

# CRIME SENSING WITH BIG DATA: THE AFFORDANCES AND LIMITATIONS OF USING OPEN-SOURCE COMMUNICATIONS TO ESTIMATE CRIME PATTERNS

MATTHEW L. WILLIAMS\*, PETE BURNAP and LUKE SLOAN

*This paper critically examines the affordances and limitations of big data for the study of crime and disorder. We hypothesize that disorder-related posts on Twitter are associated with actual police crime rates. Our results provide evidence that naturally occurring social media data may provide an alternative information source on the crime problem. This paper adds to the emerging field of computational criminology and big data in four ways: (1) it estimates the utility of social media data to explain variance in offline crime patterns; (2) it provides the first evidence of the estimation of offline crime patterns using a measure of broken windows found in the textual content of social media communications; (3) it tests if the bias present in offline perceptions of disorder is present in online communications; and (4) it takes the results of experiments to critically engage with debates on big data and crime prediction.*

**Key words:** social media, computational criminology, predictive policing, big data, broken windows, open-source communications

## *Introduction*

This paper reports on a methodological experiment with ‘big data’ in the field of criminology. In particular, it provides a *data-driven critical examination* of the affordances and limitations of open-source communications gathered from social media interactions for the study of crime and disorder. The experiment conducted was exploratory in nature, and utilized nascent ‘computational criminological’ methods (Williams and Burnap 2015) to ethically harvest, transform, link and analyse ‘big social data’ to address the classic problem of crime pattern estimation (Braga *et al.* 2012). The results presented form a preliminary basis for the critical discussion of these ‘new forms of data’ and for subsequent confirmatory analysis to be conducted. The aim of the experiment was to build big data statistical models that develop previous predictive work using social media. For example, Tumasjan *et al.* (2010) measured Twitter sentiment in relation to candidates in the German general election concluding that this source of data was as accurate at predicting voting patterns as polls. Asur and Huberman (2010) correlated frequency of posts and sentiment related to movies on Twitter with their revenue, claiming that this method of prediction was more accurate than the Hollywood Stock Market. Sakaki *et al.* (2010) found that the analysis of Twitter data produced estimates

\*Matthew L. Williams, Reader, Computational Criminology, School of Social Sciences, Cardiff University, King Edward VII Avenue, Cardiff CF10 3WT and Director, Social Data Science Lab, Data Innovation Research Institute, Cardiff University, Cardiff, UK; [williamsm7@cf.ac.uk](mailto:williamsm7@cf.ac.uk); Pete Burnap, Lecturer, Computer Science, School of Computer Science and Informatics and Director, Social Data Science Lab, Data Innovation Research Institute, Cardiff University, Cardiff, UK. Luke Sloan, Senior Lecturer, Quantitative Methods, School of Social Sciences and Deputy Director, Social Data Science Lab, Data Innovation Research Institute, Cardiff University, Cardiff, UK.

of the epicentres of earthquakes more accurately than conventional geological sensor methods. These studies illustrate how social media generates ‘naturally occurring’ socially relevant data that can be used to complement and augment conventional curated data to estimate the occurrence of offline phenomena. In our experiment, we conduct an ecological analysis of crime in London using Twitter data as a predictor to test the hypothesis that crime- and disorder-related tweets are associated with actual police crime rates. Our results provide tentative evidence that statistical models based on social media data may provide an alternative source of information on the crime pattern estimation problem. This paper adds to the evidence base and debate in the emerging field of computational criminology in four ways: (1) it estimates the utility of social media data to explain variance in offline crime patterns and compares results with conventional indicators (census variables); (2) it provides the first evidence of the estimation offline crime patterns using a measure of broken windows found in the textual content of social media communications; (3) it specifically tests if the bias present in offline perceptions and reports of crime and disorder (found between low- and high-crime areas) is present in social media; and (4) it uses the results of these experiments to critically engage with debates on big data and crime estimation.

#### *Social media communications as source of data for criminology*

The majority of individuals aged below 20 in the Western world were ‘born digital’<sup>1</sup> and will not recall a time without access to the Internet. Combined with the migration of the ‘born analogue’ generation onto the Internet, fuelled by the rise of social media, we have seen the exponential growth of online spaces for the mass sharing of opinions and sentiments. The digital revolution is generating high-volume data through multiple forms of online behaviour. The global adoption of social media over the past half a decade has seen ‘digital publics’ expand to an unprecedented level. Estimates put social media membership at approximately 2.5 billion non-unique users, with Facebook, Google+ and Twitter accounting for over half of these. These online populations produce hundreds of petabytes of information, with Facebook users alone uploading 500 terabytes of data daily. No study of contemporary society can ignore this dimension of social life. The potential value added by social media data for criminological research is that it is user-generated in real-time in voluminous amounts, and as such it can provide insight into the behaviour of specific populations on the move. This is in contrast to the necessarily retrospective snapshots provided by conventional methods such as household surveys and officially recorded data. New forms of online social data, handled by computational methods, allow criminologists to gain meaningful insights into contemporary social processes at unprecedented scale and speed, but how we marshal these new forms of data presents a key challenge (see [Williams \*et al.\* 2013](#), [Williams and Burnap 2015](#)).

In our exploratory study with big data, we make the assumption that each Twitter user is a *sensor* of offline phenomena. In the vein of [Raudenbush and Sampson \(1999\)](#), we consider these sensors, or nodes for systematic social observation, as part of a wide sensor-net covering *ecological zones* (in our case London boroughs). These sensors observe natural phenomenon—the sights, sounds and feel of the streets ([Abbott 1997](#)). As in

<sup>1</sup>Though of course this general claim is mediated by social factors such as poverty, disadvantage and spatial location (see [Boyd 2014](#)).

the case of ‘broken windows’ (Wilson and Kelling 1982), these can include minor public incivilities—drinking in the street, graffiti, litter—that serve as *signals* of the unwillingness of residents to confront strangers, intervene in a crime or call the police; cues that entice potential predators (Skogan 1990: 75). Sensors can publish information about local social and physical disorder in four ways: as victims; as first-hand witnesses; as second-hand observers (e.g. via media reports or the spread of rumour) and as perpetrators. We consider these four modes of Twitter publishing as *signatures* of crime and disorder. These social-actors-as-disorder-sensors have various characteristics. Some are activated (i.e. publish tweets) based on specific signs, while others are not (based on variation in perceptions of disorder).<sup>2</sup> Data from these sensors also includes temporal and spatial information. Sensors are not always switched ‘on’, as they may be offline, working, sleeping etc. They may also act in ways that make the data difficult to interpret and validate (e.g. using sarcasm and spreading rumours). This means they produce data that are noisier than curated data. However, the number of sensors is prodigious; over 500 million tweets are broadcast daily from over 500 million accounts; 15+ million of these emanate from the United Kingdom (Library of Congress 2013; Smith 2012).

### *The Challenges of Big Social Data for Criminology: The 6 Vs*

Criminology faces the challenge of how increasingly ubiquitous digital devices and the data they produce are reassembling its research methods apparatus. The exponential growth of social media uptake and the availability of vast amounts of information from these networks have created fundamental methodological and technical challenges. However, aside from recent papers by Chan and Bennet-Moses (2015) and Williams and Burnap (2015), big ‘social’ data have received little attention amongst criminologists, leaving the question of how as a discipline we respond to it largely unexplored. The challenges (and affordances) can be summarized as the 6 Vs: volume, variety, velocity, veracity, virtue and value.

‘Volume’ refers to the vast amount of socially relevant information uploaded on computer networks globally every second. Ninety per cent of the world’s data were created in the two years prior to 2013 (BIS 2013). This is partly due to the global adoption of social media over the past half a decade. Of the online social interactions produced on these networks, a sizable portion is relevant to criminology. For example, Williams and Burnap (2015) have examined the spread of cyberhate on Twitter following the Woolwich terror attack. A comparison with curated and administrative sources on crime reveals the scale of these new data. The most recent Crime Survey for England and Wales (CSEW, 2012–13) data file measures 113.4 megabytes in size. Since its inception in 1982 all CSEW data would not amount to more than 2 gigabytes. In terms of administrative data, the Police National Computer contains circa 9.2 million nominal records (NPIA 2009). The whole UK Data Archive currently holds between 2.2 and 15 terabytes of data. These sizes are dwarfed by the volume of social media data being produced daily that are relevant to criminology.

‘Velocity’ refers to the speed at which these new forms of data are generated and propagated. Recent social unrest illustrates how social media information can spread over large distances in very short periods of time. For example, the HMIC (2011) report

<sup>2</sup>It is important to note that sensors filter information; it is not ‘objective’ per se but instead mediated by social relations. See section Discussion for an elaboration of this point.

*Policing Public Order* highlighted how the disorder in 2011 had taken on a new dimension, which involved the use of social media. In particular, its use was implicated in the UK Uncut and university tuition fees protests in London in late 2011. At the extreme end of the spectrum, social media use was also associated with the Tunisian and Egyptian Revolutions (Lotan *et al.* 2011; Choudhary *et al.* 2012).

‘Variety’ relates to the heterogeneous nature of these data, with users able to upload text, images, audio and video. This multimodal mixed dataset can be harnessed by researchers. However, unlike qualitative and quantitative data that are often labelled, coded and structured within matrices and ordered transcripts, big ‘social’ data are messy, noisy and unstructured.

‘Veracity’ relates to the quality, authenticity and accuracy of these messy data. Triangulating social media communications with more conventional sources, such as curated data, can mitigate these problems. Instead of social media acting as a surrogate for established sources, it should instead augment them, adding a hitherto unrealized longitudinal extensive dimension to existing research strategies and designs. For the first time, this allows criminologists to study social processes as they unfold in real time at the level of populations while drawing upon gold-standard static qualitative and quantitative metrics to inform interpretations. Furthermore, Williams *et al.* (2013) show that the near ubiquitous adoption of smartphone technology and social media amongst groups that are underrepresented in official survey collection exercises means these new data sources may provide better coverage of such populations.

‘Virtue’ relates to the ethics of using this new form of data in social research. A recent survey found that 74 per cent of social media users knew that when accepting Terms of Service they were giving permission for their information to be accessed by third parties. Eighty-two per cent of respondents were ‘not at all concerned’ or only ‘slightly concerned’ about university researchers using their social media information (however, this dropped to 56 per cent for police access) (Williams 2015). We may argue therefore that researchers in this field must accept that consent has been provided, as long as researchers adhere to basic principles of social science ethics while ensuring results are presented at an aggregate level. Additional individual-level consent should be sought if researchers wish to directly quote online communications.

Finally, ‘value’ links the preceding five Vs—only when the volume, velocity and variety of these data can be computationally handled, and the veracity and virtue established, can criminologists begin to marshal them and extract meaningful information. However, to date, few academic criminological studies have collected and analysed social media data. In order to make sense of this rich material, Burnap *et al.* (2014a) advocate the establishment of interdisciplinary teams of computer and social scientists using parallel computing infrastructure. Dubbed ‘computational criminology’, this interdisciplinary methodology has its roots in computational social science (Lazer *et al.* 2009). In their pioneering article in *Science*, Lazer *et al.* argue that corporate giants such as Facebook, Google and Twitter have been using social data with advanced computing to mine and interpret it for half a decade. Until recently, academic social scientists have been left in an ‘empirical crisis’, lacking the access, infrastructure and skills to marshal these data (Savage and Burrows 2007). In this study, computer scientists and criminologists collaborated to address the 6 Vs for the purposes of offline crime estimation using Twitter data.

*Big Data and Crime Estimation*

Recent studies have attempted to integrate social media data into statistical models for crime estimation. [Bendler et al. \(2014\)](#) examined the relationship between mobile populations as recorded by Twitter's geotagging functionality and the co-location of different crime types. They found the *absence* of tweets was predictive of assaults, theft, and disturbing the peace. Similarly, [Malleon and Andresen \(2015\)](#) used Twitter data to measure mobile populations at risk from violent crime in Leeds. They used a variety of geographic analysis methods to model crime risk using tweets as signatures for mobile populations, noting that conventional estimation methods rely on outdated static data on residential populations (such as the census). They found alternative violent crime hotspots outside of Leeds city centre, not identifiable with conventional crime data sources, concluding Twitter data represent mobile populations at higher spatial and temporal resolutions than sources used by police.

The key limitation to these studies is their dismissal of tweet text, instead focussing purely on geolocation data. The content of tweets may be relevant to the estimation of crime patterns, and simple geolocation data fail to relate to any possible theoretical explanation aside from routine activities. In order to address the utility of tweet text in estimating crime patterns, [Gerber \(2014\)](#) used latent Dirichlet allocation (LDA)<sup>3</sup> on content. Tweet text was shown to improve upon models containing conventional non-social media crime predictors for stalking, criminal damage and gambling, but decrease performance for arson, kidnapping and intimidation. Although it is the first study to examine tweet content, Gerber's use of LDA is problematic given that it is an unsupervised method, meaning correlations between word clusters and crimes are not driven by prior theoretical insight ([Chan and Bennett-Moses 2015](#)). This resulted in correlations that appear relatively meaningless, (e.g. prostitution was correlated with the words 'studios', 'continental', 'village' and 'Ukrainian'). It is unclear how terms relate to crimes, and it is not easy to understand how such work can inform criminological theory or policing practice. An improved approach would involve the classification of tweet text based on a predetermined theoretical framework. We adopted such an approach in this study, using ideas from the 'broken windows' thesis to guide the classification of social media content that indicated forms of neighbourhood degeneration.

*Broken Windows and Big Data*

'Broken windows' is a well known theory in criminology. The most basic formulation of this theory is that visible signs of neighbourhood degeneration are causally linked to crime ([Wilson and Kelling 1982](#)). The broken windows thesis has received considerable attention over the past three and a half decades, resulting in empirical findings that largely support its core supposition (see [Skogan 2015](#); [Welsh et al. 2015](#)).<sup>4</sup> Most prominent in the thesis is the hypothesized relationship between *visible* forms of disorder, their deleterious impact upon residents and their additional effect of drawing offenders from outside of the neighbourhood. In particular, measures of physical disorder have included

<sup>3</sup>LDA is a form of topic modelling that automatically discovers underlying topics in documents containing text.

<sup>4</sup>Here, we do not cover debates regarding the contentious policing initiatives based on broken windows theorizing. There is a lack of agreement on the definition of broken windows policing, and many empirical studies focus on police discretion ([Braga and Bond 2008](#)) and not the measure of interest in this study: neighbourhood degeneration.



reports from residents of litter, graffiti and vandalism (Sampson and Raudenbush 2004) that are taken as signatures of the breakdown of the local social order (Skogan 2015). Such measures have conventionally been developed via community-based surveys, interviews and neighbourhood audits, but these instruments capture data in a cross-sectional fashion, often precluding longitudinal analysis at smaller temporal scales. Recently, big administrative data that exhibit longitudinal features have been mined to generate measures of broken windows. Building on the *ecometrics approach* developed by Raudenbush and Sampson (1999), O'Brien and Sampson (2015) and O'Brien *et al.* (2015) constructed and validated a measure of physical disorder using a large database from Boston's constituent relationship management (CRM) system (311 hotline) used by local residents to request city services, many of which reference physical incivilities (e.g. graffiti removal). This approach generated a large ( $n = 200,000+$ ) geospatially structured dataset that could be repurposed for the estimation of crime and disorder patterns using broken windows measures at very small temporal and spatial scales. Their findings revealed that (1) administrative records, collected for the purposes other than research, could be used to reliably construct measures of broken windows, and (ii) these measures were significantly associated with levels of crime and disorder. These represent the first studies of broken windows using administrative 'big data', and the authors conclude: 'Going further, there are private databases, such as Twitter, cell phone records, and Flickr photo collections that are also geocoded and might be equally informative in building innovative measures of urban social processes. These various resources could be used to develop new versions of traditionally popular measures, like we have done here, or to explore new ones that have not been previously accessible' (O'Brien *et al.* 2015: 35). This paper takes on this task by testing three hypotheses.

### *Hypotheses*

H1: Estimation models including social media variables will increase the amount of crime variance explained compared to models that include 'offline' variables alone.

Previous work on using social media and mobile phone data as predictors of offline phenomena, including crime, has shown that they increase the amount of variance explained in statistical models over models using conventional offline variables alone (Asur and Huberman 2010; Gerber 2014). This hypothesis tests whether this holds true for the estimation of crime patterns in the United Kingdom while accounting for temporal variation.

H2: Twitter mentions of 'broken windows' indicators will be positively associated with police-recorded crime rates in low-crime areas.

H3: Twitter mentions of 'broken windows' indicators will be negatively or not associated with crime rates in high-crime areas.

These hypotheses are based on previous research that finds offline discussions of neighbourhood degeneration and local crime issues in Partners and Communities Together meetings are not representative of local crime problems (e.g. Brunger 2011; Sagar and Jones 2013). This is in part due to patterns of low attendance in high-crime areas, and the non-representativeness of regular meeting attendees. This can result in (1) regular reporting of criminal and sub-criminal issues at such meetings in low-crimes areas, due to socially engaged attendees who are sensitive to degeneration, and (2) systematic

under-reporting of criminal and sub-criminal issues at such meetings in high-crime areas, due to lack of attendance because of a reduced sensitivity in residents to degeneration—the idea that degeneration has gone too far resulting in ‘lost neighbourhoods’ occupied by residents that have naturalized to their surroundings (Sampson 2012). Therefore, these hypotheses explicitly test whether the bias found in offline reports of crime and disorder is also present in Twitter communications.

### *Data and Methods*

#### *Data*

Variables were derived from three sources and were combined at the borough level for modelling: (1) a database of police-recorded crime provided by the Metropolitan Police Service; (2) the UK Census 2011; and (3) the social media network, Twitter. The Metropolitan Police Service provided circa 600,000 police-recorded crime records covering all London boroughs over a 12-month period between August 2013 and August 2014. The UK Census 2011 was accessed via the nomis web portal.<sup>5</sup> All UK tweets were collected via the Twitter streaming Application Programming Interface using the Cardiff Online Social Media Observatory (COSMOS) software platform<sup>6</sup> (Burnap *et al.* 2014a), resulting in circa 200 million tweets with location information covering a 12-month period. Borough level was selected as the unit of spatial analysis to maximize the number of geolocated tweets in the dataset<sup>7</sup> (see section Limitations for a discussion on spatial scale).

#### *Dependent measures*

##### *Police-recorded crime*

Nine crime categories were selected for modelling from the police recorded crime database and each were summed by 28 London boroughs and by month over the study window.<sup>8</sup> Estimating crime at the borough level allows for an *ecological analysis*. Raudenbush and Sampson (1999) show how observations collected at the level of ecological units (in our case London boroughs) can yield relationships with perceptions of disorder and fear of crime and crime patterns.

#### *Independent measures*

##### *Social media regressors*

Two regressors were derived from Twitter communications. *Frequency of Twitter Posts*—the 200 million geocoded tweets collected in the United Kingdom over the

<sup>5</sup><https://www.nomisweb.co.uk/census/2011>.

<sup>6</sup>See [www.socialdatalab.net/software](http://www.socialdatalab.net/software).

<sup>7</sup>Location information can be derived from a tweet object in several ways (see Sloan *et al.* 2013). The most accurate way is via a device’s GPS system (providing latitude and longitude) if the tweeter decides to include a precise location in a tweet. A less accurate way is extracting other location information from the tweet object generated by the user (e.g. profile location). This less accurate method is only capable of placing tweeters within broader administrative zones, such as London boroughs. To maximize the number of Twitter posts in our models, we opted to include geolocated Twitter content tagged using either method. This precluded a more fine-grained analysis below borough (as the less accurate method of geolocation is less reliable at lower geographic levels).

<sup>8</sup>Overlaps in geographic labelling between the Twitter and MPS datasets precluded the inclusion of all 32 London boroughs in the analysis.

12-month period were reduced to those geolocated in the 28 London boroughs over the study window ( $n = 8,417,438$ ) and were summed by borough and month. *Twitter Mentions of 'Broken Windows'*—tweets were classified as containing 'broken windows' indicators (e.g. mentions of neighbourhood degeneration) and were summed by borough and month. Our approach recognized that Twitter users act as *sensors* of their environment, much like a large distributed 'social sensor-net'. Some of these sensors may publish content about the changing condition of their neighbourhood, such as directly witnessing crime, disorder and decay. They may also sense degeneration as second-order witnesses (via news reports), as victims or as perpetrators of crime. Unlike O'Brien *et al.*'s (2015) econometric measurement approach, ours was a task of 'text classification' (van Rijsbergen 1979). This was due to the unstructured nature of Twitter communications, in contrast to the structured administrative<sup>9</sup> data used by O'Brien *et al.* The process followed established automatic text classification procedures adopted in our previous work with social media data (see Burnap and Williams 2015; Burnap *et al.* 2014b; Sloan *et al.* 2015; Williams and Burnap 2015). First, mentions pertaining to 'broken windows' were extracted by the authors from offline interviews with victims and non-victims in local neighbourhoods.<sup>10</sup> A coding frame for text extraction was informed by Quinton and Tuffin's (2007) evaluation of UK National Reassurance Policing Priorities that identified common concerns from local residents in six sites, which relate to 'broken windows' measures (alcohol and/or drug use; litter and dog fouling; criminal damage; speeding; parking and nuisance vehicles; anti-social behaviour and juvenile nuisance). O'Brien *et al.*'s (2015) recent work was also used to inform text extraction. They developed validated measures of 'broken windows' using large-scale administrative records and identified that reports of housing issues (e.g. poor maintenance), trash and graffiti held the strongest reliability. Second, to validate that the coding was related to 'broken windows' indicators, extracts were independently rated using a crowdsourcing approach involving 700 human annotators sampled from the CrowdFlower<sup>11</sup> crowdsourcing service. We required at least four human annotations per interview extract and only retained annotated text for which at least three human annotators (75%) agreed that extracts related to signatures of neighbourhood degeneration. Finally, the key terms contained within the verified classified text extracts were used to mine the Twitter dataset, resulting in a social media measure of 'broken windows'.<sup>12</sup> Figurative examples of tweet content containing 'broken windows' indicators included: 'New graffiti at the end of my street. How did they reach that high!?!'; 'Community allotment was vandalized today. Why would someone do this?'; 'More illegal dumping in Shoreditch. When will @hacknycouncil sort this out?!'; and 'RT if you think we should use discarded card receipts to identify litterers!' [Includes a photo of

<sup>9</sup>The administrative CRM data used in their study were pre-processed into a coding frame by humans receiving calls from the public.

<sup>10</sup>Interviews were extracted from the UK Data Archive.

<sup>11</sup>See <http://www.crowdfunder.com>.

<sup>12</sup>Retweets (RT) were included in the dataset as they were taken to mean endorsements, i.e. that residents would only retweet content indicating a breakdown in the local social order (excessive litter, graffiti etc.) if they shared the same perception. Retweets therefore act as an amplification mechanism. It is the convention on Twitter to produce a Modified Tweet (MT) (reproducing in part the original tweet with modified text indicating a difference of opinion/perception) if the intention is not to endorse. MTs were not included in the analysis if they did not indicate an endorsement.



discarded McDonalds bag with card receipt].<sup>13</sup> Both Twitter measures—frequency and measure of broken windows—were entered as time-variant regressors.

#### *Census regressors*

Measures were selected based on previous literature on crime correlates (e.g. [Young 2002](#); [Chainey 2008](#)) and included proportions of the borough populations that were black, minority ethnic, unemployed, aged 15–21 and who had no qualifications. These were entered as time-invariant regressors.

#### *Methods of estimation*

Given the requirement to incorporate the temporal variability of police-recorded crime and Twitter data with the static regressors from the census, we used linear<sup>14</sup> random- and fixed-effects regression.<sup>15</sup> This meant that we could explore correlations between independent regressors including tweets that have high temporal granularity and variability and census regressors that have very low temporal granularity with the dependent measures of police-recorded crime. We took measurements at each *consecutive* month (variable for Twitter regressors and static for census regressors) within each borough (variable for both Twitter and census regressors).<sup>16</sup> We were therefore able to conduct an ecological analysis of London police-recorded crime using Twitter data as a predictor. Random-effects (RE) assume that the boroughs error term is not correlated with the regressors, which allows for time-invariant variables to play a role as explanatory regressors (census measures). However, violation of this assumption renders RE inconsistent because of selection bias resulting from time-invariant unobservables. Fixed-effects (FE) models are based solely on within-borough variation, allowing for the elimination of potential sources of bias by controlling for stable (observed and unobserved) ecological characteristics. However, one side effect of FE models is that they cannot be used to investigate time-invariant causes of the dependent variables. We determined whether RE or FE was more appropriate using the Hausman test. Robust standard errors were used to account for heteroskedasticity.

<sup>13</sup>Publication of actual examples of tweets mentioning ‘broken windows’ indicators is precluded under Twitter Terms of Service. Twitter Terms of Service forbid the anonymization of tweet content (screen-name must always accompany tweet content), meaning that ethically, informed consent should be sought from each tweeter to quote their post in research outputs. However, this is impractical given the number of posts generated and the difficulty in establishing contact (a direct private message can only be sent on Twitter if both parties follow each other). Therefore, it is not ethical to directly quote tweets that identify *individuals* without prior consent. Furthermore, Twitter Terms of Service also requires that authors honour any future changes to user content, including deletion. As academic papers cannot be edited continuously post publication, this condition further complicates direct quotation (needless to mention the burden of checking content changes on a regular basis).

<sup>14</sup>In this study, the linear variant of the RE/FE regression model was chosen. Commonly research that estimates criminal victimization adopts negative binomial modelling to account for the Poisson skewed distribution of counts of crime (with the majority either not experiencing crime or only experiencing a few incidents) and the over-dispersion of these counts (where the conditional variance exceeds the conditional mean). However, because our counts of victimization were pooled into boroughs and months, they did not exhibit a skewed distribution, ruling out this more conventional model choice (see [Osgood 2000](#)).

<sup>15</sup>The Breusch-Pagan Lagrange Multiplier test revealed RE regression was favourable over simple OLS regression.

<sup>16</sup>We built alternative lag models to test if Twitter observations in prior months predicted offline crime rates in the later month. Results indicated a non-lagged model was preferred. This is likely due to the temporal scale chosen, where reports of crime to the police and tweets about neighbourhood disorder are likely to occur in the same month.

### Results

Table 1 reports on the results of the RE and FE models (coefficients in bold indicate those that are favoured (FE or RE) based on the Hausman tests). Model A includes only conventional census predictive regressors that have been established as correlates of certain types criminal activity in previous research (Young 2002; Chainey 2008). Model B introduces the Twitter regressors and differences in the adjusted  $R^2$  statistics<sup>17</sup> illustrate the change in variance explained by their inclusion.<sup>18</sup> Some of the conventional census regressors emerged as predictive in the RE models, and associations are in the direction expected based on previous research. Twitter regressors emerged as significantly associated with prevalence of crime in seven of the nine crime types. The addition of Twitter data increases the amount of variance explained in all models, corroborating hypothesis H1 and adding further evidence in support of the argument that social media communications can add explanatory value in estimating offline phenomena (Asur and Huberman 2010; Gerber 2014). Tweet frequency was positively associated with burglary in a dwelling, criminal damage, violence against the person and theft from shops, corroborating previous work that argues geolocation markers in Twitter data are useful in estimating crime patterns (Bendler *et al.* 2014; Malleson and Andresen 2015). Like Malleson and Andresen, this study finds the positive relationship between frequency of Twitter posts and violence against the person holds when eliminating potential sources of bias by controlling for stable (observed and unobserved) ecological characteristics. These results contradict the work of Bendler *et al.* (2014) who found a negative relationship existed between frequency of tweets and violence. However, in our models that take month and not hour as the temporal scale, it is likely that tweet frequency is acting as an indicator of population density, and not mobile population. This would account for the positive relationship with crimes that tend to occur in the absence of bystanders (burglary and criminal damage). As tweet frequency is not a key variable of interest in this paper, this is not a fundamental shortcoming of this exploratory study.<sup>19</sup>

The key variable of interest, Twitter mentions of ‘broken windows’ indicators, emerged as significantly associated with several of the crime types. However, all relationships in Table 1 suggest that an *absence* of Twitter communications containing signatures of neighbourhood degeneration (graffiti, vandalism, litter etc.) are associated with higher crime rates. In order to explore this relationship further, and to address hypotheses H2 and H3, the sample was split into low- and high-crime boroughs based on an inspection of linear plots of the panel data.<sup>20</sup> Table 2 provides the results of this analysis and shows a pattern of association that supports both hypotheses. Tweets containing mentions of ‘broken windows’ indicators are *positively correlated* with criminal damage, theft from a

<sup>17</sup>The  $R^2$  statistics are only comparable between models using the same estimation method, i.e. FE models are only comparable with other FE models, and RE models are only comparable with other RE models. However, in our study we were only concerned with comparisons between models containing conventional regressors and those containing both conventional and Twitter regressors, using the same estimation method.

<sup>18</sup>A change in variance explained following the addition of new variables means these hold a degree of explanatory power in the dependent variable.

<sup>19</sup>A forthcoming paper from this project provides an alternative analysis that increases the spatial and temporal resolution in order to test the utility of frequency of tweets in estimating crime patterns across all types.

<sup>20</sup>Boroughs above the mean were included in the high-crime group, while those below the mean were included in the low-crime group.

TABLE 1 *Random and fixed effects models for all boroughs*

	Burglary in a dwelling			Burglary in a business property			Criminal damage			
	Model A	Model B	SE	Model A	Model B	SE	Model A	Model B	SE	
	$\beta$	$\beta$	SE	$\beta$	$\beta$	SE	$\beta$	$\beta$	SE	
<i>Random</i>										
Tweet freq.		0.00***	0.00		<b>0.00</b>	<b>0.00</b>		0.00	0.00	0.00
'BW' tweets		-1.03***	0.29		<b>0.01</b>	<b>0.01</b>		-0.84**	0.26	0.26
Prop. ethnic	-0.73	0.41	0.45	<b>-0.09***</b>	<b>0.02</b>	<b>0.02</b>	-0.59	0.47	0.45	0.45
Prop. unemployed	-1.63	6.75	5.97	<b>1.09***</b>	<b>0.22</b>	<b>0.22</b>	-0.45	5.84	4.93	4.93
Prop. 15-21	0.00***	0.00	0.00	<b>0.00***</b>	<b>0.00</b>	<b>0.00</b>	0.00***	0.00	0.00	0.00
Prop. no qual	0.24	1.11	1.11	<b>-0.42***</b>	<b>0.09</b>	<b>0.08</b>	-1.10	1.29	1.23	1.23
<i>Fixed</i>										
Total tweets		<b>0.00**</b>	<b>0.00</b>		0.00	0.00		<b>0.00**</b>	<b>0.00</b>	<b>0.00</b>
'BW' tweets		<b>-0.63***</b>	<b>0.10</b>		0.02	0.02		<b>-0.43*</b>	<b>0.21</b>	<b>0.21</b>
Adjusted $R^2$	0.48	0.67		0.32	0.40		0.52	0.70		
	Theft of personal property			Theft from a motor vehicle			Theft of a motor vehicle			
	Model A	Model B	SE	Model A	Model B	SE	Model A	Model B	SE	SE
	$\beta$	$\beta$	SE	$\beta$	$\beta$	SE	$\beta$	$\beta$	SE	SE
<i>Random</i>										
Tweet freq.		0.00**	0.00		<b>0.00</b>	<b>0.00</b>		0.00	0.00	0.00
'BW' tweets		-0.28*	0.11		<b>-0.78**</b>	<b>0.26</b>		-0.34***	0.07	0.07
Prop. Ethnic	-0.17	0.27	0.26	<b>-0.60</b>	<b>0.33</b>	<b>0.36</b>	-0.83**	0.26	0.23	0.23
Prop. unemployed	11.78***	3.33	3.02	<b>2.52</b>	<b>5.93</b>	<b>5.04</b>	9.40**	3.53	2.76	2.76
Prop. 15-21	0.00	0.00	0.00	<b>0.00***</b>	<b>0.00</b>	<b>0.00</b>	0.00**	0.00	0.00	0.00
Prop. no qual	-3.62*	1.56	1.25	<b>-3.71**</b>	<b>1.33</b>	<b>1.12</b>	-1.60	0.90	0.90	0.90
<i>Fixed</i>										
Total tweets		<b>0.00</b>	<b>0.00</b>		0.00*	0.00		<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
'BW' tweets		<b>-0.11</b>	<b>0.08</b>		-0.45*	0.21		<b>-0.19</b>	<b>0.10</b>	<b>0.10</b>
Adjusted $R^2$	0.30	0.68		0.42	0.57		0.38	0.51		

TABLE 1 *Continued*

	Possession of drugs		Violence against the person		Theft from shops	
	Model A	Model B	Model A	Model B	Model A	Model B
	$\beta$	SE	$\beta$	SE	$\beta$	SE
<i>Random</i>						
Tweet freq.		0.00		0.00		0.00
'BW' tweets		-0.51***		-2.26*		-0.08
Prop. ethnic	0.98	1.12	-0.32	1.84	-0.63	0.79
Prop. unemployed	8.90	8.25	22.34	18.46	-1.36	6.59
Prop. 15-21	0.00	0.00	0.01***	0.00	0.00*	0.00
Prop. no qual	-5.48*	2.98	-8.32	4.67	-6.02*	3.17
<i>Fixed</i>						
Total tweets		<b>0.00</b>		<b>0.00**</b>		<b>0.00*</b>
'BW' tweets		<b>-0.46**</b>		<b>-0.56</b>		<b>0.05</b>
Adjusted $R^2$	0.34		0.50		0.09	
						0.42

The coefficients in bold indicate those that are favoured (FE or RE) based on the Hausman tests. Robust standard errors are presented.  
 \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

TABLE 2 Random and fixed effects models for low- and high-crime boroughs

	Burglary in a dwelling			Burglary in a business property			Criminal damage				
	High		SE	High		SE	Low		SE	High	
	$\beta$	SE		$\beta$	SE		$\beta$	SE		$\beta$	SE
<i>Random</i>											
Tweet freq.	<b>0.00**</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00**</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
'BW' tweets	<b>0.45</b>	<b>0.73</b>	<b>0.70</b>	<b>0.03</b>	<b>-0.03**</b>	<b>0.01</b>	<b>0.20</b>	<b>0.29</b>	<b>-0.26</b>	<b>1.52</b>	<b>1.39</b>
Prop. ethnic	<b>-1.25</b>	<b>1.85</b>	<b>0.25</b>	<b>0.01</b>	<b>0.26**</b>	<b>0.04</b>	<b>-1.16</b>	<b>0.66</b>	<b>-3.98**</b>	<b>15.83</b>	<b>15.83</b>
Prop. unemployed	<b>0.45</b>	<b>9.49</b>	<b>10.21**</b>	<b>0.30</b>	<b>-0.87**</b>	<b>0.37</b>	<b>7.20</b>	<b>7.36</b>	<b>29.81</b>	<b>0.00</b>	<b>0.00</b>
Prop. 15-21	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00**</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Prop. no qual	<b>1.14</b>	<b>2.19</b>	<b>-6.38**</b>	<b>0.04</b>	<b>-0.13</b>	<b>0.19</b>	<b>1.00</b>	<b>1.50</b>	<b>7.13</b>	<b>4.34</b>	<b>4.34</b>
<i>Fixed</i>											
Total tweets	<b>0.00**</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00*</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00*</b>	<b>0.00</b>	<b>0.00</b>
'BW' tweets	<b>-0.29</b>	<b>0.48</b>	<b>-1.57</b>	<b>0.03</b>	<b>0.12</b>	<b>0.09</b>	<b>0.51**</b>	<b>0.16</b>	<b>-0.26</b>	<b>0.61</b>	<b>0.61</b>
	Theft of personal property			Theft from a motor vehicle			Theft of a motor vehicle				
	High		SE	High		SE	Low		SE	High	
	$\beta$	SE		$\beta$	SE		$\beta$	SE		$\beta$	SE
<i>Random</i>											
Tweet freq.	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00*</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
'BW' tweets	<b>0.05</b>	<b>0.12</b>	<b>0.42</b>	<b>0.72</b>	<b>8.71*</b>	<b>3.80</b>	<b>0.12</b>	<b>1.42</b>	<b>-0.30*</b>	<b>0.14</b>	<b>0.14</b>
Prop. ethnic	<b>-0.01</b>	<b>0.09</b>	<b>-10.14**</b>	<b>0.77</b>	<b>4.55***</b>	<b>0.71</b>	<b>-0.41**</b>	<b>0.07</b>	<b>0.69*</b>	<b>0.34</b>	<b>0.34</b>
Prop. unemployed	<b>3.73*</b>	<b>1.90</b>	<b>39.26***</b>	<b>5.10</b>	<b>-18.25*</b>	<b>8.23</b>	<b>11.74***</b>	<b>3.35</b>	<b>-6.73*</b>	<b>3.26</b>	<b>3.26</b>
Prop. 15-21	<b>0.00**</b>	<b>0.00</b>	<b>-0.02**</b>	<b>0.00</b>	<b>0.00***</b>	<b>0.00</b>	<b>0.00***</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
Prop. no qual	<b>-0.63*</b>	<b>0.30</b>	<b>58.76***</b>	<b>5.40</b>	<b>-0.65</b>	<b>1.67</b>	<b>0.68***</b>	<b>0.16</b>	<b>-0.19</b>	<b>1.05</b>	<b>1.05</b>
<i>Fixed</i>											
Total tweets	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
'BW' tweets	<b>0.08</b>	<b>0.10</b>	<b>0.42</b>	<b>0.67</b>	<b>8.71</b>	<b>3.52</b>	<b>0.12</b>	<b>1.33</b>	<b>-0.74</b>	<b>0.66</b>	<b>0.66</b>

TABLE 2 *Continued*

	Possession of drugs				Violence Against the Person				Theft from Shops					
	Low		High		Low		High		Low		High			
	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE		
<i>Random</i>														
Tweet freq.	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>
'BW' tweets	<b>0.89*</b>	<b>0.38</b>	<b>0.83</b>	<b>1.30</b>	<b>3.45***</b>	<b>0.74</b>	<b>0.74</b>	<b>2.47</b>	<b>0.00</b>	<b>1.74</b>	<b>-0.17</b>	<b>0.09</b>	<b>0.37</b>	<b>1.64</b>
Prop. ethnic	<b>-0.28</b>	<b>0.48</b>	<b>2.45***</b>	<b>0.14</b>	<b>0.18</b>	<b>0.89</b>	<b>0.89</b>	<b>-1.13</b>	<b>-0.27</b>	<b>2.84</b>	<b>-0.27</b>	<b>0.15</b>	<b>10.22**</b>	<b>0.32</b>
Prop. unemployed	<b>5.98</b>	<b>5.27</b>	<b>48.39**</b>	<b>8.07</b>	<b>15.97</b>	<b>15.56</b>	<b>15.56</b>	<b>-49.77**</b>	<b>-1.64</b>	<b>14.86</b>	<b>-1.64</b>	<b>0.95</b>	<b>0.00*</b>	<b>0.00</b>
Prop. 15-21	<b>0.00</b>	<b>0.00</b>	<b>-0.01***</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>-0.02***</b>	<b>0.00</b>
Prop. no qual	<b>-1.60</b>	<b>1.07</b>	<b>-0.19</b>	<b>1.05</b>	<b>6.26**</b>	<b>2.94</b>	<b>2.94</b>	<b>-0.24</b>	<b>2.16***</b>	<b>15.84</b>	<b>2.16***</b>	<b>0.28</b>	<b>39.16**</b>	<b>1.12</b>
<i>Fixed</i>														
Total tweets	0.00	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.00**</b>	0.00	<b>0.00</b>	0.00	0.00	0.00	0.00
'BW' tweets	0.43	0.31	0.83	1.21	4.19**	1.22	1.22	<b>1.54</b>	-0.16	<b>1.50</b>	-0.16	0.11	0.37	1.52

The coefficients in bold indicate those that are favoured (FE or RE) based on the Hausman tests. Robust standard errors are presented.

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .



motor vehicle, possession of drugs and violence *in low-crime areas*. Conversely, tweets of this nature were *negatively correlated* with burglary in a dwelling, burglary in a business property and theft of a motor vehicle *in high-crime areas*. This pattern is in line with offline research, which suggests discussions of neighbourhood degeneration and local crime issues at community meetings are not representative of local crime problems (e.g. [Brunger 2011](#); [Sagar and Jones 2013](#)). It is possible that residents in low-crime areas are more sensitive to signs of neighbourhood degeneration and therefore feel motivated to broadcast instances of littering, graffiti and vandalism via social media, while residents in high-crime areas are less motivated to express similar observations as they are not out of the ordinary (i.e. residents have become desensitized to neighbourhood decline). The next section discusses this argument further, and critically evaluates the utility of open-source communications derived from social media for estimating crime patterns.

### *Discussion*

This exploratory study was developed out of a project that sought to innovate with new forms of data in the estimation of crime patterns. The models provide some preliminary, but nevertheless, encouraging results that indicate open-source communications, in particular from Twitter, have potential for measuring the breakdown of social and physical order at the borough level. The results show that the inclusion of Twitter data increases the amount of variance explained in the crime estimation models, lending support to the first hypothesis and supporting [Gerber's \(2014\)](#) study that estimated crime in Chicago using social media data as predictors. It was possible to create a Twitter measure of 'broken windows' using a text classification procedure that was verified by 700 human annotators in an online crowdsourcing exercise. The association of the measure with a range crime types can be explained in several ways. It is possible that tweeters sense degradation in the local area, and this is associated with increased crime rates.<sup>21</sup> If this is the case, then it would suggest further support for the broken windows thesis, that is, if we accept the proxy measure of broken windows via social media. This argument certainly seems to hold for residents in low-crime areas in relation to certain offences. But the argument does not hold for high-crime areas. This can be explained in terms of differences in disposition to report local issues of crime and disorder, a pattern found in offline settings.<sup>22</sup> This can be considered a form of reporting bias, and in our study this can have four elements: (1) varying perceptions of local signs of degeneration; (2) knowledge of Twitter; (3) tendency to use Twitter; and (4) tendency to broadcast such issues on Twitter. In relation to the first form of bias, [O'Brien et al. \(2015\)](#) found that in their use of big administrative data to develop measures of broken windows, concern for public space varied by neighbourhood and that such variation related to differences in people's perception of disorder. This suggests residents vary in their perception of neighbourhood degeneration and this could shape their responses and actions in relation to reporting. In relation to (2) and (3), [O'Brien et al.](#) also recognized

<sup>21</sup>It is of course possible that Twitter users in low-crime areas are both more likely to report crime (e.g. graffiti and vandalism) and degradation on social media *and* to the police. Twitter therefore may be acting as an informal form of reporting that is followed up (or preceded by) a formal report of a crime to the police.

<sup>22</sup>Here, we focus on a location-based explanation. Future research should explore how other explanations, such as gender, age, ethnicity and so on, mediate decisions to self-report or not.

variability in knowledge of the administrative system for reporting local issues, and propensity to use it. In relation to Twitter, we know that propensity to use the platform varies by socio-demographic and economic factors. In particular, previous work of ours shows that younger people are more likely than older people to use Twitter (Sloan *et al.* 2015). We also know that of those that do use Twitter, there are significant differences in using geolocation services by age, and that propensity to geolocate is also influenced by attendance at events and travelling (Sloan and Morgan, 2015).<sup>23</sup> Furthermore, changes in technology, such as the release of new mobile phone handsets and software updates of the Twitter app, have also been shown to impact the number of users including geolocation data in their tweets (Swier *et al.* 2015). While weighting and calibration methods may mitigate these inherent forms of bias, the same techniques cannot address the final form of bias (4), propensity to use Twitter to report neighbourhood degeneration. Hypotheses H2 and H3 specifically tested whether such bias is present, and the results show it is possible that in low-crime areas some residents have a sensitivity to local signs of degeneration and have a propensity to broadcast such signs via social media. In order to adjust for propensity to use the administrative system to report neighbourhood degeneration, O'Brien *et al.* developed a method using auxiliary measures from within the same database to estimate the extent of the bias and to help account for over- or under-reporting. A key part of achieving this was to calculate the number of residents who knew about the system by estimating the proportion of the whole population who registered to use it. Given that population figures for Twitter are not available, such corrections were not possible in this study.

### *Limitations*

There is little doubt that social media analysis marks a significant departure for criminologists. Computational methods allow for the capture of naturally occurring data at the level of populations in near-real-time, affording criminologists the ability to render populations visible and thinkable in both their locomotive (in motion) and reactive states. Furthermore, these new data may also provide access to hitherto difficult to reach, if not invisible, populations. It is well established, for example, that young male residents of urban neighbourhoods are systematically underrepresented in conventional survey methods, yet they regularly use smartphones.<sup>24</sup> However, despite an abundance of work using computational statistics, social statistical research in the big data field is nascent, and initial findings, including those found in this paper, point to potential forms of bias that are not simple to adjust for. We conclude that this is a significant limitation of using social media data for the estimation of crime patterns, and therefore if employed in predictive efforts, they must be used in conjunction with existing forms of 'trusted' data.

<sup>23</sup>Geolocated tweets from travellers may be a more problematic in London than in other cities given its popularity as a destination.

<sup>24</sup>The advent of smartphones on cheaper pay-as-you-go services resulted in over 97 per cent of the public owning mobile phones in 2009, with just under half of these using smartphone functions in 2011 (Dutton and Blank 2011). This unprecedented access has seen the socio-economic digital divide close rapidly, being filled by excluded and disenfranchised youth (Williams *et al.* 2013).

The spatial and temporal units in our models were chosen in order to examine the utility of ‘broken windows’ indicators found in Twitter data to perform an ecological study of crime in London. In particular, we felt it reasonable to assume that taking consecutive months as the temporal scale would allow for sufficient time for a Twitter report of local disorder and a crime report to coincide (hence why we did not employ a cross-lag model<sup>11</sup>). We also felt it reasonable to assume a borough-wide analysis would capture tweets about neighbourhood degeneration and associated reported crimes. These spatial and temporal scales were also required to generate a large enough number of geolocated Twitter posts for analysis. We acknowledge that the choice of scale in longitudinal and multilevel modelling can impact results. In future research, we intend on experimenting with various temporal and spatial resolutions to identify the best compromise between model fit and number of data points. Furthermore, our choice of spatial scale allowed us to perform an ecological analysis, and not an individual analysis of crime in London using Twitter data as a predictor. Ecological studies are designed to tap into proxies of social processes that lead to increases in crime. Therefore, despite the fact that RE and FE models appear to provide an advantage in assessing neighbourhood change, we do not claim causal relationships.

Our findings are based on social media data collected from a single social networking site. Despite the high number of tweets collected, arguably additional data from sites such as Facebook, the largest global social network, would have provided a more ‘census-like’ coverage of the population under study. Regretfully, data from Facebook are not freely available and the level of detail (such as geolocation) varies significantly compared to Twitter. We also used police-recorded crime as our dependent measure given that its temporal and spatial resolution (i.e. down to the second and metre) was compatible with social media data. However, we are aware that police-recorded crime is an artefact of various flawed mechanisms, such as counting rules, police discretion and reporting behaviour, resulting in the so called ‘dark figure’ of crime (the Crime Survey for England and Wales routinely shows that these data represent about half of all UK crime). Our models therefore failed to take into account all crimes in London boroughs. In future research, it would also be of interest to examine associations between estimates of disorder from social media and other offline sources, such as the CSEW and Metropolitan Police Public Attitudes Survey (METPAS), presenting an opportunity to partially validate online measures of ‘broken windows’.<sup>25</sup> Finally, we only examined crime in London, and future research should test whether the patterns found in the models reported hold for other urban areas in the United Kingdom and beyond.

While bias in social media data does question its utility in estimating offline crime patterns across *both* low- and high-crime areas (that is, until we can develop reliable adjustments), it does not rule out the use of social media in the study of crime-related topics. For example, raw social media data can be used to study conflict and abuse between users within online networks. Here, the bias is precisely what a criminologist is interested in; the propensity to send abusive content and to react is the focus of the analysis, and something that should not be adjusted for (see [Burnap and Williams 2015; 2016; Williams and Burnap 2015](#)). In the immediate term, where social media data are used to estimate offline crime patterns, account should be taken of this bias (i.e. estimates may be more reliable for low-crime areas compared with high-crime areas) and they should be used in conjunction with other sources, such as curated and administrative data, in order to mitigate any biased findings with conventional wisdom.

<sup>25</sup>We thank the anonymous reviewer for this suggestion.

### *Conclusions*

This paper has demonstrated that an association exists between aggregated open-source communications data and aggregated police-recorded crime data in London boroughs. It has also highlighted that multiple sources of bias (propensity to use Twitter, propensity to tweet about crime issues, propensity to geolocate posts etc.) are likely to be present and would require suitable adjustments to be made before reliable estimates can be drawn using Twitter data in particular. Of course, we are conscious that predictions made by machines are not always accurate. Recent grand claims have been made that big data make theory and scientific method obsolete. Yet high-profile failures of big data, such as the inability to predict the US housing bubble in 2008 and the spread of influenza across the United States using Google search terms, have resulted in many questioning the power of these new forms of data (Lazer *et al.* 2014). Many of these efforts to predict tend to examine direct and independent effects of mechanisms in statistical models. In doing so, they operate under the caveat ‘all models are wrong but some are useful’ (Box and Draper 1987: 424). It is likely that mechanisms do not operate independently and that they in practice interact, sometimes in feedback loops, in a contingent manner over time and space. Furthermore, purely data-driven approaches tend to produce models and algorithms that are over fit to the idiosyncrasies of a particular data set, leading to spurious results that often do not reflect reality (O’Brien *et al.* 2015). Social media networks offer socially relevant data at unprecedented scale and speed, but with these affordances come the challenges of ‘taming’ data, filtering out the noise and transforming content to serve research needs. This inevitably involves the use of machines to automate processes traditionally undertaken by social scientists. Machines are being used to collect, store and analyse (classify) these data, and as social scientists we must routinely test and question automated decisions. In particular, the design of algorithms that drive these automated processes must involve social scientists from the outset, especially when social science theory is being used to guide collection and analysis. If these algorithms are to be deployed in longitudinal studies or operational settings, then their components (i.e. lexicons, keywords, parts-of-speech etc.) used must be routinely refreshed and validated via established computational and social science methods. In this paper, we have attempted to address some of these issues by using theory to guide our model development, thus avoiding the default approach in big data research that is wholly data driven in the effort to predict (Chan and Bennet-Moses 2015). The key methodological recommendation from this paper is that strict checks and balances need to be put in place when dealing with big data in criminological research, such as identifying and calibrating for bias, augmenting big data with conventional sources and drawing on theory. Without theory driven big data collection, transformation and analysis, we cannot answer the substantive questions about social processes and mechanisms that concern criminologists.

### *Funding*

This work was supported by the ESRC National Centre for Research Methods Grant: ‘Social Media and Prediction: Crime Sensing, Data Integration and Statistical Modelling’ (ES/F035098/1/512589112).

## ACKNOWLEDGEMENT

We are grateful to Alex Sutherland (Cambridge University and Rand Europe) for commenting on earlier drafts of this paper.

## REFERENCES

- ABBOTT, A. (1997), 'Of Time and Space: The Contemporary Relevance of the Chicago School', *Social Forces*, 75: 1149–82.
- ASUR, S. and HUBERMAN, B. A. (2010), 'Predicting the future with social media', WI-IAT '10 Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01, Pages 492–499.
- BENDLER, J., BRANDT, T., WAGNER, S. and NEUMANN, D. (2014), 'Investigating Crime-to-Twitter Relationships in Urban Environments—Facilitating a Virtual Neighbourhood Watch', *Proceedings of the European Conference on Information Systems (ECIS)*, Tel Aviv, June 9–11.
- BIS. (2013), *Seizing the Data Opportunity: A Strategy for UK Data Capability*. Department of Business, Innovation and Skills.
- BOYD, D. (2014), *It's Complicated: The Social Lives of Networked Teens*. Yale University Press.
- BOX, G. E. P. and DRAPER, N. R. (1987), *Empirical Model-Building and Response Surfaces*. Wiley.
- BRAGA, A. A. and BOND, B. J. (2008), 'Policing Crime and Disorder Hot Spots: A Randomized Controlled Trial', *Criminology*, 46: 577–607.
- BRAGA, A., PAPACHRISTOS, A. AND HUREAU, D. (2012), *Hot Spots Policing Effects on Crime*. Campbell Systematic Reviews.
- BRUNGER, M. (2011), 'Governance, Accountability and Neighbourhood Policing in Northern Ireland: Analysing the Role of Public Meetings', *Crime, Law and Social Change*, 55: 105–20.
- BURNAP, P. and WILLIAMS, M. L. (2015), 'Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making', *Policy & Internet*, 7: 223–242.
- BURNAP, P. and WILLIAMS, M. L. (2016), 'Us and Them: Identifying Cyber Hate on Twitter across Multiple Protected Characteristics', *EPJ Data Science*, 5:11, available online at <http://link.springer.com/article/10.1140/epjds/s13688-016-0072-6>.
- BURNAP, P. WILLIAMS, M. L. RANA, O., EDWARDS, A., AVIS, N., MORGAN, J., HOUSLEY, W. and SLOAN, L. (2014a) COSMOS: Towards an Integrated and Scalable Service for Analysing Social Media on Demand, *International Journal of Parallel, Emergent and Distributed Systems*, 30: 80–100.
- BURNAP, P. WILLIAMS, M. L. and SLOAN, L. (2014b), 'Tweeting the Terror: Modelling the Social Media Reaction to the Woolwich Terrorist Attack', *Social Network Analysis and Mining*, 4: 1–14.
- CHAN, J. and BENNET-MOSES, L. (2015), 'Is Big Data Challenging Criminology?', *Theoretical Criminology*. Published online 19 May 2015, doi:10.1177/1362480615586614.
- CHAINEDY, S. (2008), 'Identifying Priority Neighbourhoods Using the Vulnerable Localities Index', *Policing*, 2: 196–209.
- CHOUDHARY, A., HENDRIX, W., LEE, K., PALSETIA, D. and LIAO, W. (2012), 'Social Media Evolution of the Egyptian Revolution', *Communications of the ACM*, 55: 74–80.
- DUTTON, W. H., and BLANK, G. (2011), *Next Generation Users: The Internet in Britain*. Oxford Internet Survey 2011, Oxford Internet Institute, University of Oxford.



- GERBER M. S. (2014), 'Predicting Crime Using Twitter and Kernel Density Estimation', *Decision Support Systems*, 61: 115–25.
- HMIC. (2011), *Policing Public Order: An Overview and Review of Progress against the Recommendations of Adapting to Protest and Nurturing the British Model of Policing*. HMIC.
- LAZER, D., KENNEDY, R., KING, G. and VESPIGNANI, A. (2014), 'The Parable of Google Flu: Traps in Big Data Analysis', *Science*, 343: 1203–5.
- LAZER, D., PENTLAND, A., ADAMIC, L., ARAL, S., BARABASI, A-L., BREWER, D., CHRISTAKIS, N. A., CONTRACTOR, N., FOWLER, J., GUTMANN, M., JEBARA, T., KING, G., MACY, M., ROY, D. and VAN ALSTYNE, M. (2009), 'Computational Social Science', *Science*, 323: 721–23.
- Library of Congress. (2013), 'White Paper: Update on the Twitter Archive at the Library of Congress', *Library of Congress*, available online at [http://www.loc.gov/today/pr/2013/files/twitter\\_report\\_2013jan.pdf](http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf) (last accessed 10 January 2013).
- LOTAN, G., GRAEFF, E., ANANNY, M., GAFFNEY, D., PEARCE, I. and BOYD, D. (2011), 'The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions', *International Journal of Communication*, 5: 1375–405.
- MALLESON, N. and ANDRESEN, M. (2015), 'The Impact of Using Social Media Data in Crime Rate Calculations: Shifting Hot Spots and Changing Spatial Patterns', *Cartography and Geographic Information Science*, 42: 112–21.
- NPIA. (2009), *NPIA Business Plan 2009–10*. National Policing Improvement Agency.
- O'BRIEN, D. and SAMPSON, R. J. (2015), 'Public and Private Spheres of Neighborhood Disorder: Assessing Pathways to Violence Using Large-Scale Digital Records', *Journal of Research in Crime and Delinquency*, 52: 486–510.
- O'BRIEN, D. T., SAMPSON, R. J. and WINSHIP, C. (2015), 'Ecometrics in the Age of Big Data: Measuring and Assessing 'Broken Windows' Using Large-scale Administrative Records', *Sociological Methodology*, 45: 101–47.
- OSGOOD, W. D. (2000), 'Poisson-Based Regression Analysis of Aggregate Crime Rates', *Journal of Quantitative Criminology*, 16: 21–43.
- QUINTON, P. and TUFFIN, R. (2007), 'Neighbourhood change: The impact of the National Reassurance Policing Programme', *Policing: A Journal of Policy and Practice*, 1: 149–60.
- RAUDENBUSH, S. W. and SAMPSON, R. J. (1999), 'Ecometrics: Toward a Science of Assessing Ecological Settings, with Application to the Systematic Social Observation of Neighborhoods', *Sociological Methodology*, 29: 1–41.
- SAGAR, T. and JONES, D. (2013), 'Priorities for the Minority? Street-based Sex Work and Partnerships and Communities Together (PACT)', *Criminology and Criminal Justice*, 13: 431–45.
- SAKAKI, T., OKAZAKI, M. and MATSUO, Y. (2010), 'Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors', *Proceedings of the 19th International Conference on World Wide Web*. ACM Press.
- SAMPSON, R. J. (2012), *Great American City: Chicago and the Enduring Neighborhood Effect*. University of Chicago Press.
- SAMPSON, R. J. and RAUDENBUSH, S. W. (2004), 'Seeing Disorder: Neighborhood Stigma and the Social Construction of "Broken Windows"', *Social Psychology Quarterly*, 67: 317–42.
- SAVAGE, M. and BURROWS, R. (2007), 'The Coming Crisis of Empirical Sociology', *Sociology*, 41: 885–99.
- SKOGAN, W. (1990). *Disorder and Decline: Crime and the Spiral of Decay in American Cities*. University of California Press.



- . (2015), 'Disorder and Decline: The State of Research', *Journal of Research in Crime and Delinquency*, 52: 464–85.
- SLOAN, L. and MORGAN, J. (2015), 'Who Tweets with Their Location? Understanding the Relationship Between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter', *PLoS One*. Published online 6 November 2015, doi:10.1371/journal.pone.0142209.
- SLOAN, L., MORGAN, J., BURNAP, P. and WILLIAMS, M. L. (2015), 'Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data', *PLoS One*. Published online 2 March 2015, doi:10.1371/journal.pone.0115545.
- SLOAN, L., MORGAN, J., WILLIAMS, M. L., HOUSLEY, W., EDWARDS, A., BURNAP, P. and RANA, O. (2013), 'Knowing the Tweeters: Deriving Sociologically Relevant Demographics from Twitter', *Sociological Research Online*, 18: 7.
- SMITH, D. (2012), 'How Many People Use the Top Social Media?', *Digital Market Ramblings*, available online at <http://expandedramblings.com/index.php/resource-how-many-people-use-the-top-social-media/> (last accessed 15 July 2012).
- SWIER, N., KOMARNICZKY, B. and CLAPPERTON, B. (2015), *Using Geolocated Twitter Traces to Infer Residence and Mobility*. Office for National Statistics.
- TUMASJAN, A.; SPRENGER, T.; SANDNER, P. G. and WELPE, I. M. (2010), 'Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment', *Proceedings of 4th ICWSM*. AAAI Press, 178–85.
- VAN RIJSBERGEN, C. J. (1979), *Information Retrieval*, 2nd ed. Butterworth.
- WELSH, B. C., BRAGA, A. and BRUINSMA, G. (2015), 'Reimagining Broken Windows: From Theory to Policy', *Journal of Research in Crime and Delinquency*, 52: 447–63.
- WILLIAMS, M. L. (2015), 'Towards an Ethical Framework for Using Social Media Data in Social Research', presented at Social Research Association Workshop, Institute of Education, 15 June 2015, available at <http://the-sra.org.uk/wp-content/uploads/ethics-in-social-media-research-matthew-williams.pdf>.
- WILLIAMS, M. L. and BURNAP, P. (2015), 'Cyberhate on Social Media in the Aftermath of Woolwich: A Case Study in Computational Criminology and Big Data', *British Journal of Criminology*, 56: 211–38.
- WILLIAMS, M. L., EDWARDS, A., HOUSLEY, W., BURNAP, P., RANA, O., AVIS, N., MORGAN, J. and SLOAN, L. (2013), 'Policing Cyber-neighbourhoods: Tension Monitoring and Social Media Networks', *Policing and Society*, 23: 461–81.
- WILSON, J. Q. and KELLING, G. L. (1982), 'The Police and Neighborhood Safety: Broken Windows', *Atlantic Monthly*, 127: 29–38.
- YOUNG J. (2002), 'Crime and Social Exclusion', in M. Maguire, R. Morgan and R. Reiner, eds, *The Oxford Handbook of Criminology*, 3rd edn. Oxford University Press.