(1) Explanatory variables for crime opportunity are significantly associate with bicycle theft, burglary, theft from the person and vehicle crimes, providing supportive evidence on integrating social disorganization theory, broken window theory and crime opportunity [79–81].
(2) Social media data could be a reliable source complementary to traditional cross-sectional surveys, according to the results of both NBR and GWR models. In another word, through supervising the real-time streams of relevant tweets data, policy makers might be able to work out instant and appropriate decisions on crime prevention and further improve policing efficiency. Furthermore, local variations in geographically weighted regression model indicated that policies or measures for property-related crimes prevention should consider varied localised predictors. For example, polices in central London could make the patrolling schedule informed by relevant tweets on incivilities to reduce bicycle thefts; whilst patrolling polices in other MSOAs might focus more only on moments when cycling flows increase, to avoid inefficient surveillance or waste of resources.
(3) College and University being tagged as frequently visited venues were found significantly related to all types of crime, indicating further enhancement on police patrolling around those facilities might exert crime-deterrent effect on potential offenders. It is also suggested that students' awareness knowledge on acquisitive crimes, guarding collaboration possibilities with local police and college staffs should be promoted. Besides, regular liaison meetings organized by local managers could be encouraged to increase the familiarity and interactions among occupants, hence, to enhance the informal social control for crime reduction purposes.
(4) Significant negative relationship had been found between ethnic diversity and crime incidences, which counterintuitively upset the traditional stereotypes about racial heterogeneity, as it is often regarded as a sign of disadvantaged social structure. This may suggest future research to explore underlying mechanisms.

Future work

There are still some limitations in this piece of work and future work could be improved towards further steps: firstly, there were 25 days during the period 01/02/2017 to 31/01/2018 failed to scrape Twitter data due to Internet problems or blackout, making the dataset incomplete to some extent. Although it won't affect the results significant due to the nature of random sampling technique applied (among 1,836,747 tweets) and the limited proportion of missing days (25 out of 365), it is still expected to conduct more concise analysis if more consistent datasets could be accessed. secondly, dissimilar the real-time tweets and foursquare data which were updated regularly, cross-sectional datasets were out of date and compiled before 2015, thus produce temporal misalignment between dependent variables and independent variables. So future collections of simultaneous datasets from same period could be encouraged if applicable; last but not the least, this study only explores the spatial pattern of crimes without considering their temporal characteristics. For example, burglary tends to occur during holiday season when people are away from their home for holiday; theft from person normally occur in daytime around crowded public transport stations and shopping malls. A future revised spatio-temporal model integrating both temporal and spatial elements into the GWR model will be more concrete for localised crime prevention strategies and measures.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

1. In order to identify the tweets in broken window theory, the relevant parts are identified through the iterative workflow in Fig. 8 and Table 7.
2. The process of building classifiers

(1) Constructing the raw training datasets and testing datasets

The stratified sampling is firstly performed to select 10,000 tweets randomly across 983 MSOAs in order to construct a basic applicable classifier. As shown in Fig. 8, a coding frame for manually annotating the sample is compiled, based on Quinton and Tuffin [82]'s identification of visible incivilities on broken windows theory from six different communities. Further referring to Burnap and Williams [15]'s method, a computationally crowdsourcing human platform is used to label the 10,000 tweets with either "YES", "NO" or "NOT SURE", based on the aforementioned coding frame. In practice, each tweet is annotated by 4 differently online workers and 158 annotators to complete the task,based on the majority vote of trusted workers for each tweet [83,15], regardless of those with score below 75% are eliminated [84].

As a result, there are only 303 tweets annotated as YES accounting for 3.03%, which is far away to the ideal training dataset towards building up a suitable training model. Consequently, random oversampling is performed to produce the suitable training and testing datasets, according to the principles for coping with imbalanced datasets [85,86]. In brief, the number of "broken windows" tweets are increased in the sample by duplicating them randomly and the proportion between BW posts and non-BW tweets are maintained intentionally around 1:3 in this study.

(2) Cleaning sample and extracting the features

Each collected tweet's text was cleaned by commonly sequent process including vectorizating it into a list of individual words (so-called tokens), removing all non-alphanumeric characters, punctuations and stop words (i.e. a, in, the), transforming all upper letters into the lower, deleting extra whitespace and stemming all tokens (e.g. connecting, connection, connected ➔ connect) in final. Afterwards, words-of-bag (WoB) approach was used to extract the features from clean texts, namely, calculating the frequency of co-occurrence between words. Usually, n-gram is used to extract features by defining the length of them (unigrams means single token and bigram means two tokens), as shown in the Fig. 9. According to Fürnkranz's [87] study on the effect of n-gram on

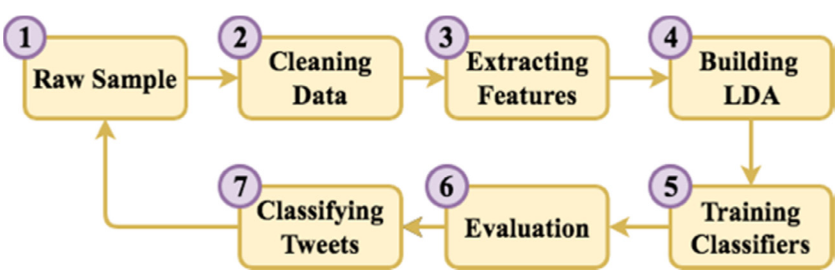

Fig. 8. Iterative workflow.

Table 7
Relevant phenomena used to annotate tweets.

Uncivil use of space	Trash	Housing Issues	Public places,Buildings and Vehicles	Other
drinking, gambling or urination on the streets	cigarettes butts, beer or liquors on the streets	bed bugs, mice infestation or pest infestation	broken windows on the buildings	infrastructure disrepair
insulting graffiti	illegal dumping	insufficient maintenance	abandoned, burned or boarded (public) buildings	strong odours
illegal parking	improper storage of trash (barrels)	any compliant residential (e.g. unsatisfactory living conditions)	abandoned cars or bicycles	noisy environment
tattooists wandering on the street	rodent activity		the broken surface of buildings	homeless people
prostitute on the streets	rats activity			gangs
any kinds of vandalism				

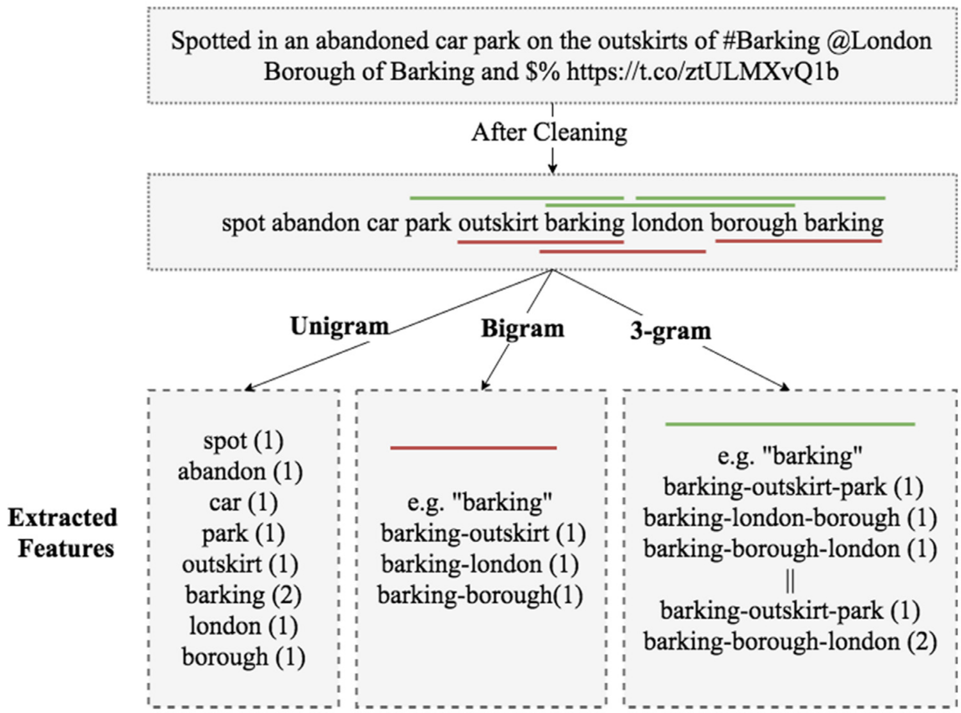

Fig. 9. Example: Extracted features from tweets by using unigram, bigram and 3-gram respectively.

text classification, the improvement reaches the peak when additionally using bigrams or 3-grams as features; majority of visible incivilities mentioned in tweets cannot be clear if only use one vocabulary. For example, it is difficult to identify whether "graffiti" refer to physical incivilities or a compliment, while tweets respectively containing "insulting graffiti" and "a graffiti artist" could be labelled as YES and NO correspondingly. Besides, Burnap and Williams [15] clusterd each tweet into grams from one to five in length for text classification and it performs well. Therefore, the final length of words sequence ranges from one to five.

The Matrix of the frequency for co-occurrence between vocabularies could be used as the input variables for Latent Dirichlet Allocation (LDA) model [88]. Two different topics (BW tweets tagged as YES and non-BW tweets tagged as NO or NOT SURE) are designated to calculate the probabilities of each tweet's belongingness.

(3) Building the classifiers then classifying the rest of tweets

Features extracted from the sample's texts and corresponding probabilities of belongingness were combined, 2/3 among which were randomly divided into training dataset with all the rest labelled as testing dataset. SVM (support vector machine) and RF (random forest decision tree) classifiers were trained and evaluated by both datasets. As a result, classifier with higher accuracy was chosen to classify the rest tweets. The primary classifier based on 10,000 tweets was iteratively informed by the results of classification for each MOSA. Namely, tweets newly tagged as YES and NO are added by the appropriation of 2:1 into the sample, which is iteratively processed in previous steps, then been log transformed for conversion.

Compliance with ethical standards

- Funding: The study was not supported by any funding.
- Conflict of Interest: The authors declare that they have no conflicts of interest.
- Ethical Approval: This article does not contain any studies involving animals performed by any of the authors.
- Informed Consent: This article does not contain any studies involving human participants performed by any of the authors.

References

[1] R.L. Matsueda, Chapter 8: Rational Choice Research in Criminology: A Multi-level Framework, Commun. Cohes. (2013) 385–458 https://faculty.washington.edu/matsueda/Papers/Rational%20Choice%20Final.pdf.
[2] D. Cornish, R. Clarke, in: The Reasoning Criminal: Rational Choice Perspectives on Offending New York, Springer-Verlag, 1986, pp. 20–45.
[3] L.E. Cohen, M. Felson, Social change and crime rate trends: a routine activity approach, Am. Sociol. Rev. 44 (4) (1979) 588–608.
[4] P.L. Brantingham, P.J. Brantingham, Environment, routine, and situation: Toward a pattern theory of crime, in: Routine Activity and Rational Choice". Advances in Criminological Theory, Transaction Publishers, 1993, pp. 259–294.
[5] C.P. Haberman, E.T. Sorg, J.H. Ratcliffe, The seasons they are a Changin': testing for seasonal effects of potentially criminogenic places on street robbery, J. Res. Crime Delinq. 55 (3) (2018) 425–459.
[6] T.A. Loughran, R. Paternoster, A. Chalfin, T. Wilson, Can rational choice be considered a general theory of crime? Evidence from individual-level panel data, Am. Soc. Criminol. 54 (1) (2016) 86–112.
[7] R.L. Matsueda, D.A. Kreager, D. Huizinga, Deterring delinquents: a rational choice model of theft and violence, Am. Sociol. Rev. 71 (1) (2006) 95–122.
[8] C.R. Shaw, H. McKay, Juvenile Delinquency and Urban Areas, University of Chicago Press, 1969.
[9] J.Q. Wilson, G.L. Kelling, Making neighborhoods safe, Atl. Mon. 263 (1989) 46–52.
[10] R.J. Sampson, S.W Raudenbush, Disorder in urban neighborhoods-does it lead to crime? U.S Dep. Justice Off. Justice Progr. National Institute of Justice 186049 (2001) 1–5.
[11] T.D. Miethe, R.F. Meier, Opportunity, choice, and criminal victimization: a test of a theoretical model, J. Res. Crime Delinq. 27 (3) (1990) 243–266.
[12] D. Weisburd, E.R. Groff, S.M. Yang, The importance of both opportunityand social disorganization theory in a future research agenda to advance criminological theory and crime prevention at places, J. Res. Crime Delinq. 51 (4) (2014) 499–508.
[13] Xinyu. Chen, Youngwoon. Cho, Suk young Jang, Crime prediction using twitter sentiment and weather, in: Proceedings of the IEEE Systems and Information Engineering Design Symposium, 2015, pp. 63–68.
[14] X.G. Wang, S. Matthew, E.D. Brown, Automatic crime prediction using events extracted from twitter posts, in: In Social Computing, Behavioral-Cultural Modelling and Prediction, Springer, 2014, pp. 1–8.
[15] P. Burnap, M.L. Williams, Cyberhate on social media in the aftermath of Woolwich: a case study in computational criminology and big data, Br. J. Criminol. 56 (2015) 211–238.
[16] J. Bendler, T. Brandt, S. Wagner, D Neumann, Investigating crime-to-twitter relationships in urban environments—facilitating a virtual neighborhood watch, in: Proceedings of the European Conference on Information Systems (ECIS), Tel A viv, 2014.
[17] Matthew.S. Gerber, Predicting crime using Twitter and kernel density estimation, Decision Support Systems 61 (2014) 115–125.
[18] W. Bernasco, R. Block, Robberies in Chicago: a block-level analysis of the influence of crime generators, crime attractors, and offender anchor points, J. Res. Crime Delinq. 48 (1) (2011) 33–57.

[19] R.J. Sampson, S.W. Raudenbush, . Earls, Neighborhoods and violent crime: a multilevel study of collective efficacy, Science 277 (5328) (1997) 918–924.
[20] W. Steenbeek, JR. Hipp, A longitudinal test of social disorganization theory: feedback effects: among cohesion, social, social control, and disorder, Criminology 49 (3) (2011) 833–871.
[21] W.G. Skogan, Disorder and crime. Edited by David P. Farrington and Brandon C. Welsh, Oxford Handbook of Crime Prevention (2012) 1–17.
[22] R. Macmillan, A. Nierobisz, S. Welsh, Experiencing the Streets: harassment and perceptions of safety among women, J. Res. Crime Delinq. 37 (2000) 306–322.
[23] C.E. Ross, J. Mirowsky, Disorder and decay: the concept and measurement of perceived neighborhood disorder, Urban Aff. Rev. 34 (1999) 412–432.
[24] A.E. Liska, in: Perspectives on Deviance, Prentice- Hall, 1978, pp. 20–24.
[25] J.Q. Wilson, L.K. George, The police and neighbor- hood safety: broken windows, Atl. Mon. 127 (1982) 29–38.
[26] L.M. Berger, D. Borenstein, An agent-based simulation of car theft: further evidence of the rational choice theory of crime, Econ. Anal. Law Rev. 4 (1) (2013) 103–119.
[27] C.P. Rosbough, Property crime at the Atlanta international airport: an examination of the rational choice theory and the 9/11 intervention, J. Appl. Secur. Res. 7 (3) (2012) 354–374.
[28] W. Evans, O. Emily, COPS and crime, J. Public Econ. 91 (2007) 181–201.
[29] M.J. Lin, More police, less crime: evidence from U.S. state data, Int. Rev. Law Econ. 29 (2009) 73–80.
[30] D. Weisburd, T. Cody, H. Joshua, E. John, Is problem oriented policing effective in reducing crime and disorder? Criminol. Pub. Policy 9 (2010) 139–172.
[31] K.R. Williams, R. Hawkins, Perceptual research on general deterrence: A critical review, Law Soc. Rev. 20 (1986) 545–572.
[32] R.J. Sampson, J.H. Laub, Turning points and the future of life-course criminology: reflections on the 1986 criminal careers report, J. Res. Crime Delinq. 53 (3) (2016) 321–335.
[33] J. Schwartz, D. Steffensmeier, H. Zhong, J. Ackerman, Trends in the gender gap in violence: reevaluating NCVS and other evidence, Criminology 47 (2009) 401–426.
[34] D. Larsson, Exposure to property crime as a consequence of poverty, Scand. Stud. Criminol. Crime Prev. 7 (1) (2006) 45–60.
[35] E.R. Louderback, S.S. Roy, Investigating social disorganization and routine activity theories and testing the effectiveness of neighborhood crime watching programs: case study of Miami-Dade County, 2007–15, Br. J. Criminol. 58 (4) (2018) 968–992.
[36] A.P. Askey, R. Taylor, E. Groff, A. Fingerhut, Fast Food Restaurants and Convenience Stores: Using Sales Volume to Explain Crime Patterns in Seattle, Crime & Delinquency 64 (14) (2018) 1836–1857 First published in 2017.
[37] E. Demeau, G. Parent, Impacts of crime attractors and generators on criminality in montreal, Can. J. Criminol. Crim. Justice 60 (3) (2018) 387–412.
[38] Roman, C.G. Schools as generators of crime: routine activities and the sociology of place. Doctoral thesis, American University, Washington, DC. 2002.
[39] B.C. Welsh, A.A. Braga, G.J.N Bruinsma, Reimagining broken windows: from theory to policy, J. Res. Crime Delinq. 52 (4) (2015) 447–463.
[40] R.J. Sampson, W.B. Groves, Community structure and crime: testing social disorganization theory, Am. J. Sociol. 94 (1989) 774–802.
[41] W.G. Skogan, in: Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods, Free Press, 1990, pp. 30–40.
[42] D. O'Brien, J.S. Robert, "Public and private spheres of neighborhood disorder: assessing pathways to violence using large-scale digital records, J. Res. Crime Delinq. 52 (2015) 486–510.
[43] N. Malleson, M. Andresen, The Impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns, Cartogr. Geogr. Inf. Sci. 42 (2015) 112–121.
[44] T. Althoff, P. Jindal, J. Leskovec, Online Actions with Offline Impact: How Online Social Networks Influence Online and Offline User Behavior, in: Maarten de Rijke (Ed.), Proceedings of the WSDM '17 Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, Association for Computing Machinery, New York, United States, 2017, pp. 537–546. https://dl.acm.org/doi/10.1145/3018661.3018672.
[45] M.L. Williams, P. Burnap, L. Sloan, Crime sensing with big data: the affordances and limitations of using open-source communications to estimate crime patterns, Br. J. Criminol. 57 (2017) 320–340.
[46] D. Quercia, U.P. Fabra, Mining urban deprivation from foursquare: implicit crowdsourcing of city and use, IEEE Pervasive Comput. 13 (2) (2014) 30–36.
[47] C. Kadar, J. Iria, P.I. Cvijikj, Exploring foursqaure-derived features for crime prediction in New York City, in: Proceedings of the The 5th International Workshop on Urban Computing (UrbComp 2016), San Francisco, CA, USA, ACM Press, 2016, pp. 1–8.
[48] M.A. Boni, M.S Gerber, Man, AND cybernetics. predicting crime with routine activity patterns inferred from social media, in: Proceedings of the IEEE International Conference on Systems, Budapest, Hungary, 2016, doi:10.1109/SMC.2016.7844410.
[49] M. Draca, M. Stephen, W. Robert, Panic on the streets of London: Police, crime and the July 2005 terror attacks, Am. Econ. Rev. 101 (2011) 2157–2181.
[50] L.J. Krivo, R.D. Peterson, D.C. Kuhl, Segregation, racial structure, and neighborhood violent crime, Am. J. Psychol. 114 (6) (2009) 1765–1802.
[51] S. Openshaw, P.J. Taylor, The modifiable areal unit problem, in: (Quantitative Geography: A British View), Geo Books, 1981, pp. 23–40.
[52] A. Tumasjan, T. Sprenger, P.G. Sandner, I.M Welpe, Predicting elections with twitter: what 140 characters reveal about political sentiment, in: Proceedings of 4th ICWSM, AAAI Press, 2010, pp. 178–185.
[53] L. Mitchell, M.R. Frank, K.D. Harris, P.S. Dodds, C.M. Danforth, The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place, PLoS One 8 (5) (2013) 1–15 e64417.
[54] Q. Le, T. Mikolov, Distributed representations of sentences and documents, in: Proceedings of the 31st International Conference on Machine Learning, Beijing, China, ICML, 2014, pp. 1188–1196.
[55] M. Neiss, Does public transit affect crime? The addition of a bus line in Cleveland, J. Econ. Polit. 22 (1) (2015) 59–74.
[56] T. Ravalin, T. Tevis, Social disorganization theory and crime rates on California community college campuses, Commun. Coll. J. Res. Pract. 41 (1) (2017) 27–41.
[57] A Alesina, A. Devleeschauwer, W. Easterly, S. Kurlat, R. Wacziarg, Fractionalization, J. Econ. Growth 8 (2003) 155–194.
[58] Y. Irvin-Erickson, N.L. Vigne, A spatio-temporal analysis of crime at Washington, DC metro rail: stations' crime-generating and crime-attracting characteristics as transportation nodes and places, Crime Science 14 (2015) 4–14.
[59] R. Block, D. Sean, The environs of rapid transit stations: a focus for street crime or just another risky place? in: Preventing Mass Transit Crime, Criminal Justice Press, 1994, pp. 237–257. 1996.
[60] K. Ihlanfeldt, Rail transit and neighborhood crime: the case of Atlanta, Georgia, South. Econ. J. 70 (2) (2003) 273–294.
[61] A.L. Nielsen, M.T. Lee, R. Martinez, Integrating race, place and motive in social disorganization theory: lessons from a comparison of black and Latino homicide types in two immigrant destination cities, Criminology 43 (2005) 837–872.
[62] A.T. Murray, I. McGuffog, J.S. Western, P. Mullins, Exploratory spatial data analysis techniques for examining urban crime, Br. J. Criminol. 41 (2001) 309–329.
[63] D. Weisburd, W. Bernasco, G. Bruinsma, in: Putting Crime in its Place: Units of Analysis in Geographic Crimi- Nology, Springer, 2009, pp. 65–80.
[64] L. Anselin, Local indicators of spatial association—LISA, Geogr. Anal. 27 (1995) 93–115.
[65] A. Getis, J.K. Ord, Local spatial statistics: An Overview, Spatial analysis: modelling in a GIS environment (1996) 261–277.
[66] C. Brunsdon, A.S. Fotheringham, M. Charlton, Geographically weighted regression: a method for exploring spatial non-stationarity, Geogr. Anal. 28 (1996) 281–298.
[67] H. Zhang, W. Song, Addressing issues of spatial spillover effects and non-stationarity in analysis of residential burglary crime, GeoJournal 79 (2014) 89–102.
[68] A. Alesina, E. La Ferrara, Who trusts others? J. Public Econ. 85 (2002) 207–234.
[69] A. Alesina, E. Zhuravskaya, Segregation and the quality of government in a cross section of countries, Am. *Econ. Rev.* 101 (2011) 1872–1911.
[70] A.T. Leigh, Inequality and ethnic heterogeneity, Econ. Rec. 82 (2006) 268–280.
[71] S.A. Churchill, E. Laryea, Crime and Ethnic Diversity: Cross-country Evidence, Crime & Delinquency 65 (2) (2019) 239–269 First published in 2017.
[72] Transport for London (TfL), and Reports, Strategic cycling analysis, TfL Corporate Publications (2017) https://content.tfl.gov.uk/strategic-cycling-analysis.pdf.
[73] R.M. Tewksbury, E. Ehrhardt, College Students' lifestyle and self-protective behaviors: futher considerations of the guardianship concept in routine activity theory, Crim. Justice Behav. 30 (3) (2003) 302–327.
[74] E.R. Groff, B. Lockwood, Criminogenic facilities and crime across street segments in philadelphia: uncovering evidence about the spatial extent of facility influence, J. Res. Crime Delinq. 5 (3) (2014) 277–314.
[75] D. Mcilhatton, W. Mcgreal, P. Taltavul de La Paz, A. Adair, Impact of crime on spatial analysis of house prices: evidence from a UK city, Int. J. Hous. Mark. Anal. 9 (4) (2016) 627–647.
[76] C. Lum, Violence, drug markets and racial composition: Challenges stereotypes through spatial analysis, Urban Stud. 48 (13) (2011) 2715–2732.
[77] A. Sagovsky, S.D. Johnson, When does near repeat burglary victimization occur? Aust. N. Z. J. Criminol. 40 (1) (2007) 1–26.
[78] M. Townsley, R. Homel, J. Chaseling, Infectious burglaries. A test of the near repeat hypothesis, Br. J. Criminol. 43 (3) (2003) 615–633.
[79] M.A. Andresen, A spatial analysis of crime in vancouver, British Columbia: a synthesis of social disorganization and routine activity theory, Can. Geogr. Le G´eogr. Can. 50 (4) (2006) 487–502.
[80] K.J. Rice, R.C.S. William, "Socioecological models of auto-motive theft: integrating routine activity and social disorganization approaches, J. Res. Crime Delinq. 39 (2002) 304–336.
[81] D.L. Weisburd, G. Elizabeth, S.M. Yang, in: The Criminology of Place: Street Segments and Our Understanding of the Crime Problem, Oxford University Press, New York, 2012, pp. 1–45. https://www.london.gov.uk/business-and-economy-publications/working-paper-72-economic-analysis-londons-housing-market.
[82] P. Quinton, R. Tuffin, Neighborhood change: the impact of the national reassurance policing programme, Polic.A J. Policy Pract. 1 (2007) 149–160.
[83] V.Z. Kolhatkar, H.G. Hirst, Interpreting anaphoric shell nouns using antecedents of cataphoric shell nouns as training data, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Washington,, 2013, pp. 300–310.
[84] M.B. Thelwall, G. Paltogou, D. Cai, A. Kappas, Sentiment strength detection in short informal text, J. Am. Soc. Inf. Sci. Technol. 61 (2010) 2544–2558.
[85] H.M. Nguyen, E.W. Cooper, K. Kamei, Borderline over-sampling for imbalanced data classification, Int. J. Knowl. Eng. Soft Data Paradigms. 3 (1) (2011) 4–21.
[86] A.Dal. Pozzolo, O. Caelen, Y.A.L. Borgne, S. Waterschoot, G. Bontempi, Learned lessons in credit card fraud detection from a practitioner perspective, Expert Syst. Appl. 41 (10) (2014) 4915–4928.
[87] J. Fürnkranz, A study using n-gram features for text categorization, Austrian Research Institute for Artifical Intelligence 3 (1998) 1–10.
[88] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.